LNCS 9891

Katrien Verbert
Mike Sharples
Tomaž Klobučar (Eds.)

# Adaptive and Adaptable Learning

11th European Conference on Technology Enhanced Learning, EC-TEL 2016
Lyon, France, September 13–16, 2016, Proceedings

Springer

# Lecture Notes in Computer Science 9891

*Commenced Publication in 1973*
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

More information about this series at http://www.springer.com/series/7409

Katrien Verbert · Mike Sharples
Tomaž Klobučar (Eds.)

# Adaptive and Adaptable Learning

## 11th European Conference on Technology Enhanced Learning, EC-TEL 2016
Lyon, France, September 13–16, 2016
Proceedings

Springer

*Editors*
Katrien Verbert
KU Leuven
Leuven
Belgium

Tomaž Klobučar
Jožef Stefan Institute
Ljubljana
Slovenia

Mike Sharples
The Open University
Milton Keynes
UK

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-45152-7 ISBN 978-3-319-45153-4 (eBook)
DOI 10.1007/978-3-319-45153-4

Library of Congress Control Number: 2016948275

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

# Preface

The 11th edition of the European Conference on Technology-Enhanced Learning (EC-TEL) was held in Lyon (France) during September 13–16, 2016. This volume collects all peer-reviewed contributions that were included in the exciting program of this year's conference.

In the 11th year of its existence, EC-TEL has become the major interdisciplinary venue for the community of technology-enhanced learning (TEL) researchers in Europe and worldwide. Furthermore, EC-TEL is a shared opportunity for researchers, practitioners, educational developers, and policy makers to address current challenges and advances in the field. Since 2006, EC-TEL has provided a reference point for relevant state-of-the art research in TEL; first in Crete (Greece, also in 2007), and then in Maastricht (The Netherlands, 2008), Nice (France, 2009), Barcelona (Spain, 2010), Palermo (Italy, 2011), Saarbrücken (Germany, 2012), Paphos (Cyprus, 2013), Graz (Austria, 2014), and Toledo (Spain, 2015).

In these uncertain and turbulent times, it is essential for individuals and organizations continually to adapt and change. The theme of EC-TEL 2016 was "Adaptive and Adaptable Learning." It highlighted developments in learning systems that adapt to the needs, interests, and abilities of each learner, toward a vision of learning that is personalized yet social. Effective technology enhanced learning must also be adaptable – resilient, flexible, and sustainable to meet rapidly changing needs, technologies, contexts, and policies. The conference explored how research in collaborative and personalized learning could be combined with new developments in analytics, interaction design, social, mobile and ubiquitous technologies, and visualization techniques, to enhance learning for everyone.

Drawing on the core TEL disciplines of computer science, education, psychology, cognitive science, and social science, research contributions presented at EC-TEL 2016 addressed topics such as adaptive and adaptable learning, collaborative knowledge building, motivation and engagement, collaborative learning, game-based learning, lifelong learning, intelligent learning systems, recommender systems, learning design, learning analytics, assessment for learning, social computing and social media, massive open online courses (MOOCs), and wearable and pervasive technologies.

This 2016 edition was again extremely competitive, given the high number of submissions generated. A total of 148 valid paper submissions were received. Of these, 102 were full papers. All submissions were assigned to at least three members of the Program Committee (PC) for review. One of the reviewers had the role of leading reviewer and initiated a discussion in the case of conflicting reviews. All reviews as well as the discussions were checked and discussed within the team of PC chairs, and additional reviews or meta-reviews were elicited if necessary. From this process, 26 submissions were selected as full papers (resulting in an acceptance rate for full papers of 25 %). Additionally, 23 papers were chosen as short papers, eight as demonstrations, and 33 as posters. Table 1 shows the detailed statistics.

**Table 1.** Acceptance rate in different submission categories

| | | Published as | | | |
|---|---|---|---|---|---|
| Submitted as | | Full Paper | Short Paper | Poster Paper | Demo Paper |
| Full Paper | 102 | 26 | 16 | 21 | 2 |
| Short Paper | 31 | | 7 | 8 | |
| Poster Paper | 9 | | | 4 | |
| Demo Paper | 6 | | | | 6 |
| Sum | 148 | 26 | 23 | 33 | 8 |

The dedicated work of all the PC members as well as the additional reviewers must be acknowledged. Only with their help was it possible to deal with the high number of submissions and still meet all deadlines as originally planned.

Keynote presentations completed this competitive scientific program. Pierre Dillenbourg from the EPFL Center for Digital Education, Switzerland, gave a presentation on "How Does TEL Research Inform the Design of Educational Robots?" and Vincent Aleven from Carnegie Mellon University presented on "Adaptivity in Learning Technologies: Kinds, Effectiveness, and Authoring." A keynote from the European Commission covered policy aspects of technology enhanced learning.

A plenary panel session was held on the theme of the conference – Adaptive and Adaptable Learning. Two invited panelists from the artificial intelligence and education community, Benedict du Boulay and Rose Luckin, joined the researchers from the TEL community.

Demonstrations and posters had a pronounced role in the conference program. A plenary session was organized as a "TEL demo shootout" in which the demonstrations were presented to arouse the audience's curiosity and highlight the unique aspects. Later on, the demonstrations were shown in action, giving participants the opportunity for hands-on experience, sparking discussions between researchers, practitioners, and educational developers, providing a basis to vote for the best demo. A plenary session was dedicated to an exhibition of posters, to foster discussion about work in progress and research issues. Representatives from the industry also presented and discussed their contributions to the field in the industry track.

The TEL community proposed and organized a set of stimulating workshops as part of the conference. In all, nine workshops were selected from the proposals and were organized. Some of them continue a series of well-established workshops on motivational and affective aspects in TEL and on awareness and reflection in TEL. Others, like Pedagogical Grounded Learning Analytics Design, were new for 2016. A doctoral consortium was organized concurrently with the workshops, which provided an opportunity for PhD students to discuss their work with experienced TEL researchers.

We would like to thank the many contributors for creating a stimulating conference of high quality. These include foremost the authors, the PC members and reviewers, and the conference chairs, who all contributed to the program. We would also like to thank an enthusiastic and dedicated local organization team who made EC-TEL a smooth and memorable experience. The conference was partially supported by the European Association of Technology-Enhanced Learning (EATEL), Springer, and EasyChair.

September 2016

Mike Sharples
Katrien Verbert
Tomaž Klobučar

# Organization

## Executive Committee

### General Chair

| | |
|---|---|
| Tomaž Klobučar | Jožef Stefan Institute, Slovenia |

### Program Chairs

| | |
|---|---|
| Mike Sharples | The Open University, UK |
| Katrien Verbert | KU Leuven, Belgium |

### Steering Committee Representative

| | |
|---|---|
| Ralf Klamma | RWTH Aachen University, Germany |

### Demo and Poster Chairs

| | |
|---|---|
| Marco Kalz | Open University, The Netherlands |
| Viktoria Pammer-Schindler | TU Graz and Know-Center, Austria |

### Workshop Chair

| | |
|---|---|
| Christian Glahn | University of Applied Sciences HTW Chur, Switzerland |

### Industry Chairs

| | |
|---|---|
| Franck Tarpin Bernard | SBT Group, France |
| Pierre Dubuc | OpenClassrooms, France |

### Doctoral Consortium Chairs

| | |
|---|---|
| Ulrike Lucke | Universität Potsdam, Germany |
| Katherine Maillet | Institut Mines-Télécom, Télécom Ecole de Management, France |

### Dissemination Chairs

| | |
|---|---|
| Christine Michel | INSA de Lyon, France |
| Karim Sehaba | Université Lumière Lyon 2, France |

### Local Organization Chair

| | |
|---|---|
| Élise Lavoué | Université Jean Moulin Lyon 3, France |

## Program Committee

Marie-Helene Abel
Mohammad Al-Smadi
Carlos Alario-Hoyos
Liaqat Ali
Luis Anido Rifon
Inmaculada Arnedillo-Sánchez
Juan I. Asensio-Pérez
Antonio Balderas
Merja Bauters
Jason Bernard
Yolaine Bourda
Manuel Caeiro Rodríguez
Sven Charleer
Mohamed Amine Chatti
Michel Christine
Miguel Ángel Conde
John Cook
Audrey Cooke
Ulrike Cress
Alexandra Cristea
Mihai Dascalu
Paul De Bra
Carlos Delgado Kloos
Stavros Demetriadis
Christian Depover
Michael Derntl
Philippe Dessus
Darina Dicheva
Stefan Dietze
Yannis Dimitriadis
Vania Dimitrova
Juan Manuel Dodero
Peter Dolog
Hendrik Drachsler
Benedict Du Boulay
Martin Ebner
Iria Estévez-Ayres
Baltasar Fernandez-Manjon
Carmen Fernández-Panadero
Christine Ferraris
Angela Fessl
Beatriz Florian-Gaviria
Serge Garlatti
Muriel Garreta Domingo
Dragan Gasevic
Sebastien George
Denis Gillet
Fabrizio Giorgini
Carlo Giovannella
Christian Glahn
Sabine Graf
Monique Grandbastien
Andrina Granić
David Griffiths
Begona Gros
Franka Grünewald
Joerg Haake
Andreas Harrer
Ángel Hernández-García
Davinia Hernández-Leo
Sharon Hsiao
Seiji Isotani
Isa Jahnke
Marco Kalz
Nikos Karacapilidis
Michael Kickmeier-Rust
Andrea Kienle
Barbara Kieslinger
Ralf Klamma
Styliani Kleanthous
Joris Klerkx
Tomaž Klobučar
Johannes Konert
Vitomir Kovanovic
Milos Kravcik
Mart Laanpere
Lydia Lau
Élise Lavoué
Effie Lai-Chong Law
Derick Leony
Elisabeth Lex
Tobias Ley
Andreas Lingnau
Martin Llamas-Nistal
Ulrike Lucke
Rose Luckin

Vanda Luengo
George Magoulas
Katherine Maillet
Nils Malzahn
Jean-Charles Marty
M. Antonia Martínez-Carreras
Manolis Mavrikis
Martin Memmel
Agathe Merceron
Riichiro Mizoguchi
Yishay Mor
Pedro J. Muñoz Merino
Rob Nadolski
Carmen L. Padrón-Nápoles
Viktoria Pammer-Schindler
Pantelis Papadopoulos
Abelardo Pardo
Kai Pata
Mar Perez-Sanagustin
Zinayida Petrushyna
Niels Pinkwart
Kaska Porayska-Pomsta
Michael Prilla
Christoph Rensing
Bart Rienties
Adolfo Ruiz Calleja
Demetrios Sampson
Olga C. Santos
Andreas Schmidt
Ulrik Schroeder
Karim Sehaba
Stylianos Sergis
Mike Sharples
Bernd Simon
Peter Sloep
Sergey Sosnovsky
Marcus Specht
John Stamper
Slavi Stoyanov
Kairit Tammets
Stefano Tardini
Pierre Tchounikine
Stefaan Ternier
Vladimir Tomberg
Christoph Trattner
Stefan Trausan-Matu
Katrien Verbert
Jo Wake
Fridolin Wild
Raphael Zender

## Additional Reviewers

Vasiliki Aidinopoulou
Muhammad Anwar
Cecilia Avila Garzon
Oliver Blunk
Mutlu Cukurova
Jort de Vreeze
Nour El Mawas
Danny Flemming
Christoph Greven
Julio Guerra
Wayne Holmes
Christopher Krauss
Amna Liaqat
Daniyal Liaqat
Tobias Moebert
Jelena Nakic
Manuel Palomo-Duarte
Patricia Santos
Hendrik Thüs
Ina von der Beck
Matthias Weise

# Contents

## Full Papers

**Short Papers**

**Demo Papers**

**Poster Papers**

# Full Papers

# A Semantic-Driven Model for Ranking Digital Learning Objects Based on Diversity in the User Comments

Entisar Abolkasim(✉), Lydia Lau, and Vania Dimitrova

University of Leeds, Leeds LS2 9JT, UK
{scl0ena, L.M.S.Lau, V.G.Dimitrova}@Leeds.ac.uk

**Abstract.** This paper presents a computational model for measuring diversity in terms of variety, balance and disparity. This model is informed by the Stirling's framework for understanding diversity from social science and underpinned by semantic techniques from computer science. A case study in learning is used to illustrate the application of the model. It is driven by the desire to broaden learners' perspectives in an increasingly diverse and inclusive society. For example, interpreting body language in a job interview may be influenced by the different background of observers. With the explosion of digital objects on social platforms, selecting the appropriate ones for learning can be challenging and time consuming. The case study uses over 2000 annotated comments from 51 YouTube videos on job interviews. Diversity indicators are produced based on the comments for each video, which in turn facilitate the ranking of the videos according to the degree of diversity in the comments for the selected domain.

**Keywords:** Diversity model for learning · Semantics · User comments analytics · Video rating

## 1 Introduction

Videos are considered one of the main resources for learning. For instance, YouTube was ranked the second most popular social resource that has been used for informal learning by students [1]. One of the challenges that faces the learners and tutors is the tremendous amount of videos available in social environments (e.g. 300 h of video are uploaded to YouTube every minute[1]). Finding the right videos can be time consuming, especially if the learner is seeking knowledge in ill-defined domains such as culture or body language.

Social interactions around videos (e.g. user's textual comments, likes, dislikes, etc.) offer a rich source of information about the video itself, the users, and the subject domain. These interactions can provide access to diverse perspectives on the subject domain and users can learn from each other vicariously.

In "The Wisdom of Crowds", Surowiecki argues that one of the elements to have a wise crowd is to have a diverse crowd [2]. A diverse crowd could provide different

[1] http://www.statisticbrain.com/youtube-statistics/.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 3–15, 2016.
DOI: 10.1007/978-3-319-45153-4_1

perspectives or expertise by users from different backgrounds. This research aims to analyse the social cloud (e.g. YouTube videos with associated user comments, user profiles and other metadata) for the identification and ranking of suitable videos. Combining social computing and semantic techniques, this paper attempts to answer the following research questions:

**Q1** What metrics can be used to measure diversity in user comments?

**Q2** How to rank videos based on diversity in user comments?

The rest of the paper is structured as follows: Sect. 2 positions this research in related techniques used to analyse user comments and introduces a diversity framework that informed the development of the model for this paper. Section 3 introduces the proposed semantic-driven diversity model and the steps to operationalise the model. Implementation of the model as the **S**emantic-**D**riven **D**iversity **A**nalytic **T**ool (SeDDAT) is presented in Sect. 4. Section 5 shows the results from the application of SeDDAT in a study with YouTube videos. Section 6 concludes and presents future directions.

## 2 Related Work

**Techniques for Classification and Ranking of Videos.** Data mining techniques have been used to exploit the richness of user interactions around videos, especially user comments, for various purposes. For example, a mechanism for filtering comments was proposed by Serbanoiu & Rebedea to identify relevant comments on YouTube videos using classifications and ranking approaches [3]. Similarly, using classification techniques a study by Siersdorfer et al. shows that community feedback on already rated comments can help to filter and predict ratings for possibly useful and unrated comments [4]. Using the state-of-the-art in learning to rank approaches, the user interactions or "social features" were shown to be a promising approach for improving the video retrieval performance in the work introduced by [5]. For improving video categorisation, a text-based approach was conducted to assign relevant categories to videos, where the users' comments among all the other features gave significant results for predicting video categorisation [6]. Underpinned by data mining techniques, Ammari et al. used user comments on YouTube videos to derive group profiles to facilitate the design of learning simulated environments [7]. Galli et al. conducted a study that used different data mining techniques to analyse user comments to introduce a re-ranking method which produced a new ordered list of videos that is originally provided by the YouTube recommender [8].

**Semantics Techniques for Diversity Modelling.** Semantics offers a great potential for diversity modelling by providing an explicit structure to position the model within the domain of interest. A new research stream in exploration of diversity of individual's views in social media platform has emerged. A formal framework has been developed for extracting individual viewpoints from semantic tags associated with user comments [9]. Research has shown that linked data can be a useful source for enriching user modelling interactions when bringing new user dimensions, such as cultural variations

[10]. New work has also emerged on the interpretation and analysis of social web data with a strong focus on cultural differences - for example, a comparison between Twitter and Sina Weibo [11]. Likewise, recent work has also shown how data analytics can benefit the workforce engagement in enterprise contexts [12].

**Framework for Understanding Diversity.** An extensive study by Andy Stirling on measures for diversity shows how diversity has gained interest in different disciplines such as ecology, economics and policy [13]. His study shows that diversity has been measured based on three different dimensions, using Stilring's terminology, variety, balance and disparity. These dimensions have been used in three different ways to indicate the level of diversity: one concept diversity (e.g. variety only as in ecology); or dual concept diversity by combining two dimensions (e.g. variety and balance as used in economics), or triple concept diversity as a combination of variety, balance and disparity (e.g. as an aggregated value of the three dimensions as proposed by Stirling). The Stirling framework has been used in different domains, such as cultural diversity for policy and regulation [14], cultural diversity in the cinema, television and book industries [15–17], and spread of subjects in interdisciplinary research [18]).

Informed by the Stirling diversity framework, this research uses the semantic annotations of user comments on videos to facilitate video ranking according to diversity.

## 3 A Semantic-Driven Diversity Model

The diversity dimensions, *variety, balance and disparity* are defined as follow [19, p. 709]:

- *Variety* is "the number of categories into which system elements are apportioned".
- *Balance* is "a function of the pattern of apportionment of elements across categories".
- *Disparity* is "the manner and degree in which the elements may be distinguished".

Underpinned by semantic techniques, these dimensions will be used separately and in combination as indicators to measure diversity in user comments against an ontology representing a domain of interest, which will be labelled as *domain diversity*.

### 3.1 Preliminaries

**Basic Components.** The main input of the proposed model for measuring diversity is a set of textual comments $\boldsymbol{T} = \{\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_n\}$ which have been created by users $\boldsymbol{U} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ while interacting with a set of digital objects $\boldsymbol{D} = \{\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_k\}$.

**Social Cloud Components.** Every digital object $\boldsymbol{d}$ has a set of users $\boldsymbol{U}(\boldsymbol{d}) = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{m_d}\}$ who commented on $\boldsymbol{d}$, where every user $\boldsymbol{u}_i \in \boldsymbol{U}(\boldsymbol{d})$ has written at least one comment on $\boldsymbol{d}$.

Every comment $t \in T$ is associated with a user $u_t \in U$ and a digital object $d_t \in D$ where $u_t$ has made $t$ while interacting with $d_t$ in a social space. The textual comments created by a user $u \in U$ are denoted with $T(u) = \{t_1, t_2, \ldots, t_{n_u}\}$; it is assumed that $T(u) \neq \emptyset$. Similarly, the textual comments associated with a digital object $d \in D$ are denoted with $T(d) = \{t_1, t_2, \ldots, t_{n_d}\}$

It is assumed that some data are available to characterise the digital objects and the users. A digital object $d \in D$ can have some metadata that represents key features, e.g. title, author, media type (e.g. video, text, and image), and date. These metadata are presented as a vector $metadata(d) = \langle f_1, f_2, \ldots, f_{n_d} \rangle$. Similarly, it is assumed that for every user $u \in U$ some profile data is collected, e.g. user age, gender, nationality, expertise. This is captured in a user profile vector $userProfile(u) = \langle p_1, p_2, \ldots, p_{n_u} \rangle$.

**Semantic Underpinning.** As the starting point for the semantic-driven analytics pipeline, the textual comments would be semantically annotated using an ontology $\Omega$ representing the domain of interest. The set of annotated comments will be used for the diversity analysis.

***Domain Ontology.*** The ontology $\Omega$ is structured as $\Omega = \ <E_\Omega, H_\Omega>$ , where:$E_\Omega$ is a set of ontology entities $E_\Omega = C_\Omega \cup I_\Omega$, where $\mathrm{C}_\Omega$ is a set of classes that represent the domain categories, $\mathrm{I}_\Omega$ is a set of instances representing the individuals which belonging to the classes, and $C_\Omega \cap I_\Omega = \emptyset$.

$\mathrm{H}_\Omega$ is a set of hierarchical relationships between entities $\mathrm{H} = \{\mathrm{subClassOf}, \mathrm{instanceOf}\}$, where $subClassOf(e_i, e_j), e_i, e_j \in C_\Omega, e_i \neq e_j$ defines that $e_i$ is a subclass of $e_j$; and $instanceOf(e_i, e_j), e_i \in I_\Omega, e_j \in C_\Omega$ defines that $e_i$ is an instance of class $e_j$.

***Semantic Annotation.*** Every comment $t \in T$ is tagged with a set of entities $E_t = \{e_1, e_2, \ldots, e_{n_t}\}$, where $E_t \subseteq E_\Omega$. The set of ontology entities associated with all comments in $T = \{t_1, t_2, \ldots, t_n\}$ is denoted as $E = \bigcup_{i=1..n} E_{t_i}$.

### 3.2 Metrics for Domain Diversity

Measuring diversity requires the identification of the system elements and categories of the system elements [19]. For this paper, the system elements are $E$ - the entities used in annotating the user comments. The categories in which system elements can be apportioned are $C_\Omega$ - domain ontology classes. Therefore, the diversity dimensions - *variety, balance and disparity* of *domain diversity* of the digital objects, are defined as follows:

***Variety v.*** The number of ontology super classes (i.e. domain categories) into which the entities from annotation (i.e. system elements) are apportioned.

$$E_c = \{\forall e \sqsubset \mathrm{c} | c \in C_\Omega \wedge E_c \subseteq E$$

$$K = \{\forall c || E_c| > 0\}$$

$$\boldsymbol{v} = |K| \tag{1}$$

***Balance b.*** The proportions $\boldsymbol{p_i}$ of entities from annotation across the ontology super classes that are identified for variety $\boldsymbol{K}$. Shannon Entropy index is used for this research. An alternative, Shannon Evenness, is not used as it will give infinity results when variety is equal to one.

$$\boldsymbol{b} = \sum\nolimits_{i=1}^{v} p_i \ln p_i, \text{where} \tag{2}$$

$$p_i = \frac{|E_c|}{|c|}$$

***Disparity d.*** The manner and degree in which the entities from annotations may be distinguished. This investigates how scattered/dispersed the entities from annotations are within their super classes, which could be referred to as *disparity within categories*. An internal validation index Ball-Hall [20], based on clustering, is adapted to measure the dispersion *dis(c)* within each super class where a semantic distance measure (shortest path [21]) is used to calculate the distances between entities for each super class.

$$\boldsymbol{d} = \frac{1}{v} \sum\nolimits_{i=1}^{v} dis(c_i), \text{where,} \tag{3}$$

$$dis(c) = \frac{1}{|E_c|} \sum\nolimits_{j=1}^{|E_c|} \left( \min_{\forall \boldsymbol{p}} (path_p(e_j, m_c)) \right)^2,$$

and $m_c$ is the medoid[2] of category $c$.

## 4 An Overview of SeDDAT- Semantic-Driven Diversity Analytics Tool

***Implementation:*** The semantically-driven model is operationalised using Java, Jena APIs and SPARQL queries resulting in the semantic-driven diversity analytics tool, SeDDAT. It is depicted on the right hand side of Fig. 1.

***Input:*** SeDDAT takes as an input the annotated user comments, ontology that represents the domain of interest and used for annotating the comments, user profile and

[2] A medoid is the most centrally located item in a cluster that has minimal average distances to all the other items in the cluster [22].

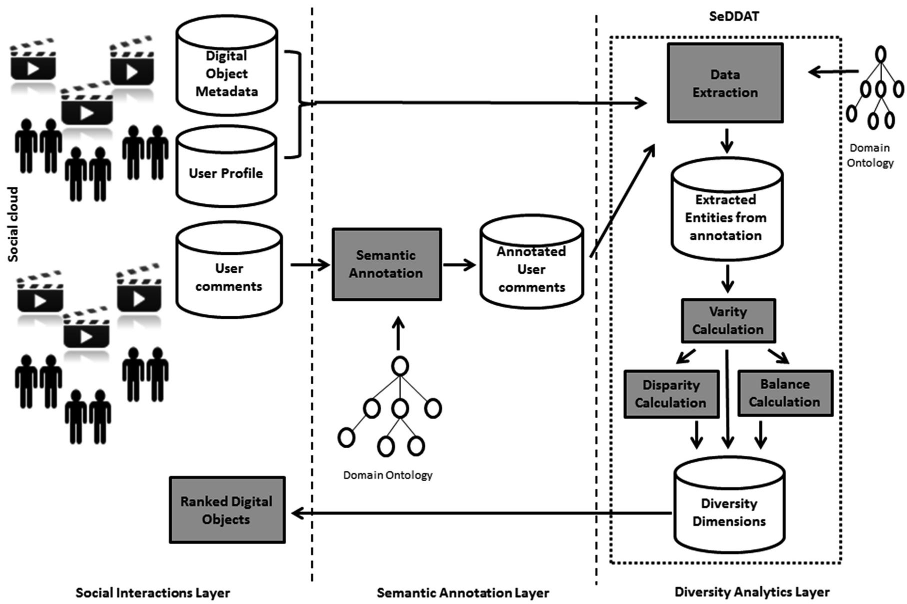



**Fig. 1.** The process of producing ranked digital objects according to diversity in user comments.

digital object metadata. To calculate domain diversity, SeDDAT retrieves the entities from an xml file and then uses the extracted entities for further calculations.

***Output:*** Given the domain ontology, the algorithms of this tool calculate a vector of the three diversity dimensions (variety, balance and disparity) for each digital object.

Figure 1 shows how SeDDAT is used for measuring the diversity in user comments. The process goes through three layers: the social interactions layer, where the social cloud (user comments, user profile, and digital object metadata) is collected; the semantic layer, where a selected domain ontology is used to annotate the user comments; and the diversity analytics layer, where SeDDAT extracts the entities used in the annotations of the user comments, calculates the diversity of these entities that are mapped against the domain ontology, and ranks the digital objects according to the selected metrics.

## 5 A Case Study- Application of SeDDAT on Video Ranking

In order to test the proposed diversity model, SeDDAT was used on a set of videos about job interviews. Apart from the verbal communications, body language is one of the aspects that may influence the outcome of the interaction between the interviewer and interviewee. In an increasingly inclusive and diverse society, it is important to understand the different possible interpretations of the body language signals to avoid misunderstanding. This study aimed to test the usefulness of the diversity metrics in the selection of videos that contain the most diverse range of comments relating to body language in job interview. There is an assumption that the higher the diversity, the higher the potential of a video for broadening and deepening the learners' awareness.

### 5.1 Input Dataset

The input dataset was an xml file, obtained from a previous study by Despotakis [23]. It contains (**a**) videos metadata: video ID, URIs of the YouTube videos on job interviews with associated title, category, author, duration, (**b**) user profiles: nickname, age, gender, location, and occupation, and (**c**) annotated comments: comments with associated ontology entities and their URIs. A body language ontology was used to semantically annotate the comments (an automated process).

The assumption for SeDDAT is that the ontology and the semantic annotations of the comments are sound. Only a subset of the data was used for this study:

– 51 videos were randomly selected from over 200 videos.
– 2949 associated comments were extracted.
– 1223 unique entities from annotations were extracted.

### 5.2 The Domain Ontology

Body language[3] ontology, which was used to semantically annotate the comments and assist the process of calculating the diversity dimensions, has *eight* domain categories (top super classes): body motion; body position; body language; body language signal meaning; body sense function; object; kinesics; and nonverbal communication (see Fig. 2).

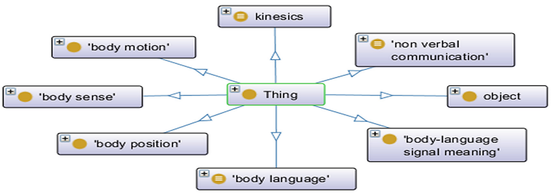



**Fig. 2.** A protégé snapshot of the domain categories (top super classes) of the selected domain ontology.

### 5.3 Results

The extracted entities from annotations were passed through the three algorithms designed to calculate the diversity dimensions as shown in Fig. 1. The results (data associated with each video as well as diversity dimensions) were saved in a spreadsheet

[3] http://imash.leeds.ac.uk/ontology/amon/BodyLanguage.owl

**Table 1.** Sample results of seven YouTube videos sorted by video IDs (smallest to largest).

| Video ID | #Comments | #Entities | Variety | Balance | Disparity |
|---|---|---|---|---|---|
| 103 | 25 | 6 | 2 | 0.32 | 39.4 |
| 190 | 5 | 2 | 1 | 0.01 | 60.5 |
| 209 | 74 | 48 | 4 | 0.68 | 20.08 |
| 363 | 4 | 16 | 3 | 0.39 | 25.28 |
| 402 | 425 | 105 | 6 | 1.14 | 10.65 |
| 403 | 293 | 68 | 4 | 0.85 | 14.83 |
| 788 | 45 | 35 | 5 | 0.75 | 15.95 |

for further analysis. Table 1 shows the diversity dimensions of a sample of seven YouTube videos with some of the associated data: video ID and number of comments and entities from annotations.

### 5.4 Analyses and Discussion

A combination of quantitative and qualitative analysis of the results was conducted to acquire a deeper understanding on the nature of diversity highlighted. Inspired by Rafols et al. [18], this study used more than one indicator for diversity in user comments. Each diversity dimension was used separately to rank the videos and then in combination. Answers to the following questions were sought:

*Q1 What does it mean to be ranked top or bottom based on variety?*
*Q2 What does it mean to be ranked top or bottom based on balance?*
*Q3 What does it mean to be ranked top or bottom based on disparity?*
*Q4 What if the three diversity dimensions are used in combination for ranking?*

**(1) Ranking Based on Variety.** Videos with high variety indicate that the comments have covered most or all of the high level aspects of the domain (i.e. the entities from annotations are apportioned to different domain categories). Therefore, to identify videos that covered a variety of domain aspects, the video ordering can be based on the highest to smallest values for variety. As can be seen in Table 2, comments on the top video 402 covered six domain categories (body sense function; body position; object; body language; body motion; and body language signal meaning) compared to the bottom ranking video 190 that had comments covering only one domain category (body language signal meaning).

**(2) Ranking Based on Balance.** Videos with a high value in balance mean that comments covered evenly the aspects of the domain (i.e. the entities from annotations are well proportioned across domain categories). See Table 3 for the list of videos sorted based on balance. The video 402 was ranked top, because the proportions $p_i$ of its entities are higher compared to the other videos. Table 4 shows the proportions, as defined in formula 2 in Sect. 3.2, of the two top videos 402 and 403. For example, body language signal meaning has a total of 1336 entities (classes and instances), and the proportions of videos 402 and 403 are 52 and 40 respectively.

**Table 2.** The sample videos ordered top to bottom according to variety.

| Video ID | #Comments | #Entities | Variety | Balance | Disparity |
|---|---|---|---|---|---|
| 402 | 425 | 105 | 6 | 1.14 | 10.65 |
| 788 | 45 | 35 | 5 | 0.75 | 15.95 |
| 209 | 74 | 48 | 4 | 0.68 | 20.08 |
| 403 | 293 | 68 | 4 | 0.85 | 14.83 |
| 363 | 4 | 16 | 3 | 0.39 | 25.28 |
| 103 | 25 | 6 | 2 | 0.32 | 39.4 |
| 190 | 5 | 2 | 1 | 0.01 | 60.5 |

**Table 3.** The sample videos ordered top to bottom according to balance.

| Video ID | #Comments | #Entities | Variety | Balance | Disparity |
|---|---|---|---|---|---|
| 402 | 425 | 105 | 6 | 1.14 | 10.65 |
| 403 | 293 | 68 | 4 | 0.85 | 14.83 |
| 788 | 45 | 35 | 5 | 0.75 | 15.95 |
| 209 | 74 | 48 | 4 | 0.68 | 20.08 |
| 363 | 4 | 16 | 3 | 0.39 | 25.28 |
| 103 | 25 | 6 | 2 | 0.32 | 39.4 |
| 190 | 5 | 2 | 1 | 0.01 | 60.5 |

**Table 4.** The proportions of videos 402 and 403 across the eight domain categories.

| Video ID | Body language signal meaning (1336) | Body position (33) | Body motion (118) | Body language (429) | Object (256) | Nonverbal communication (1) | Kinesics (1) | Body senses function (6) |
|---|---|---|---|---|---|---|---|---|
| 402 | 52 | 1 | 13 | 4 | 32 | 0 | 0 | 4 |
| 403 | 40 | 0 | 9 | 0 | 17 | 0 | 0 | 2 |

**(3) Ranking Based on Disparity.** Videos with high disparity indicate that the comments cover distinctive aspects within the domain categories (i.e. the entities from annotating the comments are widely scattered within their domain categories). Therefore, to identify videos that triggered distinct domain aspects around their content, the videos can be order largest to smallest according to their disparity value as can be seen in Table 5.

Ranking based on disparity shifted the top videos (e.g. videos 402 and 403) that were ranked based on variety or balance to the bottom. Similarly, the video 190 that was ranked bottom for variety and balance came top here.

To investigate this further, the ranked videos were inspected closely using the **(a)** video content, **(b)** number of comments, **(c)** number of entities from annotations, and **(d)** samples of user comments. Also, a correlation between the number of user comments and the diversity dimensions was conducted.

**Table 5.** The sample videos ordered top to bottom according to disparity.

| Video ID | #Comments | #Entities | Variety | Balance | Disparity |
|---|---|---|---|---|---|
| 190 | 5 | 2 | 1 | 0.01 | 60.5 |
| 103 | 25 | 6 | 2 | 0.32 | 39.4 |
| 363 | 4 | 16 | 3 | 0.39 | 25.28 |
| 209 | 74 | 48 | 4 | 0.68 | 20.08 |
| 788 | 45 | 35 | 5 | 0.75 | 15.95 |
| 403 | 293 | 68 | 4 | 0.85 | 14.83 |
| 402 | 425 | 105 | 6 | 1.14 | 10.65 |

As can be seen in Table 6, the number of comments correlates significantly with the diversity dimensions. The comments correlate positively with variety and balance and negatively with disparity. For example, video 402, which had the highest number of comments (i.e. 425), presents seemingly the appearance (dress code and makeup) appropriate for working in a certain company, but the comments covered most of the domain aspects related to body language (highest variety), and more evenly compared to other videos (highest balance). On closer inspection, the majority of the comments converged around ‘racial’ theme triggered by watching the video or by discussing the company’s policy, which might be the cause of the low disparity value.

**Table 6.** The correlation between the number of comments and diversity dimensions.

**Correlations**

| | | | #Comments | Variety | Balance | Disparity |
|---|---|---|---|---|---|---|
| Spearman's rho | #Comments | Correlation Coefficient | 1.000 | .656** | .648** | -.185 |
| | | Sig. (2-tailed) | . | .000 | .000 | .199 |
| | | N | 50 | 50 | 50 | 50 |
| | Variety | Correlation Coefficient | .656** | 1.000 | .876** | -.566** |
| | | Sig. (2-tailed) | .000 | . | .000 | .000 |
| | | N | 50 | 50 | 50 | 50 |
| | Balance | Correlation Coefficient | .648** | .876** | 1.000 | -.546** |
| | | Sig. (2-tailed) | .000 | .000 | . | .000 |
| | | N | 50 | 50 | 50 | 50 |
| | Disparity | Correlation Coefficient | -.185 | -.566** | -.546** | 1.000 |
| | | Sig. (2-tailed) | .199 | .000 | .000 | . |
| | | N | 50 | 50 | 50 | 50 |

**. Correlation is significant at the 0.01 level (2-tailed).

A high number of domain-related comments is likely to result in a high number of entities from annotations, but what is important is that the entities from annotating the comments must be: apportioned to many domain categories to be ranked high based on variety, or well proportioned across the domain categories to be ranked high based on balance, or widely dispersed within the domain categories to be ranked high based on disparity.

A visual inspection of the coverage of domain categories by entities was conducted. Figure 3 shows two snapshots of the dispersion of the entities from annotations of videos 190 and 402 within the domain category body language signal meaning.

The snapshots are obtained using the framework ViewS[4] implemented by Despotakis [23]. As can be seen in Fig. 3 on the left side, the two entities of video 190 are widely scattered within the domain category (i.e. the semantic distance between the entities is high). On the other hand, the entities on video 402 are closely distributed within the domain category (i.e. the semantic distance is low).

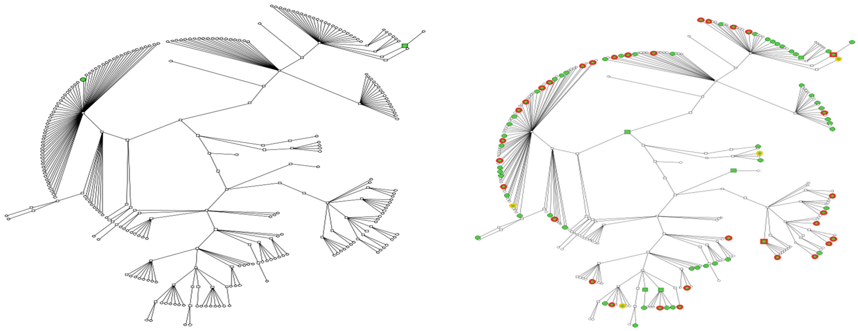


**Fig. 3.** The dispersion of the entities within the domain category body language signal meaning for the videos 190 and 402.

**(4) Ranking Based on a Combination of Diversity Dimensions.** One way of ranking based on the combined diversity dimensions is to rank based on variety first, then balance and then disparity (e.g. largest to smallest). This was raised by the question *"How to differentiate videos with the same variety index?"* such as, videos 403 and 209 in Table 5.

## 6 Conclusion and Future Work

Combining social computing and semantic annotation techniques, this paper presented a novel mechanism to rank videos based on the diversity in user comments of these videos. The proposed ranking tool harvests and utilises the richness of the social cloud, specifically the comments, to benefit tutors and learners by identifying the videos that have the potential to diversify the learner's perspectives.

In the future, this research will extend to the other components of the social cloud, such as user profiles and videos' metadata, to (**a**) better understand the diversity of the learners and the users who commented on the videos, and (**b**) enhance the ranking and recommendation. For example, the user profile can help to understand the user/commenter diversity, which in turn can be used with the user's own comments on videos that he/she has previously watched to nudge him/her to videos that diversify the current knowledge.

[4] A graph in ViewS shows the entities (classes and instances) of a domain category (super class). The colored (darker) shapes are the entities from annotating the comments on the video and the uncolored ones are the entities not present in the user comments.

Moreover, the effectiveness of using the Stirling diversity index [19], calculated by aggregating the three diversity dimensions (variety, balance, and disparity), will be investigated, where other indexes for measuring the diversity dimensions will be explored as appropriate.

## References


1. Yakin, I., Gencel, I.E.: The utilization of social media tools for informal learning activities: a survey study. Mevlana Int. J. Educ. **3**(4), 108–117 (2013)
2. Surowiecki, J.: The Wisdom of Crowds. Random House (2004)
3. Serbanoiu, A., Rebedea, T.: Relevance-based ranking of video comments on YouTube. In: Proceedings of 19th International Conference on Control Systems and Computer Science, CSCS 2013, pp. 225–231 (2013)
4. Siersdorfer, S., Chelaru, S., Nejdl, W., San Pedro, J.: How useful are your comments? - analyzing and predicting YouTube comments and comment ratings. In: Proceedings of the 19th International Conference on World Wide Web, vol. 15, pp. 891–900 (2010)
5. Chelaru, S., Orellana-Rodriguez, C., Altingovde, I.S.: How useful is social feedback for learning to rank YouTube videos? World Wide Web **17**(5), 997–1025 (2014)
6. Filippova, K., Hall, K. B.: Improved video categorization from text metadata and user comments. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information – SIGIR 2011, p. 835 (2011)
7. Ammari, A., Lau, L., Dimitrova, V.: Deriving group profiles from social media to facilitate the design of simulated environments for learning. In: Proceedings 2nd International Conference Learning Analytics Knowledge – LAK 2012, p. 198 (2012)
8. Galli, M., Gurini, D. F., Gasparetti, F., Micarelli, A., Sansonetti, G.: Analysis of user-generated content for improving YouTube video recommendation. In: CEUR Workshop Proceedings, vol. 1441 (2015)
9. Despotakis, D., Dimitrova, V., Lau, L., Thakker, D.: Semantic aggregation and zooming of user viewpoints in social media content. In: Carberry, S., Weibelzahl, S., Micarelli, A., Semeraro, G. (eds.) UMAP 2013. LNCS, vol. 7899, pp. 51–63. Springer, Heidelberg (2013)
10. Denaux, R., Dimitrova, V., Lau, L., Brna, P., Thakker, D., Steiner, C.: Employing linked data and dialogue for modelling cultural awareness of a user. In: Proceedings of the 19th International Conference on Intelligent User Interfaces, IUI 2014, pp. 241–246 (2014)
11. Gao, Qi, Abel, F., Houben, G.-J., Yu, Y.: A comparative study of users' microblogging behavior on Sina Weibo and Twitter. In: Masthoff, J., Mobasher, B., Desmarais, M.C., Nkambou, R. (eds.) UMAP 2012. LNCS, vol. 7379, pp. 88–101. Springer, Heidelberg (2012)
12. Bozzon, A., Efstathiades, H., Houben, G.-J., Sips, R.-J.: A study of the online profile of enterprise users in professional social networks. In: WWW 2014 Companion Proceedings of the 23rd International Conference on World Wide Web, pp. 487–492 (2014)
13. Stirling, A.: On the economics and analysis of diversity, 28 (1998)
14. UNESCO Institute for Statistics (UIS), Measuring The Diversity Of Cultural Expressions: Applying the Stitling Model of Diversity in Culture,vol. 6 (2011)
15. Farchy, J., Ranaivoson, H.: Do public television channels provide more diversity than private ones. J. Cult. Manag. Policy **1**, 50–63 (2011)
16. Benhamou, F., Peltier, S.: Application of the stirling model to assess diversity using UIS cinema data. UNESCO Inst. Stat., pp. 1–73 (2010)

17. Benhamou, F., Peltier, S.: How should cultural diversity be measured? an application using the French publishing industry. J. Cult. Econ. **31**, 85–107 (2007)
18. Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., Stirling, A.: How journal rankings can suppress interdisciplinary research: a comparison between Innovation Studies and business & management. Res. Policy **41**(7), 1262–1282 (2012)
19. Stirling, A.: A general framework for analysing diversity in science, technology and society. J. R. Soc. Interface **4**(February), 707–719 (2007)
20. Despotakis, D.: Modelling viewpoints in user generated content (2013)

# Social Facilitation Due to Online Inter-classrooms Tournaments

Roberto Araya(✉), Carlos Aguirre, Manuel Bahamondez, Patricio Calfucura, and Paulina Jaure

Centro de Investigación Avanzada en Educación, Universidad de Chile, Santiago, Chile
roberto.araya.schulz@gmail.com

**Abstract.** In this paper we explore the impact of an inter-classrooms math tournament implemented through internet. The strategy is to increase learning through intra-classroom collaboration generated by inter-classroom competition. Ten fourth grade classes with all their students from eight schools participated. During previous weeks students practiced on-line and played a cloud based board game designed to learn word problems. Afterwards, all students participated on an inter-classroom tournament. They played on-line synchronously during 60 min. The game was played in dyads formed from different schools. The list of each classroom average score was published every 5 min on each student computer. We found an important social facilitation effect: a significant improvement on the performance of male students weak on math, and therefore a reduction on the performance gap between mathematically weak and strong male students. The improvement of female students weak on math was also significant but lower.

**Keywords:** Game based learning · Learner affect · Motivation and engagement · ICT inclusion for learning

## 1 Introduction

There is ample evidence that schools have not changed dramatically over the last few centuries [4, 15]. Even after the introduction of textbooks, students continue to spend their class time by primarily listening to lectures and taking notes. Why does education seem so immune to transformations? Labaree [15] argues that education is a far more complex domain than other areas. For example, he compares a typical nuclear power facility with a school. Since every component of a nuclear facility is causally interrelated with the others, it is much easier to trace the source of any deficiencies and fix them accordingly. Schools, conversely, are composed of completely independent units: isolated classrooms. If one classroom performs well, it does not immediately produce an effect on parallel classrooms. Superintendents and principals generally track mean performance across classrooms and, on average, good and bad performances cancel each other out. Therefore, on the whole, the school remains highly stable. In this paper, we provide some empirical evidence to suggest that this situation can be radically transformed by information technology, game-based learning, and, in particular, by online inter-classrooms tournaments.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 16–29, 2016.
DOI: 10.1007/978-3-319-45153-4_2

On the other hand, from a psychological point of view, learning requires several cognitive resources: working memory, long-term memory, attention, unconscious and conscious mechanisms, representation mechanisms and metacognitive processes [20]. When starting with a problem, perceptual pattern detection and nature of problem recognition processes are activated. A strategy is then selected from long-term memory. This selection depends on the familiarity with the problem. If the problem is unfamiliar, basic procedures are explored and used. After developing fluency through extensive practice, attentional resources are then freed up. Strategy discovery processes then become activated, leading to the combination of old strategies and the construction of new ones. Therefore, learning requires practice. With each new level of complexity, practice is required to ensure proficiency and free up attentional resources in order to start a new cycle of discovery and further learning.

However, practice requires strong motivation. Therefore, the main challenge facing the teacher is motivating their students. In order to do so, teachers need effective tools with which to connect with their students and engage them in learning activities. Play is a natural tool and is ideal for repetitive practice [5]. While playing, students are constantly practicing. Social play is even more engaging than playing alone, and it is arguably even more natural. It is an ecologically-valid educational strategy used by mammals and several other animals [17]. In a classic study from 1898 [23], Triplett found that cyclists were faster when competing against others than when racing alone. This effect is called social facilitation and has subsequently been found in other tasks and other animals [26, 27]. Brains have evolved for action [11], but actions with others are more engaging. The hunter-gatherer brain is particularly well-adapted to collaborating and learning from others in order to compete with neighboring groups. Intergroup competition may therefore be older than our species' heavy reliance on cultural evolution [13]. Tribal warfare is a chronic occurrence [7] and by no means the exception. However, cooperation is a very powerful weapon for competition. It evolved not due to benevolence, but because it provides an advantage when it comes to survival [12]. "Us against Them" situations generate a strong motivation to learn together, compare strategies, help each other, improve and keep trying. In these situations, learning is a pressing matter and an urgent need, as well as being more meaningful. The brain immediately perceives the benefit of practicing, and the benefit is not decades away in the future. This is an important emotional effect that can boost performance. These findings from anthropology and evolutionary psychology suggest that there is a big opportunity for team games in education. Team games capture these biologically primary motives and could therefore be used to increase motivation and learning of academic contents.

Social play is hardwired for learning, but it is better suited to acquiring biologically primary skills [9], such as hunting, fighting and mating. It is not always obviously suited to academic knowledge, such as fractions or word problems. Academic contents are biologically secondary knowledge [9]. They are the product of several millennia of cultural development, and are not easily grasped. They require thousands of hours of intensive practice and guided instruction. Furthermore, when there are several children playing simultaneously, managing them and making sure they are learning is a complicated task. Even in games with very well-defined and widely-understood rules, the challenge of classroom management is far more complex than in a traditional lecture-based class.

Nonetheless, there is a long history of using team games in classrooms for academic subjects [14, 16, 21] and tournaments [21]. For example, Slavin [21] proposes Team-Games-Tournaments (TGT), in which every week students from a class compete against members of other teams from the same class. In mathematics, Edwards et al. [8] measured the effect of a non-simulation (no attempt to simulate aspects of reality), non-computer based math game played intra class by teams of 4, competing in a tournament over the course of 9 weeks. Ninety six 7$^{th}$ graders from two low ability and two average ability classes were taught equations, and met twice per week. One low ability class and one average ability class participated in the tournament, while the other two classes were control groups who were taught following traditional classroom methods. Significant interaction and improvement was obtained in the low ability class, and learning rates were more similar in the experimental classes than in the control ones. This is a game where math skills are needed for winning, and which allows for peer tutoring. During the game, the students receive immediate feedback, while each individual's score is made publicly available.

From the teacher's point of view, team games provide a unique environment for teaching. The teacher can easily form emotional connections with the students, empathize with them and be their leader. Our brains have also evolved to follow a leader in our conflict against other tribes. This opportunity is optimized when teams comprise the whole class. In this case, the students can truly trust their teacher as there is no conflict of a teacher helping rival teams. Instead, they only provide academic and emotional support to their own class. Empirical evidence shows that students learn the most in classrooms where the students feel they can trust the teacher [6]. In inter-class competitions, students can truly trust their teacher. Therefore, they should be more open to receiving instructions and feedback from the teacher. Additionally, in inter-class tournaments, students identify as members of their class. With massive online synchronous tournaments between classes, we can recreate the powerful "Us" against "Them" environment and, therefore, activate ancestral intra-group collaboration and social motivation mechanisms.

In this paper, we reveal empirical evidence from a game played by teams. Each team is formed by all of the students in a class. This is an innovation and a challenge. According to the cooperative learning literature [14, 16, 21], large teams are not efficient for academic learning. The larger the group size, the fewer the members that can participate [14]. Edwards et al. [8] suggest that when teams have more than 5 members, it does not allow the majority of the students to participate. In this paper, we explore the effect of large teams. Some classes have more than 30 students. Another difficulty is that classes are of different sizes, ranging from 20 to 40 students. Moreover, the classes are not homogenous. Instead, they comprise students of very different levels of ability. Mixed-ability classes are an extra challenge when the teacher has to make sure that all of the students have to learn.

To the best of our knowledge, games involving teams made up of a whole class are not used for academic learning. In [3] we reflect on several years of experience with massive computer-based team tournaments and in [1, 2] we look at massive online multiplayer tournaments for mathematical modeling that are held once or twice a year, with teams from hundreds of schools competing against each other. However, these were teams of 12 students selected from a class or from several classes in the same grade level at each school.

## 2 Methods

Fourteen entire 4$^{th}$ grade classes from 12 schools prepared for an inter-classroom tournament at the end of 2015. All of the schools are in low socio economic status (SES) communities. Prior to the tournament, the schools participated in three training sessions. The first training session was held during the fourth week of October. In total, 271 students practiced word problems using a non-game mode of the ConectaIdea web-based computer platform. This is a platform where the classroom teacher and a remote teacher track student performance in real time, detect which students are having difficulty, and provide just-in-time support using a chat function included in the platform. Later, during a session held in November, 282 students played the *Espiral Mágico* game within their class. Subsequently, in another session held one or two weeks later in November, 255 students again played *Espiral Mágico* within their class (Fig. 1). *Espiral Mágico* is an online board game included in the ConectaIdea platform. This game is designed to help students learn how to tackle word problems. After these 3 training sessions, the tournament was held on December 9$^{th}$, involving 217 students (87 girls and 130 boys) from 10 classes, with an average age of 9.99 years and a standard deviation (SD) of 1.90 years. The average class size was 21.7 students, with the smallest class comprising 17 students and the largest 29. During the tournament, all of the classes played against each other synchronously for 60 min. Four of the classes could not join the tournament due to scheduling difficulties. Therefore, the statistics and results that are presented below are taken from the 10 classes that participated in the tournament. During the tournament, the students played the same game that they had played in the two final training sessions, though this time it was played using the inter-classroom tournament mode. In this particular mode, the game is played by pairs of students from different schools that compete against each other, but the score for each class is the average score for all of the students in the class. Each class' score is continuously updated and displayed as a ranking every 5 min on each student's computer.

Why use a board game? According to the National Mathematics Advisory Panel [24], board games are "particularly effective in improving low-income preschoolers' numerical knowledge and reducing discrepancies in the numerical knowledge that children from low- and middle-income homes bring to school." They are engaging and effective for classroom context [18]. *Espiral Mágico* is a board game that has been designed to practice word problems using the curricular content selected by the teacher. Why word problems? As stated in the National Mathematics Advisory Panel report [24], word problems are the most challenging curricular content in elementary school mathematics, and they are an essential prerequisite for learning algebra. Furthermore, two topics from the curriculum were selected for the tournament: "properties of 0 and 1 for multiplication and divisions", and "solving equations and inequations using addition and subtraction with natural numbers up to 100". Therefore, the word problems that were presented required the use of these two curricular contents. Examples of such

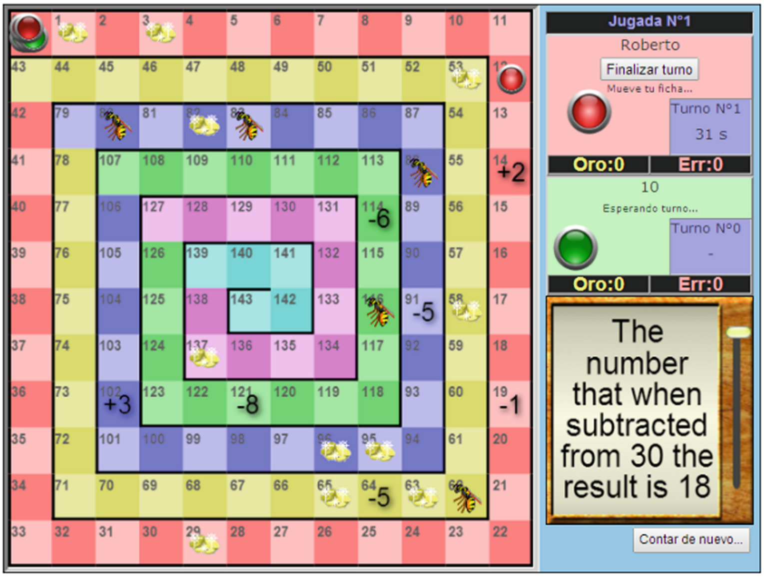



**Fig. 1.** Screen shot of the *Espiral Mágico* board game posing a word problem. It is a spiral path that starts in the upper-left corner, where the start cell is located. The path ends at the center of the board, where the goal is located. It runs clockwise. The word problem is located in the bottom right of the screen. The solution gives the number of positions that the player has to move one of her three beads. The player chooses the most convenient bead. If, after moving the bead, it ends up in a cell containing a number in large writing, then she has to move the bead forwards or backwards the corresponding number of squares. If the bead ends up in square with a monster then her bead goes to the start cell. If her bead ends up in space containing one of their rival's counters, then that rival's bead goes back to the start cell. Therefore, the player has to strategically select which bead to move. The player that reaches the goal cell first gets extra points and the play is over.

word problems include (Fig. 1): "The number that when subtracted from 30 the result is 18", "The number that when added to 18 equals 36", "the result of adding 12 and 0, divided by 10", etc.

The tournament is run by a tournament administrator. This is an independent teacher that remotely presents the game and the participating schools, cheers the teams along and constantly announces the updated ranking. He has the support of a video streaming engineer who broadcasts one way video to every class, and he also has the support of a chat manager, who answers questions from teachers and students (Fig. 2). Every class connects to a web page at a certain pre-defined time. Each class plays as a team. Every 5 min, a ranking with each class' score is published. However, the score of each individual student is also recorded and specific feedback is automatically given to each student according to his particular performance. As [14, 21] underline, individual accountability and team goals are two critical features in collaborative learning. During the tournament, each student solved an average of 11.8 word problems, with a SD of 4.3 problems.

**Fig. 2.** From left to right, (i) one engineer tracks the video streaming and supervises the students' connection to the game, one teacher manages the video streaming chat, and one teacher introduces and runs the tournament via video streaming and announces the class ranking every 5 min, (ii) one of the ten participating classes; (iii) another of the ten participating classes. The teacher that runs the tournament can be seen in the projections on the classroom walls.

In each class we define students that are weak at math as those with a grade point average (GPA) that is below their class average. The rest of the students are defined as being strong at math. From herein after, we will refer to them as the weak and strong students, respectively. The GPA is calculated based on several online tests taken by the students throughout the year using the ConectaIdeas platform. All of the classes took the same tests. The scale goes from 1 (the minimum) to 7 (a perfect score). The mean GPA in math for the students that participated in the tournament was 4.5, with a SD of 0.86. Overall, the tournament featured 113 weak students and 104 strong students. The mean GPA for the weak students was 3.93, with a SD of 0.63. On the other hand, the mean GPA for the strong students was 5.17, with a SD of 0.56. Note that there is a significant gap between these two means. As shown in Fig. 3, the strong students scored 1.24 points more than the weak students. If we measure this gap in terms of the SD of the GPA for the weak students, the result is 1.24/0.63 = 1.97 SD. Measured in terms of the SD of the mean of the GPA for the weak students, this gap is 20.5 SD. This is not only statistically significant, it is a huge gap.

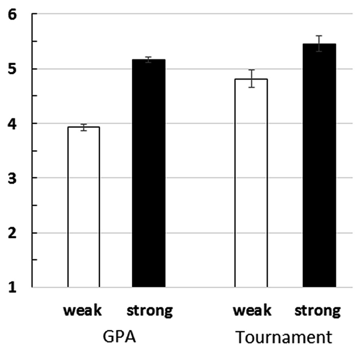



**Fig. 3.** GPA and performance during the tournament by weak and strong students, shown with confidence intervals.

# 3 Results and Discussion

The mean performance by the students during the tournament was 5.1, with a SD of 1.65. As shown in Fig. 3, the mean performance by the weak students was 4.82, with a SD of 1.74. On the other hand, the mean performance by the strong students was 5.45, with a SD of 1.47. Therefore, the mean performance by the weak students was 0.64 points lower than the mean performance by the strong students. This is 50.8 % of the aforementioned gap between their respective GPAs. Expressed in terms of the SD of the performance by the weak students, the gap is 0.64/1.74 = 0.36 SD. Although a significant gap, it is much lower than the gap of 1.97 SD between the GPAs. In fact, the gap that was witnessed during the tournament is only 18.4 % of the gap between the students' GPAs. This result means that during the tournament the gap between the strong and weak students is significantly reduced. This is an important finding. However, there can be several possible explanations.

## 3.1 Possible Explanations Based on Differences in Difficulty Level

One possible explanation for this result is that the math included in the game is easier than the math used in the online tests to calculate the GPA. If this explanation were valid, then all of the students should have done better. Since the strong students are close to the maximum score, their improvement during the tournament is less than the improvement made by the weak students, and therefore the gap is reduced. However, the mean performance by the strong students in the tournament did not differ much from the mean of their GPA. Although it is a statistically significant difference, they are still far from 7, the perfect score. In other words, they only performed slightly better. Their mean improved by only 0.28 points compared to their GPA. In terms of SD, this improvement is 19 % of the SD of the scores from the tournament and 50 % of the SD of the GPA. Moreover, 34.6 % of the strong students performed worse during the tournament than their respective GPA.

Another possible explanation is that there were two different kinds of math questions used in the *Espiral Mágico* game: very easy and very difficult. If this were the case, then there would be no medium-difficulty problems, and therefore all of the students would have performed similarly. However, the variation in student performance for the tournament was very high compared to the variation in GPAs. The SD of the performance by strong students during the tournament was 1.47. This is much higher than the SD of their GPAs, which was 0.56. Similarly, the SD of the performance by the weak students during the tournament was 1.74, which is much higher than the SD of their GPAs, which was 0.64. Therefore, this second explanation does not match the data. After ruling out these two possibilities, the most plausible explanation is that the weak students learned more in the tournament than the strong ones.

## 3.2 Possible Explanation Based on the Game-like Nature of the Activity

One possible cause of the significant improvement by the weak students is the game-like nature of the activity. To explore this hypothesis we shall analyze the data from the training sessions. As mentioned previously, three training sessions were held before the tournament. In the first session, the students used the same ConectaIdea web platform with the same type of word problems as those used in the tournament, although the platform was not set to game mode. The platform provided word problems and gave instant feedback. The teacher tracked the students' performance in real time using her tablet and provided support to those who were struggling the most. In this training session 146 students participated that later also participated on the tournament. 75 were weak students and 71 were strong students. The mean performance by the weak students was 4.61, with a SD of 1.79. The mean performance by the strong students was 5.63, with a SD of 1.42. This means that in this non-game-based activity there is a significant difference in performance between the weak and strong students; a gap of 1.02 points. This is similar to the gap between the mean GPA for the weak and strong students. Moreover, as shown in Fig. 4, there is a significant correlation between student GPA and performance in this non-game-based activity: student performance = 0.93 GPA + 0.80 with an $R^2$ of 0.20. However, the correlation between GPA and student performance during the tournament is close to zero, with an $R^2$ of 0.04. Thus the results from this training session confirm that the *Espiral Mágico* game was not easier than the normal tests and that it did not mostly contain only easy or difficult problems. They also suggest that the game-based nature of the activity makes a difference to the weak students.

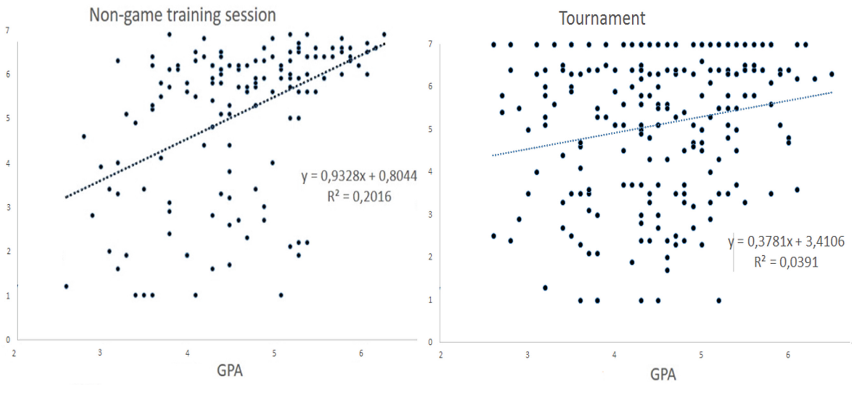



**Fig. 4.** GPA and performance during the non-game-based training session by the 146 students who participated in the training session, and GPA and performance during the tournament by the 217 students who participated in the tournament.

### 3.3 Possible Explanation Based on the Inter-classroom Nature of the Tournament

The hypothesis regarding the game-based nature of the activity includes two probable causes for the improvement made by the weak students: the game-based nature of the activity and/or the inter-classroom nature of the tournament. In order to try to disambiguate these two possible causes, we shall now analyze the next two training sessions. These were sessions where the students played the same game using the same type of word problems. In these two warm up tournaments, the students did not play against students from other classes. Instead, the students played one-on-one against another student from their class, without forming teams. 145 students participated in the second training session. The mean performance by the students during the training session was 4.76, with a SD of 1.92. This was slightly higher than the mean GPA, which was 4.49. Here, the difference is just 0.27 points, which corresponds to 14 % of the SD of the students' performance during this training session. The gap between the weak and strong students was 0.71 points. 156 students participated in the third training session. In this case, the mean performance by the students was 4.30, with a SD of 1.59, which is 0.19 points lower than the mean GPA. This difference is 12 % of the SD of the students' performance during this final training session. The gap between the weak and strong students was 1.04 points. These facts therefore suggest that it is not just the game-based element that leads to significant improvement by weak students; instead it may be the social nature of competing between classes. The "Us" against "Them" ancestral mechanism, which is more strongly activated in inter-class tournaments, appears to be the most important driver of motivation and improvement among weak students.

In order to explore this motivational mechanism we can get insights from the evolutionary psychology literature. According to Geary [10], boys tend to form larger groups, which is normal when preparing for inter-tribal conflicts. Girls instead tend to form much smaller groups, with more intense and lasting relations. Thus, boys are more easily motivated by large group collaboration in preparation for inter-group conflicts. Therefore, a prediction from evolutionary psychology is that if the inter-classrooms nature of the tournament is indeed the mechanism that boosts performance among weak students, then there should be a gender difference in the improvement made by weak students. Since the tournament provided us with information on weak and strong male and female students, then we can confirm or refute this prediction.

There were 89 girls that participated in the tournament, 42 with a weak GPA and 47 with a strong GPA. As shown in Fig. 5, the mean GPA among the weak female students was 4.00, with a SD of 0.56, while the mean GPA among the strong female students was 5.18, with a SD of 0.60. In other words, there was a gap of 1.18 points between the two groups. This gap represents 2.1 SD of the GPA for the weak female students. On the other hand, the mean performance during the tournament by the weak female students was 4.37, with a SD of 1.97. The mean performance during the tournament by the strong female students was 5.53, with a SD of 1.42. This performance is slightly higher than the mean performance by the strong male students, which was 5.33, with a SD of 1.52. The gap between the weak and strong female students during the tournament was 1.16 points. This is very similar to the gap in their GPA. However, it now represents only 0.59 SD of the GPA for the weak female students. This means that the gap during the tournament is 28 % of the gap in GPA. In the case of the boys, the weak male students had a mean GPA of 3.89, with a SD of 0.68, and the strong male students had a mean GPA of 5.15, with a SD of 0.52. This means that the gap is 1.26 points, which is 1.85 SD of the GPA for the weak male students. During the tournament, the mean performance by the weak male students was 5.08, with a SD

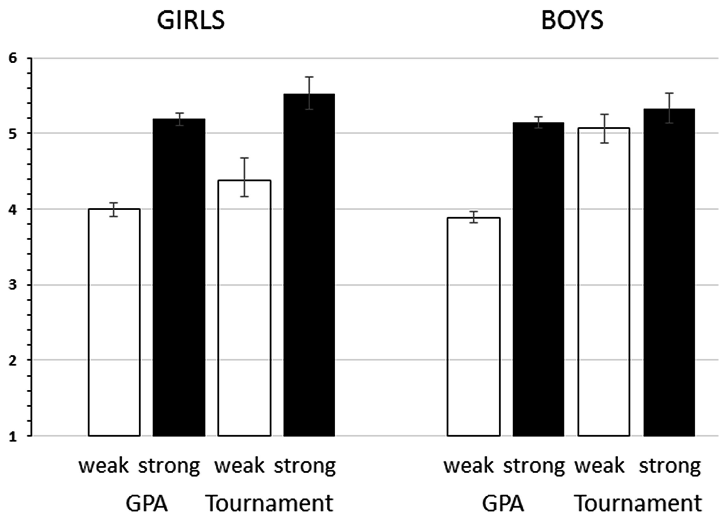



**Fig. 5.** GPA and performance during the non-game-based training session, and GPA and performance during the tournament, shown with confidence intervals.

of 1.53, while the mean performance by the strong male students was 5.33, with a SD of 1.52. In this case, the gap is just 0.25 points. It is much lower than the gap in GPA. In terms of SD, the gap in the tournament is 0.16 SD of the GPA for the weak male students. Therefore, the gap during the tournament is 9 % of the gap in GPA. This is a huge decrease; much bigger (3 times) than the decrease in the gap for the girls. The empirical evidence regarding the gender difference for the improvement made by weak students therefore seems to confirm the hypothesis that this improvement is mainly caused by the inter-classroom nature ("Us" against "Them") of the game used during the tournament, and not just the game-based nature of the activity itself.

An independent study during a similar tournament held in December 2014 confirms the hypothesis that the main motivation comes from playing against a rival from another class. In this 2014 tournament, eight $4^{th}$ grade classes competed during the official tournament. While playing the game, a quick, two-question survey was conducted. The first question was: Who are you playing against? The second question asked the students to select one of the 6 options that were listed, regarding their preference for doing math exercises. From a total of 159 $4^{th}$ graders that competed in the tournament, 128 answered the survey. As shown in Table 1, 56.3 % prefer doing exercises by playing a social game against a rival; particularly if the rival is from another school. However, it is interesting to note that girls selected the option of doing exercises on their own in their notebook much more than boys. Theses gender differences agree with predictions from evolutionary studies. According to Geary [10], "selection pressures favored the evolution of motivational and behavioral dispositions in boys and men that facilitate the development and maintenance of large, competitive coalitions and result in the formation of within-coalition dominance hierarchies".

**Table 1.** Preferences for practicing math problems

| | Males | Females | Total |
|---|---|---|---|
| I prefer doing exercises using the *Espiral Mágico* Board Game with a rival from another school | 33.80 % | 25.90 % | 30.50 % |
| I prefer doing exercises using the *Espiral Mágico* Board Game with a rival from my class | 29.70 % | 20.40 % | 25.80 % |
| I prefer doing exercises using the *Espiral Mágico* Board Game with a bot rival | 9.50 % | 5.60 % | 7.80 % |
| I prefer doing exercises on my own on the computer | 12.20 % | 16.70 % | 14.10 % |
| I prefer doing exercises on my own in my notebook | 5.40 % | 24.10 % | 13.30 % |
| I prefer doing exercises on the whiteboard | 9.50 % | 7.40 % | 8.60 % |

## 4 Conclusions and Practical Implications

There is a long tradition of collaborative learning and team-based activities. From 1898 to 1989, over 500 experimental and 100 correlational cooperative learning studies have been conducted [14]. According to Slavin et al. [22] cooperative learning is very effective in elementary mathematics education. However, it is not used much in schools. Mevareth and Kramarski [16] argue that the main reason that co-operative learning has not always fulfilled its potential is the difficulty of guiding students in how to monitor, control and evaluate their learning. Without this guide, metacognition is not promoted and therefore student interactions are ineffective. Slavin [16] has another potential explanation. He cites observational studies which document that cooperative learning is still informal and does not include group goals or individual accountability.

However, with synchronous online inter-class tournaments there is a real opportunity to overcome these difficulties. According to Johnson et al. [14], in cooperative learning the most important element is positive interdependence: “a clear task and a group goal must be given so that students must believe that they sink or swim together”. With the synchronous online inter-class tournament there is a common goal, which is shared by the whole class. This key element is explicitly highlighted by publishing the class rankings while students are competing against other classes. In fact, in the implementation we have described, the ranking is published every 5 min in order to continuously remind students of the shared goal. Another key element is individual and group accountability [14]. In these tournaments, the platform also keeps track of the performance of each individual student. The game also includes instant feedback and metacognition. For example, in the training sessions the teacher can freeze the game, as can be done in basketball, and can pose open questions that can be answered as free text. The teacher is therefore transformed into a coach, who is constantly providing the class with cognitive and emotional support. The emotional connection with the students is therefore hugely facilitated. On the other hand, the tournament presenter also has a critical role in promoting metacognition. This role is particularly intensive in the training sessions, where the game is played before the official tournament. The presenter comments on the strategies developed by students from different classes, encourages the comparison of strategies, as well as encouraging students to reflect on the mathematical concepts and methodologies.

The results from a 2015 online inter-school tournament are very promising. 217 students prepared over the course of three training sessions and then participated in the tournament. There was a huge decrease in the performance gap between strong and weak students. This decrease was caused by an improvement among the weak students. It seems that the ancestral and social game-based nature of inter-group conflict is a very important motivational mechanism for these students. The improvement was more significant among male students. This is an interesting finding, which agrees with predictions taken from evolutionary psychology.

Traditional mathematics classes dedicate a significant amount of time to practice. Notebooks and worksheets are full of exercises. However, the proportion of personalized feedback received from the teacher or peers is very low. Web-based games facilitate a practice strategy, with constant, personalized feedback, detailed monitoring of each student's progress, balanced coverage of the curriculum, as well as opportunities for metacognitive reflection and social learning. However, online inter-school tournaments provide a unique and critical benefit: the classroom is no longer isolated. Classrooms can be connected to each other in an active, synchronized network. In this case, each class competes against the other classes. Therefore, the ancient tribal hunter-gatherer emotions and group identity sentiments are activated and with them emerges intra-class collaboration and a high level of engagement.

According to [19], games and gamification are the experimental petri dish for 21$^{st}$ century social thought, and they represent a rethinking of the assessment mechanisms used in schools to make them more effective and more democratic. However, most of the motivational mechanisms that have been used in gamification are aimed at the individual [25]. The ancestral inter-group motivational mechanisms have been underused in education. At most, they have been used with small teams belonging to the same class. The experience obtained from inter-classroom, synchronized online tournaments opens the door to new opportunities. It provides a strategy for connecting classrooms, for reducing teacher and classroom isolation, and for implementing new forms of learning and engagement that were previously impossible without the latest information technology. The impact is very interesting and powerful. It attracts the attention and motivation of students, particularly those who are weaker at math and harder to motivate. This extra motivation and energy boosts their performance and reduces the academic gap with students that are stronger at math. The data suggests a very important hypothesis: part of the academic gap is due to motivation, not ability.

**Acknowledgements.** We are thankful to Gonzalo Navarrete, mayor of the district of Lo Prado and President of the Comisión de Educación de la Asociación Chilena de Municipalidades; to Maximiliano Ríos, Director of Corporación Lo Prado; to Jorge Morkoff from Santa Rita school; to Basal Funds for Centers of Excellence Project FB 0003 from the Associative Research Program of CONICYT.

## References

1. Araya, R., Jimenez, A., Bahamondez, M., Dartnell, P., Soto-Andrade, J., Calfucura, P.: Teaching modeling skills using a massively multiplayer on line mathematics game. World Wide Web J. **17**(2), 213–227 (2014). Springer Verlag
2. Araya, R., Jiménez, A., Bahamondez, M., Dartnell, P., Soto-Andrade, J., González, P., Calfucura, P.: Strategies used by students on a massively multiplayer online mathematics game. In: Leung, H., Popescu, E., Cao, Y., Lau, R.W., Nejdl, W. (eds.) ICWL 2011. LNCS, vol. 7048, pp. 1–10. Springer, Heidelberg (2011)
3. Araya R.: What is inside this box: look at these other opened boxes for clues. In: Fifth Conference of the European Society for Research in Mathematics Education. Group 1: The role of Metaphors (2007)
4. Cuban, L.: Inside the Black Box of Classroom Practice. Harvard Education Press, Cambridge (2013)
5. Devlin, K.: Mathematics Education for a New Era: Video Games as a Medium for Learning. Peters/CRC Press, Natick (2011)
6. DeSteno, D.: The Truth About Trust. Random Press, New York (2014)
7. Diamond, J.: The World Until Yesterday. Penguin Press, London (2012)
8. Edwards, K., De Vries, D., Snyder, J.: Games and teams: a winning combination. Report 135. Center for Social Organization of Schools. Johns Hopkins University (1972)
9. Geary, D.: Educating the evolved mind: conceptual foundations for an evolutionary educational psychology. In: Carlson, J.S., Levin, J.R. (eds.) Psychological Perspectives on Contemporary Educational Issues. Information Age Publishing, Greenwich (2007)
10. Geary, D., Byrd-Craven, J., Hoard, M., Vigil, J., Numtee, C.: Evolution and development of boys social behavior. Dev. Rev. **23**, 444–470 (2003)
11. Gee, J.: The Anti-education Era. Creating Smarter Students Through Digital Learning. Palgrave Macmillian, New York (2013)
12. Greene, J.: Moral Tribes. The Penguin Press, New York (2013)
13. Henrich, J.: The Secret of Our Success. Princeton University Press, Princeton (2016)
14. Johnson, D., Johnson, R., Johnson, E.: Circles of Learning. Cooperation in the Classroom. Interaction Book Company, New York (1984)
15. Labaree, D.: Someone has to Fail. The Zero-Sum Game of Public Schooling. Harvard University Press, Cambridge (2010)
16. Mevarech, Z., Kramarski, B.: Critical Maths for Innovative Societies. OECD Publications, New York (2014)
17. Pellegrini, A.: The Role of Play in Human Development. Oxford University Press, New York (2009)
18. Ramani, G.B., Siegler, R.S., Hitti, A.: Taking it to the classroom: number board games as a small group learning activity. J. Educ. Psychol. **104**, 661–672 (2012)
19. Ramirez, R.; Squire, K.: Gamification for Education Reform (2014)

20. Siegler, R., Araya, R.: A computational model of conscious and unconscious strategy discovery. In: Kail, R.V. (ed.) Advances in Child Development and Behavior, vol. 33, pp. 1–42. Elsevier, Oxford (2005)
21. Slavin, R.: Co-operative learning: what makes groupwork work? In: Dumont, H., Istance, D., Benavides, F. (eds.) The Nature of Learning: Using Research to Inspire Practice. Educational Research and Innovation. OECD Publishing, Paris (2010)
22. Slavin, R., Lake, C.: Effective programs in elementary mathematics: a best-evidence synthesis. Rev. Educ. Res. **78**(3), 427–515 (2008)
23. Triplett, N.: The dynamogenic factors in pacemaking and competition. Am. J. Psychol. **9**(4), 507–533 (1898)
24. U.S. Department of Education. Foundations for Success: Report of the National Mathematics Advisory Panel (2008)
25. Vassileva, J.: Motivating participation in social computing applications: a user modeling perspective. User Model. User-Adapt. Interact. **22**, 177–201 (2012)
26. Zajonc, R.B.: Social facilitation. Science **149**, 269–274 (1965)
27. Zajonc, R.: Attitudinal effect of mere exposure. J. Pers. Soc. Psychol. **7**(4), 461–472 (1968)

# How to Attract Students' Visual Attention

Roberto Araya(✉), Danyal Farsani, and Josefina Hernández

Centro de Investigación Avanzada en Educación,
Universidad de Chile, Santiago, Chile
roberto.araya.schulz@gmail.com,
danyalfarsani@corpotalk.co.uk,
josefina.hernandez@ciae.uchile.cl

**Abstract.** Attracting students' visual attention is critical in order for teachers to teach classes, communicate core concepts and emotionally connect with their students. In this paper we analyze two months of video recordings taken from a fourth grade class in a vulnerable school, where, every day, a sample of 3 students wore a mini video camera mounted on eyeglasses. We looked for scenes from the recordings where the teacher appears in the students' visual field, and computed the average duration of each event. We found that the student's gaze on the teacher lasted 44.9 % longer when the teacher gestured than when he did not, with an effect size (Cohen's d) of 0.69. The data also reveals different effects for gender, subject matter, and student Grade Point Average (GPA). The effect of teacher gesturing on students with a low GPA is higher than on students' with a high GPA. These findings may have broad significance for improving teaching practices.

**Keywords:** Eye gaze · Hand gestures · Video analysis · Classroom practices

## 1 Introduction

"Culture hides more than it reveals, and strangely enough what it hides, it hides most effectively from its own participants" [20]. This quote fits very well with a Persian proverb and well-known aphorism that has been cited in many ethnographic papers: "a fish is the last creature to discover water". Being immersed in and surrounded by water makes it invisible and almost impossible to notice for the fish. Thus, this paper attempts to scrutinize and reveal the "visibility" and "familiarity" of everyday classroom interactions from the students' perspective, which is often invisible and unfamiliar to us as educators. Our goal is to investigate and reveal some insights into student gazes, trying to achieve an understanding of the situation by closely attending to and documenting the particulars from the students' perspective. Our approach follows Brown's [5] observation that the processes that lead to knowledge construction are habitually and locally situated in nature, as well as Seeley et al.'s [31] observation that "ignoring the situated nature of cognition, education defeats its own goal of providing useable, robust knowledge".

Understanding patterns of classroom interaction between teacher and students, as well as between students themselves, has been an area of interest for teachers. Many ethnographic studies have been conducted to understand the meaning-making practices


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 30–41, 2016.
DOI: 10.1007/978-3-319-45153-4_3

that naturally and normally occur in mainstream schools [10], complementary schools [8] and, in particular, mathematics classrooms [35]. In most studies concerning classroom interaction, an "outsider" enters a classroom, video tapes the lesson, takes field notes and pretends to be a fly on the wall. The outsider's visit to the classroom can last weeks or even months. There is a good chance that the outsider's presence impacts on what is observed. There is therefore a major issue concerning the extent to which an observer affects the situation under observation [9].

Observing involves interpretation by the observer "who has desires and prejudices, sensitivities and propensities" [26]. As observers in the classroom, we observe what we are prepared to observe and we notice what we are sensitized to notice [26]. This fact makes the observer part of the observation. Furthermore, with classroom observation there is no objective place to stand; all observation involves standing somewhere, which subsequently influences what is seen. A classroom observer makes 'choices' and 'decisions' [11] concerning the timing and setting of the observation. For example, during the video recording process we make choices influenced by our "identities and intentions, choices that are also affected by our relationship with the subject" [7]. "The focus of the video camera is selective" [3] and "every camera position excludes other views of what is happening" [17]. Moreover, video recordings produce rich data but only capture a partial view of the social interaction [14]. In practice, recordings that are generated through the lens of a single camera do not capture the whole classroom interaction. The data that is obtained from a single camera has a single focus of attention, whereas students and teachers are capable of focusing on multiple aspects of a complex setting [28]. Therefore, the video recording process can be problematic because there are choices which influence when and whom to record.

Even though there are epistemological issues concerning the validity of data while having a researcher who does not normally belong to the naturally occurring setting, we cannot study the classroom practices outside of its naturally occurring context. "We cannot study the social behavior of a fish by taking him out of water. The child is a child in his world" [4]. Therefore, with the right approach, classrooms can be a natural laboratory for studying situated learning [2]. With this in mind, our approach in this study was to ask students to wear an eyeglass with a mini video camera mounted on it. This way, without having the presence of an "outsider", we would able to observe and document the classroom interaction from the students' own ontological orientations. This approach would enable us to detect who is looking at whom and for how long their visual attention is maintained. Our particular interest for this paper is to identify the students' visual attention on the instruction when the teacher is gesturing, versus when they are not. Furthermore, we want to identify whether the duration of the gaze pattern is different for different subjects. Our study satisfies the fundamental test of research [33], i.e. our results have predictive power. For example, gestures made by the teacher in situated learning are more effective in attracting the students' visual attention. It is particularly important to note that we are able to make predictions, despite the presence of the difficulties suggested by Rudolph, i.e. this is an observation study of classroom practices, which is a very complex environment that depends on several variables, as well as being a "far different phenomena from those studied in controlled laboratory settings" [30].

### 1.1 The Role of Gestures in Teaching and Learning

Understanding classroom communication has been a curious and an interesting area of research [32]. However, classroom communication is not only restricted to verbal messages, which has traditionally been the focus of study. An important aspect of communication in teaching and learning is its multimodal nature [13], which is embodied in learning [25]. The term multimodality refers to the complex repertoire of semiotic resources that interlocutors draw on in different social encounters. For example, a multimodal approach involves looking at language and other means of making meaning, such as images, text, graphic symbols, and gestures [22]. In recent years, researchers have looked at how gestures are used in facilitating language production, as well as promoting learner comprehension [1, 24].

Speakers adjust the frequency and size of their gestures in accordance with how likely a gesture is to benefit their listener's understanding [21]. Gestures have been found to complement the verbal message beyond their semantic meaning. Farsani [12] examined a teacher's gestural representation that helped convey both the concept and a new mathematic register to bilingual learners with different levels of English proficiency. In describing the mathematical notion of 'isosceles triangle', the classroom teacher pointed to his eyes as he uttered 'isosceles' in his speech. Phonologically, there is a strong connection between 'isosceles' and how it is pronounced (eye-sosceles). Therefore, the teacher's gesture acted as a mnemonic device to help remember key terminology and a mathematical concept by emphasizing how our eyes are like the two identical sides of an isosceles triangle. Gestures may be more helpful to listeners with weak verbal skills than to listeners with strong verbal skills [21]. However, it is important not to overinflate gestures as interpretive resources. As with any other form of data in qualitative research, gestures are transitory, ephemeral, partial and incomplete, and need to be considered and evaluated in relation to their accompanying verbal message. Particular attention must therefore be paid to 'hearing gestures' [16], in the same way that it is important to 'hear' speech. Speakers often change the size of their gestures, as though they intend for larger gestures to be particularly communicative, while also producing larger gestures when they are describing information that is unknown to their listener or when they are particularly motivated to communicate clearly [21]. Given this, we are therefore interested in discovering to what extent students glean information from their teacher's gestures. What is really happening from the students' perspective when a teacher is gesturing? Where exactly do learners place their visual attention when a teacher is gesturing?

Traditional research into classroom interaction often fails to acknowledge the reciprocal visual attention between students and the classroom teacher. Only recently has technology allowed researchers to look into the 'black box' of classroom practices. For example, a gaze tracker has enabled researchers to document and identify fine-grain information about learners' visual attention in real-time or moment-to-moment activities as they are engaged in their routine classroom practices [15, 29, 34]. As a methodology, gaze tracking seems to be a promising tool for fine-grain analysis of meaning-making practices during classroom communication, as well as student attention. To our knowledge, no previous study has ever analyzed the duration of students' visual attention while the teacher is naturally and spontaneously gesturing during his instructional talk. Furthermore, we are interested in examining whether the attention on the teacher is more sustained if the teacher makes gestures, in comparison to instances when no gestures are made.

## 2 Process of Data Collection and Analysis

The data that emerges in this paper is part of a larger dataset which investigated the interactional patterns in a classroom by examining the gaze between students and the classroom teacher [2]. From September 26th, 2012 until November 27th of 2012, a fourth grade classroom teacher and a sample of three students selected each day were asked to wear a mini video camera, which was mounted on eyeglass frames. The original eyeglass lenses were removed so as to minimize the weight and facilitate the original view. Each day, the teacher and students had to wear the eyeglass for approximately six hours. Students took the eyeglasses off during breaks and lunch time, as well as when they went to the bathroom. The class consisted of 36 students (21 boys and 15 girls) and the average age of the students was 10.5 years. Both the parents of the students and students themselves gave signed consent to wear these eyeglasses, as well as agreeing to allow any information that was obtained to be disseminated both in professional conferences and in journal articles.

The recordings were manually downloaded at the end of each day. A total of 12,133 min of interactional data was recorded, from which 2,600 min came from the teacher and the rest from the students. In this study, we were primarily interested in looking at instances where the students visually focused on the teacher as he was conveying the instructional information. Considering that the videos had a recording quality of thirty frames per second (30 fps), every second, or every thirty frames, a frame was sampled from the videos. Each frame that was obtained was processed using the OpenCV software in order to detect the presence of faces. A total of 24,148 faces were detected and each face was saved as an image file. Each facial image was then processed semi-automatically using the Google Picasa software in order to identify the subject. Picasa initially identified around 60 % of the faces, and after a few iterations of training, where the software asked us to confirm some of the automatic identifications; it ended up identifying 80 % of the persons. The remaining images, mostly the low-resolution images of faces, were subsequently identified manually.

Of all of the detected and identified faces, there were a total of 857 frames (still images) where students looked at the classroom teacher. In another study [2], we analyzed all the frames, including the gazes among peers. In this particular study we were interested in instances where students kept their focus on the teacher, therefore of the 857 frames where students looked at the classroom teacher, we decided to discard frames where there were other faces (or distractions) present in the same frame. For example, in Fig. 1, we are not sure whether the attention is on the teacher or the other student (given that the OpenCV software identified and saved both faces). We therefore primarily looked at frames where only the teacher's face was present. Furthermore, visibility of the teacher's gestural enactment was essential. Instances where the teacher's hands were blocked due to an obstacle (e.g. a chair, a desk or the student sitting in front) were not considered. Clarity of the frames was also important; therefore if


**Fig. 1.** In this frame, the eye gaze and the attention could be on the teacher, the other boy (who is also wearing the eyeglasses), or on both

a student moved his/her head fast and suddenly, it often generated a blurred frame, which was also discarded. With this restriction, we obtained 264 frames from the total of 857 that were initially generated by the computer.

Furthermore, given the fact that two consecutive frames from the same video camera where the student is looking at the teacher (consecutive meaning that they are only one second apart) do not represent two independent gazes, rather the same gaze that was maintained for more than one second, we define a 'scene' as two or more consecutive frames coming from the same student's camera. Of the 264 frames where the students were looking directly at the teacher, we found 83 scenes, with the shortest containing only two frames, and the longest containing up to ten frames. Of the 83 scenes, 43 scenes correspond to when the teacher made some kind of gesture and 40 scenes correspond to when no gesture was made during his instructional talk.

Using these restrictive categorizations made our interpretation and analysis of the frames more effective. Each of our team members examined every scene in order to look for subtle and silent hand gestures [27]. Reading still images [23] was, indeed, an integral part of the analysis, noting what each student and teacher did, moment by moment. For example, Fig. 2 shows two frames where the classroom teacher is using his gesture space to convey his instructional talk. His gestures can be spontaneous as well as deliberate, synchronous or asynchronous. Gestures could be used to: align with prosodic prominence patterns in his speech (as politicians often do), pantomime to accentuate his verbal message visually, or point to objects or information on the board. The gestures that are employed here could be used for disciplinary remarks and/or pedagogical practices. On the other hand, Fig. 3 illustrates two consecutive frames where the students' gaze is maintained on the teacher, but the teacher is not gesturing.

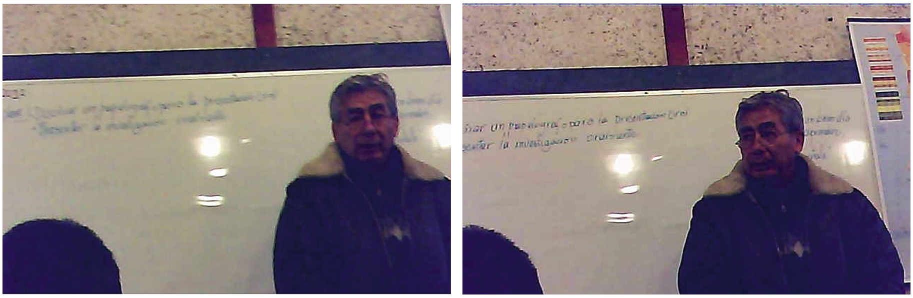


**Fig. 2.** Two consecutive frames illustrating when the teacher is not gesturing

**Fig. 3.** Two consecutive frames illustrating when the teacher is gesturing

## 3 Results and Discussion

In this paper we consider 'attention' to be the focus of the student's gaze. Of course, this may not always be the case. It is possible for a student to focus on a visual target (teacher) without paying attention to it (i.e. 'looking without observing') and, conversely, paying attention to something without directly focusing on it ('observing without looking') [18]. There is a very crucial and subtle difference between the two. However, in this paper we are assuming that the duration of the students' visual gaze on the teacher is the same as the duration of the students' visual attention on the teacher. With this in mind, we analyzed the moments when a teacher made a gesture as he conveyed his instructional information, in comparison to instances when no gestures were made. What we were particularly interested in identifying was: (1) for each gaze, who is the subject that is looking at the teacher, in terms of their gender and Grade Point Average (averaged annual school grades in the subjects in study, GPA); (2) the duration of the students' visual attention on the instruction while the teacher was gesturing versus when they were not; and (3) whether the duration of a gaze pattern was different for different subjects, specifically in mathematics lessons.

First let us review the mean duration of the students' visual attention on the teacher when no gesture was made. In this sense, there were 40 scenes, each scene containing a minimum of 2 consecutive frames and a maximum of 8. The analysis revealed a mean of 2.58 s, with a standard deviation of 1.24. In contrast, the other group consisted of moments when the teacher gestured during his instructional talk, with a total of 43 scenes comprising 161 frames. In this group, there was a minimum of two consecutive frames and a maximum of 10 consecutive frames, with a mean of 3.74 s and a standard deviation of 2.04. The difference in the amount of time that students gaze at the teacher when he is gesturing versus when he is not gesturing indicates a 44.9 % increase in the students' visual attention for moments where the teacher gestured, with a p-value of 0.002. We estimated the effect size for different lengths of gaze by calculating Cohen's d, where Cohen's d is defined as:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}, \tag{1}$$

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}. \tag{2}$$

This gave us a value of 0.69.

We also analyzed the students according to their GPA, whereby students with an above-average GPA are defined as having a "High GPA" and students with a below-average GPA as having a "Low GPA". We obtained 63 scenes where high GPA students looked at the teacher and only 20 scenes where low GPA students looked at him. Even so, of the 63 scenes when High GPA students look at the teacher, only 46 % correspond to when he is gesturing. In contrast, of the 20 scenes from when the low GPA students look at the teacher, 70 % correspond to when he gestures. The p-value of the difference between these proportions is 0.031, with a Cohen's d of 0.48. This means that low GPA students focus their attention on the teacher more when he gestures than High GPA students, compared with when he does not gesture.

Let us now compare and contrast the students' visual attention in mathematics lessons, compared to all other subjects. For mathematics lessons, we found 29 scenes: 13 from instances where no gestures were made and 16 from moments in which the teacher made gestures. We found a mean of 2.69 s with a standard deviation of 1.18 for when no gestures were made, versus 3.94 s for instances when gestures were made, with a standard deviation of 2.41 (p-value 0.1, Cohen's d 0.64). Even though these differences are not statistically significant, there is a 46.8 % increase in the students' visual attention for moments in which the teacher made gestures in the mathematics classrooms. Furthermore, it appears that in mathematics lessons students visually pay more attention to the teacher (regardless of whether he is making a gesture or not) than in any other subject. In addition to this, if the teacher gestures in a mathematics class, 69 % of the students will look at him, whereas when he gestures in other subjects, only 50 % will look at him. The difference in these proportions has a p-value of 0.049, and Cohen's d of 0.38 (Tables 1 and 2).

**Table 1.** Effect of teacher's gestures on the students' visual attention according to GPA, gender and subjects: Effect on the proportion of gazes on the teacher when the teacher gestures

| | **High GPA (number of scenes)** | **% of High GPA Students** | **Low GPA (number of scenes)** | **% of Low GPA Students** | **Comparison** | **P-Value of difference between the proportions (one tail)** | **Cohen's D for the difference between the proportion** |
|---|---|---|---|---|---|---|---|
| **Teacher gestures** | 29 | 46% | 14 | 70% | ***High GPA vs. Low GPA:*** | **0.031** | 0.48 |
| **Teacher does not gesture** | 34 | 54% | 6 | 30% | | | |
| ***Total*** | ***63*** | ***100%*** | ***20*** | ***100%*** | | | |
| | **Boys (number of scenes)** | **% of Boys** | **Girls (number of scenes)** | **% of Girls** | | | |
| **Teacher gestures** | 35 | 56% | 8 | 40% | ***Boys vs. Girls:*** | **0.115** | 0.31 |
| **Teacher does not gesture** | 28 | 44% | 12 | 60% | | | |
| ***Total*** | ***63*** | ***100%*** | ***20*** | ***100%*** | | | |
| | **Math (number of scenes)** | **% of Math** | **All other subjects (number of scenes)** | **% of all other subjects** | | | |
| **Teacher gestures** | 20 | 69% | 27 | 50% | ***Math vs. All other subjects:*** | **0.049** | 0.38 |
| **Teacher does not gesture** | 9 | 31% | 27 | 50% | | | |
| ***Total*** | ***29*** | ***100%*** | ***54*** | ***100%*** | | | |

**Table 2.** Effect of teacher's gestures on the students' visual attention according to GPA, gender and subjects: Effect on the length of gazes focused on the teacher (measured in seconds) when the teacher gestures

| | AVG gaze when teacher does not gesture (s) | SD | Total number of scenes | AVG gaze when teacher gestures (s) | SD | Total number of scenes | Comparison | P-Value of diff. between length of gazes (two tail) | Cohen's D of diff. between the length of gazes |
|---|---|---|---|---|---|---|---|---|---|
| **All students** | 2.58 | 1.24 | 40 | 3.74 | 2.04 | 43 | ***Gesture vs. No Gesture for all students*** | **0.002** | 0.69 |
| **High GPA** | 2.58 | 1.33 | 34 | 3.69 | 2.05 | 29 | ***Gesture vs. No Gesture for High GPA Students*** | **0.013** | 0.65 |
| **Low GPA** | 2.50 | 0.55 | 6 | 3.86 | 2.07 | 14 | ***Gesture vs. No Gesture for Low GPA Students*** | **0.136** | 0.76 |
| **Boys** | 2.43 | 0.74 | 28 | 3.80 | 2.15 | 35 | ***Gesture vs. No Gesture for Boys*** | **0.002** | 0.82 |
| **Girls** | 2.92 | 1.98 | 12 | 3.50 | 1.51 | 8 | ***Gesture vs. No Gesture for Girls*** | **0.489** | 0.32 |
| **Math** | 2.69 | 1.18 | 13 | 3.94 | 2.41 | 16 | ***Gesture vs. No Gesture in Math*** | **0.100** | 0.64 |
| **Other Subjects** | 2.52 | 1.28 | 27 | 3.63 | 1.82 | 27 | ***Gesture vs. No Gesture in all other subjects*** | **0.012** | 0.71 |

## 4 Conclusions and Practical Implications

Our results not only support previous findings [2], but also reveal more about the nature of students' visual attention with regards to teacher gestures. Although this study only featured one fourth grade classroom, belonging to a district with one of the lowest levels of socioeconomic status in Chile, as well as only one teacher and three students selected every day (with all students wearing the eyeglasses at least twice), the findings have a more general predictive power. As Wieman [33] noted, "a good qualitative study that examines only a few students or teachers in depth will allow one to recognize, and hence more accurately predict, some factors that will be important in educational outcomes and important in the design of larger quantitative experiments in similar populations". Although our study is on a small scale, we can generate precise quantitative predictions with reasonable accuracy regarding what is likely to be observed in student behavior within situated classrooms.

This paper suggests that students paid more attention to the teacher when the instructional talk was accompanied by gestures. Specifically, if a teacher gestures, this would have a higher effect on students with a Low GPA than students with a higher GPA, as well as on boys. On this matter, a future study would be to analyze if this effect is maintained if the teacher were female instead of male. Given that this study counted only with one male teacher, this may affect the pupil's ability to relate to him, given that students may be sensitive to role models with the same gender. The teacher's gestures in mathematics lessons played an even more crucial role in capturing the students' visual attention. It appears that the students' visual attention on the teacher in mathematics lessons was higher than in other subjects. The teacher's gestures, therefore, appeared to act as nonverbal amplifiers for maintaining the students' visual attention for longer.

The implications of this study raise awareness of how technology can be used to understand fine-grain meaning-making practices during classroom interactions that can be relevant in transforming practice [19]. We would like to conclude this section by reflecting on a recent observation that was made by Castañer and her colleagues [6]. They believe that, regardless of a teacher's experience and qualifications, it is always worth questioning the forms, styles and quality of the messages that are conveyed verbally and nonverbally in their professional teaching practices. We believe that optimization of these very subtle and silent nonverbal messages can have a direct, positive impact on the teaching and learning process. One recommendation and practical application is to incorporate nonverbal training in teacher education courses, both for pre-service and in-service teachers, in order to raise awareness of the communicative function of nonverbal language. In other words, we must not only consider the pedagogical effects of gestures in teaching and learning, but also how these can be used for disciplinary purposes.

The findings of this study open a new window of investigation and give rise to the following future research questions: would we obtain the same results if we had conducted this same experiment in countries where people are known to gesture greatly, or, to the contrary, countries where people are known to be less expressive? And also, what other nonverbal variables affect the flow of interaction between teacher and students, as well as among students themselves?

**Acknowledgements.** We are thankful to all the Santa Rita School staff; in particular, the enthusiasm and collaboration of the fourth grade teacher Stenio Morales, that was the subject of this paper. We also thank Paulina Sepúlveda and Luis Fredes for the development of the software; to Avelio Sepúlveda, Johan van der Molen, and Amitai Linker for preliminary statistical analysis; to Marylen Araya and Manuela Guerrero for manual classification of faces obtained from the videos; and to Ragnar Behncke for his participation in the design of the measurement strategy and data gathering process. We also thank the Basal Funds for Centers of Excellence Project BF 0003 from CONICYT.

## References

1. Alibali, M.W., Nathan, M.J.: Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. Journal of the Learning Sciences **21**(2), 247–286 (2012)
2. Araya, R., Behncke, R., Linker, A., van der Molen, J.: Mining social behavior in the classroom. In: Núñez, M., Nguyen, N.T., Camacho, D., Trawinski, B. (eds.) ICCCI 2015. LNCS, vol. 9330, pp. 451–460. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24306-1_44
3. Bezemer, J., Mavers, D.: Multimodal transcription as academic practice: a social semiotic perspective. Int. J. Soc. Res. Methodol. **14**(3), 191–207 (2011)
4. Birdwhistell, R.: Kinesics and Context: Essays in Body Motion Communication. Penguin, Harmondsworth (1970)
5. Brown, J.S., Collins, A., Duguid, P.: Situated cognition and the culture of learning. Educ. Researcher **18**, 32–42 (1989)
6. Castañer, M., Camerino, O., Anguera, M.T., Jonsson, G.K.: Kinesics and proxemics communication of expert and novice PE teachers. Qual. Quant. **47**(4), 1813–1829 (2013)
7. Collier, M.: Approaches to analysis in visual anthropology. In: van Leeuwen, T., Jewitt, C. (eds.) Handbook of Visual Analysis, pp. 35–60. Sage publications, Thousands Oaks (2001). London, New Delhi
8. Creese, A., Bhatt, A., Bhojani, N., Martin, P.: Fieldnotes in team ethnography: researching complementary schools. Qual. Res. **8**, 223–242 (2008)
9. Emerson, R.M., Fretz, R.I., Shaw, L.L.: Writing Ethnographic Fieldnotes. The University of Chicago Press, Chicago, London (1995)
10. Erickson, F.: What makes a good ethnography "ethnographic"? Coun. Anthropol. Educ. Newslett. **4**(2), 10–19 (1973)
11. Erickson, F.: Qualitative methods in research in teaching. In: Linn, R.L., Erickson, F. (eds.) Research in Teaching and Learning, vol. 2, pp. 119–161. Macmillan, New York (1990)
12. Farsani, D.: Deictic gestures as amplifiers in conveying aspects of mathematics register. In: Proceedings of the 9th Conference of European Society for Research in Mathematics Education, Prague, Czech (2015a)
13. Farsani, D.: Making multi-modal mathematical meaning in multilingual classrooms. Unpublished Ph.D. thesis, University of Birmingham (2015b)
14. Flewitt, R.: Using video to investigate preschool classroom interaction: education research assumptions and methodological practices. Vis. Commun. **5**(1), 25–51 (2006)
15. Garcia, E., Hannula, H.: Using gaze tracking technology to study student visual attention during teacher's presentation on board. In: Proceedings of the 9th Conference of European Society for Research in Mathematics Education, Prague, Czech (2015)
16. Goldin-Meadow, S.: Hearing Gesture: How Our Hands Help Us Think. Harvard University Press, Cambridge (2003)

17. Goodwin, C.: Practices of seeing: visual analysis: An ethnomethodological approach. In: van Leeuwen, T., Jewitt, C. (eds.) Handbook of Visual Analysis, pp. 157–182. Sage publications, Thousands Oaks (2001). New Delhi, London
18. Gullberg, M., Kita, S.: Attention to speech-accompanying gestures: eye movements and information uptake. J. Nonverbal Behav. **33**, 251–277 (2009)
19. Gutierrez, K., Penuel, R.: Relevance to practice as a criterion for rigor. Educ. Res. **43**(1), 19–23 (2014)
20. Hall, E.T.: The Silent Language. Fawcett, Greenwich (1959)
21. Hostetter, A.: When do gestures communicate? A meta-analysis. Psychol. Bull. **137**(2), 297–315 (2011)
22. Jewitt, C.: Multimodal discourses across the curriculum. In: Martin-Jones, M., de Mejia, A., Hornberger, N. (eds.) Encyclopedia of Language and Education, Volume 3: Discourse and Education, pp. 357–367. Springer (2008)
23. Kress, G., van Leeuwen, T.: Reading Images: The Grammar of Visual Design. Routledge, London (1996)
24. Lee, O., Fradd, S.H.: Science for all, including students from non-english language backgrounds. Educ. Res. **27**, 12–21 (1998)
25. Lindgren, R., Johnson-Glenberg, M.: Emboldened by embodiment: six precepts for research on embodied learning and mixed reality. Educ. Res. **42**, 445–452 (2013)
26. Mason, J.: Researching Your Own Practice: The Discipline of Noticing. Routledge, UK (2002)
27. Mehrabian, A.: Silent Messages. Wadsworth, Belmont (1971)
28. Pimm, D.: From Should to Could: reflection on possibilities of mathematics teacher education. Learn. Math. **13**(2), 27–32 (1993)
29. Rosengrant, D.: Using eye-trackers to study student attention in physical science classes. Bull. Am. Phys. Soc. **58** (2013). Working paper
30. Rudolph, J.L.: Why understanding science matters the IES research guidelines as a case in point. Educ. Res. **43**(1), 15–18 (2014)
31. Seeley, J., Collins, A., Duguid, P.: Situated cognition and the culture of learning. Educ. Res. **18**(1), 32–42 (1989)
32. Sinclair, J., Coulthard, M.: Towards an Analysis of Discourse. Oxford University Press, London (1975)
33. Wieman, C.E.: The similarities between research in education and research in the hard sciences. Educ. Res. **43**(1), 12–14 (2014)
34. Yang, F.Y., Chang, C.Y., Chien, W.R., Chien, Y.T., Tseng, Y.H.: Tracking learners' visual attention during a multimedia presentation in a real classroom. Comput. Educ. **62**, 208–220 (2013)
35. Zevenbergen, R.: Ethnography in the mathematics classroom. In: Malone, J., Atweh, B. (eds.) Aspects of Postgraduate Supervision and Research, pp. 19–38. Lawrence Erbaum & Association, Mahwah (1998)

# Creating Effective Learning Analytics Dashboards: Lessons Learnt

Sven Charleer[1(✉)], Joris Klerkx[1], Erik Duval[1], Tinne De Laet[2], and Katrien Verbert[1]

[1] Department of Computer Science, KU Leuven, Leuven, Belgium
{Sven.Charleer,Joris.Klerkx,Katrien.Verbert}@kuleuven.be
[2] Tutorial Services of Engineering Science, KU Leuven, Leuven, Belgium
Tinne.DeLaet@kuleuven.be

**Abstract.** Learning Analytics (LA) dashboards help raise student and teacher awareness regarding learner activities. In blog-supported and inquiry-based learning courses, LA data is not limited to student activities, but also contains an abundance of digital learner artefacts, such as blog posts, hypotheses, and mind-maps. Exploring peer activities and artefacts can help students gain new insights and perspectives on learning efforts and outcomes, but requires effort. To help facilitate and promote this exploration, we present the lessons learnt during and guidelines derived from the design, deployment and evaluation of five dashboards.

**Keywords:** Learning analytics · Learning dashboards · Information visualisation · Guidelines · Collaboration

## 1 Introduction

Learning Analytics (LA), or the collection and analysis of traces that learners leave behind, can help to understand and optimise (human) learning and the environments in which it occurs [40]. Furthermore, it can help raise awareness of personal and peer learning activities, help reflect on and make sense of learner traces, and impact behaviour [44]. Learning Dashboards (LD) often visualise **efforts** such as artefacts produced, time spent, social interaction, resource use, and exercise and test result [45]. However, only focusing on effort can have a detrimental effect on motivation [36]. A collection of efforts is part of progress towards a larger goal, such as learning a language, passing an exam, etc. Throughout our case studies, we have learnt that it is essential that students are continuously aware of the impact of their efforts towards these intended **learning outcomes**.

LA provides ways of taking these learner traces to help raise awareness of personal and peer learning activities, help reflect on and make sense of learner traces, and impact behaviour [44]. On the one hand, educational data mining techniques try to help students by making decisions on their behalf (like intelligent tutoring systems [4] and educational data mining systems (EDM) [38] do).


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 42–56, 2016.
DOI: 10.1007/978-3-319-45153-4_4

As such, they automatically use students' efforts to produce information regarding outcomes. For example, they show students that they have a calculated chance on passing the course, or they show which paper to read next. We however believe that it is highly important to empower students rather than automating the learning process. Indeed, technology can support the student to play a more active role in the LA reflection process role [12] instead. Clearly visualising the path from effort to outcome should thus be supported by all LA dashboards. There is also a certain philosophical or ethical side to this notion of two approaches. As Klerkx et al. [22] frame it: *"If learners are always told what to do next, then how can we expect them to possess the typical 21st century skills of collaboration, communication, critical thinking and creativity? Or, at a more fundamental level, can we expect students who are always told what to do next to become citizens equipped with the knowledge, skills and attitude to participate fully in society?"*

To facilitate this empowerment, this paper looks at what data needs to be accessible to students and how this data should be visualised. To discover knowledge relevant to their learning process, the empowerment should happen in their everyday lives, in and outside the classroom. We therefore also take into account the different contexts in which their learning occurs [17], and how we can leverage these contexts to promote students to explore the path from effort to outcome. We summarise this through two research questions:

– How should we visualise relevant data to facilitate students exploring the path from effort to outcomes? (RQ1)
– How can we promote students, inside and outside the classroom, to actively explore this effort to outcomes path? (RQ2)

We start by explaining the different course settings, the five dashboards and the evaluation setups in Sect. 2. Based on the design, deployment and evaluation of these dashboards, Sect. 3 explores the lessons learnt. Conclusions and remarks on future work are presented in Sect. 4.

## 2 Course Setting

We first explain the course settings in which our dashboards were deployed. We briefly discuss how the traces are collected, and present an overview of the dashboards and their evaluations.

### 2.1 Blog-Supported Courses

Blogging has become more popular in learning environments [46] as it facilitates assessment, reflection, interaction and collaboration among students, and improves participation in learning activities [24]. It allows students to develop their ideas and receive contributions from peers through blog comments [21,31]. During the face-to-face Master course "Human-Computer Interaction" of 2013 and 2014 at KU Leuven, students were required to use blogs to report progress, share opinions and knowledge [23], and provide feedback to peers through blog comments. Twitter was used as a communication channel for e.g. quick questions about the topic of the course or for sharing reading material. These on-line activities often generate an abundance of data. A typical course results in 140–300 blog posts, 600–1400 blog comments and 300–500 tweets.

### 2.2 Inquiry-Based Learning Courses

Contrary to a traditional passive role in a classroom, in Inquiry-Based Learning (IBL), learners assume an active role as explorer and scientist with a focus on learning how to learn. Teachers try to stimulate learners to pose questions and create hypotheses regarding a specific topic, perform independent investigations, gather data to confirm and discuss their findings and generate conclusions. In the on-line weSPOT Inquiry Environment[1], a teacher can set up an inquiry regarding a specific research topic. The students then use this on-line environment to create hypotheses, join discussions, generate mind-maps and conclusions. By taking pictures, recording videos, and registering measurements through a mobile application integrated into the IBL environment, students collect data in the field to support their hypotheses [27,33].

### 2.3 Learning Analytics Traces

To collect the learning traces from these learning environments, we use the architecture presented by Santos et al. [38]. For the blog-supported courses, trackers connect to the RSS feeds of the student blogs and utilise the Twitter API to track the activities related to the course hash-tag. The content of these activities, together with relevant meta-data (time of activity, student identification) is pushed to the Learning Record Store (LRS), which stores the data following a simplified xAPI format [38].

Through exposed REST services[2] of the weSPOT Inquiry Environment, the trackers access the learner artefacts (e.g. hypothesis created, picture taken, mind-map created) and meta-data (e.g. time of the activity, user identification, peer and teacher rating), and store the data in the LRS. The LRS exposes a set of REST services for data retrieval[3], which the dashboards use to request the relevant learner traces to populate the LD visualisations.

### 2.4 Deployed Dashboards

Five dashboards were developed and deployed during the course of three years. Each dashboard builds upon findings of the previous, taking into account the stakeholders and the specific learning context in which it will be deployed. They

1 http://inquiry.wespot.com.
2 http://goo.gl/37mr4D.
3 https://github.com/weSPOT/wespot_datastore.

**Table 1.** Overview of the dashboards and evaluation setups.

| Details | Navi Badgeboard (A) | Navi Surface (B) | Class View (C) | LARAe (D) | LARAe.TT (E) |
|---|---|---|---|---|---|
| References | [7,36] | [7] | [9] | [8] | [6] |
| Course setting | Master in Engineering course (16 weeks) | | | Master in Engineering course (16 weeks)<br>Multiple IBL courses at European Secondary schools | |
| Data | 142 blog posts<br>549 comments<br>548 tweets | | | 254 blog posts<br>1326 comments<br>352 tweets | Test IBL data |
| Activities accessible | x | x | x | x | x |
| Artifacts accessible | | | x | x | x |
| Learner path visualised | | | | | x |
| Visualisation methods | Abstraction of course goals through badges | Abstraction of course goals through badges | Overview + detail | Abstraction augmented with meta-data (rating, social interactions,..)<br>Overview + detail | Visual exploration |
| Main focus | Abstraction | Collaboration | Access to artifacts | Workflow Integration<br>Access to artifacts | Collaboration<br>Learner activity path |
| Research questions | RQ1 | RQ1,RQ2 | RQ1 | RQ1,RQ2 | RQ1,RQ2 |
| **Evaluation** | **Navi Badgeboard (A)** | **Navi Surface (B)** | **Class View (C)** | **LARAe (D)** | **LARAe.TT (E)** |
| Questionnaires | x | x | x | x | x |
| Usage Tracking | x | | | x | |
| Etnographic Field Study | | x | x | x | |
| Interviews | | x | x | x | |
| Focus group | | | | | x |
| Prototype Evaluation | | x | x | x | x |
| Pilot Run | x | | | x | |
| Student Participants | 22 Master students | 14 Master students | | 38 Master Students<br>Secondary school students | |
| Expert Participants | | | 6 with teaching responsibilities | 5 with teaching responsibilities<br>Teachers at secondary schools | 15 with teaching responsibilities and pedagogical research experience/interest |

are built as low-fidelity prototypes at first, with four high-fidelity dashboard prototypes deployed in authentic settings during pilot studies [25]. The dashboards were developed using web technologies such as D3.js, Processing.js, and Node.js. Table 1 provides an overview of the dashboards, their course setting and evaluations. Screenshots of the dashboards can be found in Fig. 1.

**Dashboard A:** Navi Badgeboard [7,35] presents the user with per student dashboards containing an overview of achieved (in colour) and still achievable (greyed out) goals through badges. Students can position themselves among peers through the number next to each badge, indicating the amount of students who have achieved this goal. A high number next to a grey badge thus indicates that the student is one of the few students without the badge. A low number next to a coloured badge indicates that the student is one of the few to have earned this badge. The dashboard is designed to work on mobile devices and desktop browsers.

**Dashboard B:** Navi Surface [7,35] is a multi-user interactive visualisation designed for large multi-touch tabletop displays. Out of a list of student names and course badges, both students and teacher can simultaneously drag badges and student names onto the screen. The dashboard then visualises the badge reward relationships by drawing lines from students to badges, providing an overview and comparison of achievements to drive the conversation.

**Dashboard C:** The Class View dashboard [9] is designed for large desktop monitors, interactive whiteboards and large touch displays. Four modules visualise the LA data in different ways: a student-badge matrix shows how many times a specific student has been awarded a specific badge. Activities and badges are visualised over time through five different bar charts, displaying the amount of activity done and badges awarded per day. Selecting a day will show the list of activities or badges awarded that day. In turn, selecting one of these activities visualises the content behind the analytic data (e.g. the text of the blog post). Another list of bar charts shows the number of awards given per badge. Two modules allow for the filtering of the above data. The user can set a time-range and split the data by grouping students. This facilitates student comparison, with each visualisation module displaying each group's data in different colours.

**Dashboard D:** The LARAe dashboard [8] visualises blog and Twitter activities of students. Following the "Overview+Context" approach, the overview shows circles coloured by age representing activities and are grouped by activity type (blog, blog comment, tweet, retweet) and by student group/staff. Selecting an activity updates the context part of the visualisation, showing a thread containing the activity content (e.g. the text of the blog post) and its related activities (e.g. blog comments). The activities in this thread are also highlighted in the overview, giving a visual overview of the distribution of people engaging with the selected activity. The number in each circle indicates the amount of activity (e.g. the number of comments on a blog post). The dashboard is designed to run on large displays, desktop computers and tablets.

**Dashboard E:** LARAe.TT's [6] main objective is to facilitate collaborative exploration of the learner paths, i.e. the chronological sequence of all activities and artefacts generated. The visualisation displays a horizontal line per activity thread, e.g. the creation of a hypothesis by a learner followed by every comment on, rating on, and edit of the hypothesis. The chronology is maintained across threads, enabling the user to see the impact an event might have had on other parallel threads. Figure 1E shows an example of this cross-thread relationship: Angela's comment on Geoff's hypothesis results in Geoff accessing her data collection and changing his hypothesis and conclusion.

## 3 Lessons Learnt and Guidelines

As presented in Table 1, the dashboards have been evaluated in several user studies with different evaluation methods, including interviews, focus groups, questionnaires, analysis of actual behaviour, and ethnographic field studies. The number of students and teachers participating in these studies range from six users for smaller studies to 43 users for larger evaluations. In this section, we present lessons learnt and guidelines from these user studies.

### 3.1 How Should We Visualise Relevant Data to Facilitate Students Exploring the Path from Effort to Outcomes? (RQ1)

**Abstract the LA Data:** LA data can be visualised in multiple ways [15,41], depending on the audience and desired message. LA prediction systems create notifications and visualisations to warn users and impact retention [1,16], while structural and content analysis help teachers gain insights at higher levels [31]. The data can also be abstracted or aggregated, providing students with awareness of efforts [29] and outcomes [35]. There are many ways of dealing with the abundance of LA data, so that both teachers and students can make sense of it. Overview approaches are a good basis for facilitating further and deeper exploration of the LA data.

With dashboard A and B [37], the abstract overview through the use of badges (see Fig. 1) had more impact on student motivation than our previous aggregate version that visualised the data through tables and numbers [37]. The badges still sufficed for the teacher to intervene or start a discussion in the classroom by projecting dashboard A on the wall. An interactive tabletop dashboard B visualising the reward relationship between students and badges served as enough incentive for students to actively explore and discuss their achievements with peers [7].

*Guideline: Aggregating or abstracting the information can help create progress awareness towards specific learning outcomes. These "overview" presentations of the learner traces can serve as a first incentive to trigger students into further LA data exploration.*

**Provide Access to the Learner Artefacts:** By limiting dashboard visualisation to an abstracted overview, teachers and students need to access the original, external environment in which the activities occur to gain further insights (e.g. the on-line learning environment, the individual blog posts). By doing so, the user loses the advantages of the LA overview, and it becomes more difficult to link effort to learning outcome (e.g. which blog posts resulted in a badge). During dashboard B's evaluation, students could still reflect on their personal progress through memory recall, but when trying to make sense of peer data, the lack of access to the blog posts inhibited further discussion. By adding artefacts directly to the LA dashboard, we can retain the connection between effort and outcome.

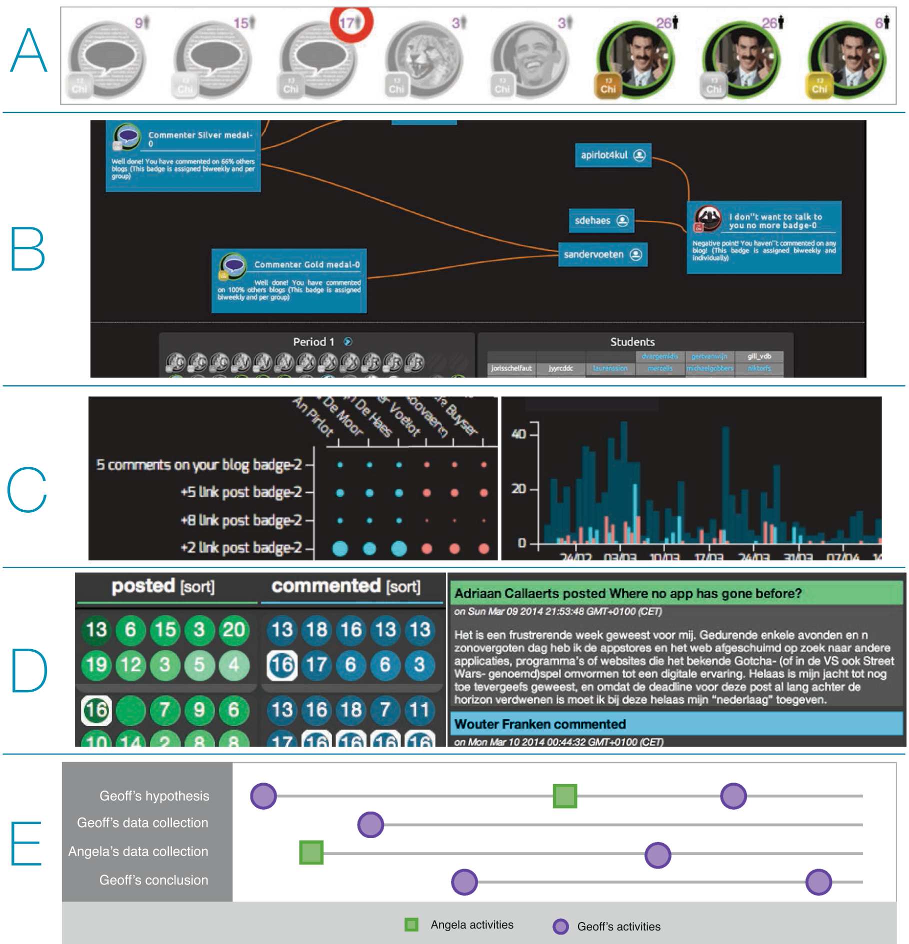



**Fig. 1.** **A.** Navi Badgeboard: visualising course goals through badges. **B.** Navi Surface: collaborative exploration of LA data. **C.** Two Class View modules: comparing two groups of students (red and blue) through the student-badge matrix and the total activity per day graph. **D.** LARAe: integration of LA with student workflow. **E.** LARAe.TT: collaborative exploration of the learner paths. (Color figure online)

The visual information-seeking mantra of "Overview first, zoom and filter, then details-on-demand" [39] is the basis used in dashboard C, D and E: the abstraction layer becomes a gateway to further exploration of the LA data (see Fig. 2). Teachers and students reported this functionality to be valuable: further exploration in the learner artefacts makes the LA dashboards applicable for e.g. evaluations with the student, or finding relevant learner artefact examples of peers for self-improvement.

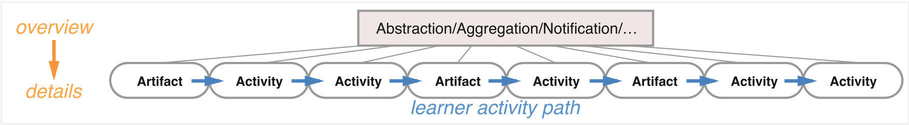



**Fig. 2.** Facilitating exploration of the abundance of learning traces and student learning paths through overview to details and facilitating learning path exploration.

Few examples of LDs provide access to the learner artefacts. Fulantelli et al. [14] support LA visualisations with direct access to the artefacts, but use is limited to teachers. When artefacts are made available to students, the selection is usually made for them: Shum et al. [42] automatically filter the large amount of assets to provide students with relevant resources. Bull et al. [5] provide assessment feedback to the student that can be linked with artifacts as evidence.

Many LA systems already store the learner artefacts [11,14,30], but limit its access to teachers [14]. We believe it is important for future dashboards to make personal and peer artefacts also available to students, as shown by our studies of dashboards B, C, D and E [6–9,35]. The findings are confirmed in studies of [19,30].

*Guideline: To empower the student and promote exploration of the effort to outcome path, LDs should allow manual exploration of the artefacts.*

**Augment the Abstracted Data:** Abstractions present the essentials, and thus lower the cognitive efforts required by students. Students could access peer's personal overviews in dashboard A, but rarely did so. However, the simplified, abstracted personal overview left room for the integration of peer information: every badge rewarded in the class was included to the personal overview, including the number of times each badge was awarded in class (see Fig. 1A). This was regarded as a valuable asset for students: they reported the presence of peer data in the personal overview helped position themselves among their peers and played an important motivational role.

In a blog-course setting, dashboard D [8] provides an overview of each blog post generated, and augments each data point (blog post) with the age of the blog post and number of comments the blog post has received (see Fig. 1C). This helped teachers and students find learner artefacts worthy of their attention: 55 % of students considered a high number and thus active thread as interesting, while 18 % reported they would avoid such threads. Teachers reported inactive threads were a sign for need of intervention. 7 % of students would use the numbers for self-assessment (e.g. low numbers on personal artefacts could indicate low quality). In the IBL setup, learner artefacts can be rated by teachers and peers. This information was visualised per artefact data point, providing a good overview of both the quantity and quality of learner outcomes per student, and helped peers in finding valuable (highly rated by peer or teacher) hypotheses, conclusions, discussions.

Extra meta-data regarding the LA traces can serve as indicators to guide users to relevant information, without forcing a predefined decision. Huang et al. [30] use location, time and peer information as a way for students to find relevant data. Doug and Makryannis [10] suggest reputation meta-data to support judgement on quality of artefacts. By leveraging meta-data to extend simple dashboards, students can be exposed to peer information without much user effort (e.g. class badge rewards of dashboard A). Interpretable indicators (e.g. social activity count in dashboard D) can help explore and find relevant artefacts.

*Guideline: While abstraction can help tackle the abundance of learner traces, augmentation approaches should be taken into account to help improve judgement of quality and exploration of the abundance of LA data.*

**Provide Access to Teacher and Peer Feedback:** For teachers, it is important to remain up to date with student efforts and outcomes, but also to provide students with timely feedback [2,32]. Providing public access to teacher feedback was well received by both students and teachers. As mentioned above, visualising ratings of the IBL learner artefacts provides teachers with a clear view of the quality of the student contributions. Students can use these ratings as guides to find quality example artefacts to learn from.

In the blog-supported courses, feedback is given through blog comments. Dashboard D helped students quickly access all teacher feedback across the entire course. Students reported that having access to teacher feedback given to peers helped them to "be ahead of the game". While the important feedback is usually repeated in face-to-face sessions, students mentioned "by then it might be too late". Teachers, when working with multiple colleagues on the same course, reported the feedback visualisation helps keep track colleague activity, resulting in a better feedback consistency, and preventing redundant feedback.

LA-supported feedback is often related to EDM systems, where informative and explanatory notifications and visualisations attempt to change student behaviour [1,16]. Clear evidence of dashboards that help teachers intervene when necessary, is provided in [9,26]. Bull et al. [5] successfully use artefacts as evidence for assessment feedback. Our evaluation participants showed interest in using the dashboard to support evaluation. But as shown in [28], incorporating teacher feedback into the LA traces can play an important role as well.

*Guideline: Make teacher feedback and feedback given by peers accessible in LDs and incorporate (intermediate) assessment data to raise awareness and to support reflection.*

**Visualise the Learner Path:** Until now, we have explored the vertical path of overview to details: abstraction as a way to facilitate teacher and student to drill down and explore the abundance of learner traces. A quality learner artefact does not necessarily indicate a good understanding of the matter, and only provides a narrow view of the student's process [3]. We define the *learner path* as the sequence of student activities and artefacts: An artefact created and the activity that happens on an artefact (e.g. a rating, a comment) can impact the next one: a comment by a peer can influence the next blog post, the creation of a mind-map might result in a new hypothesis.

While the vertical path from overview to details can help navigate the LA data, this horizontal learner path (see Fig. 2) can help provide deeper insights into students' learning [13]. We have explored this concept in dashboard E [6], where we visualise the sequence of an entire class across multiple activity types (see Fig. 1E). Teachers reported that visualising this path can help students backtrack through their IBL process, reflect, and make sense of it. But it can also assist students in exploring peer paths, to discover different approaches and improve their own methods: when discovering an interesting inquiry conclusion posted by a peer, both teacher and student can access and reflect on every learner activity that helped arrive at that specific solution.

*Guideline: LD design should try to give insight into the learning path to support reflection, peer comparison and self-regulated learning.*

### 3.2 How Can We Promote Students, Inside and Outside the Classroom, to Actively Explore This Effort to Outcomes Path? (RQ2)

**Integrate the Dashboard into the Work-Flow:** During dashboard A's deployment, the Master in Engineering students reported that their high workload did not leave much room for LDs. Google Analytics logs showed that students would access the dashboard the evening before class. The successful dashboard features were those with low requirements on effort and time: a quick glance was sufficient to raise student awareness of personal and class progress [7].

With dashboard D, we attempted to integrate the dashboard into the student work-flow. As reading and commenting on peer blogs is part of the course activities, dashboard D [8] provides direct access to the learner artefacts (blog posts), teacher and peer feedback, and augments the data with blog post age and activity to help students navigate. Simply put, the dashboard replicates RSS[4] reader functionality, but leverages LA data to facilitate richer exploration to provide further insights. Dashboard D was used by 55 % of the blog-supported course students on a weekly basis. During the IBL pilots, dashboard D was reported to be used in the classroom for weekly coaching tasks, while it also became part of the student's time management tool set.

Kapros et al. [20] integrated LA visualisations into an LMS and empowered learning and development managers by providing context next to LA visualisations. But this LA contextualisation can also benefit students. For example, Course Signals' traffic light representation of the chance of success was successfully integrated and accepted into the student's course homepage [1]. Dashboard D leveraged LA to support students' learning activities (e.g. finding, reading and reacting to relevant posts, accessing feedback), improving not only their

[4] https://en.wikipedia.org/wiki/RSS.

work-flow, but also exposing them to LA data more often. Wise [47] identified a similar need for better integration into existing work-flows.

*Guideline: It is important to incorporate LD use in the work-flow of students and to tailor LDs depending on the context in which learning occurs* [17]. *Therefore, while designing dashboards, keeping in mind the specific user needs, the course setting, and the target location and technologies available results in a better user acceptance, which in turn can help raise usage and improve impact.*

**Facilitate Collaborative Exploration of LA Data:** Dashboard A was developed as a desktop application, but was several times projected on a wall in the classroom when the teacher deemed intervention necessary. Problematic students would be highlighted, and the students would get the opportunity to explain their (lack of) activity. In this situation, the teacher drives the visualisation and students can contribute to the discussion. However, students cannot interact with the visualisation directly, only through the teacher.

Leveraging the affordances of large interactive tables, we can facilitate collaborative sense-making [18] as students and teachers can simultaneously interact with and explore the LA traces. To the best of our knowledge, no examples exist of LA data visualised on such devices.

Dashboard B limited the visualisation to badges. This abstract view of the data was sufficient to trigger exploration and discussion, but only happened when students grouped around the tabletop. They would reflect on their own and peer achievements, and come up with arguments for their lack of achieving certain badges. However, students who approached the tabletop by themselves were not motivated to explore the LA data. Students interacting in group experienced the system as "fun", and reported they would like to use it together with teachers.

Dashboard E visualises an overview of the class' *learner paths* and *learner artefacts*. The collaborative aspect was well-received and resulted in many scenarios teachers considered interesting: a teacher can initiate a discussion and ask students around the tabletop to explain their reasoning. Teachers can use other students examples to inspire struggling students. Participants also mentioned that it can help students self-support their learning activities without teacher intervention: a student can explore peer activities and find "peer experts" on specific topics the student struggles with.

While many LDs visualise social and group interactions [26,34], few dashboards are created with collaborative sense-making of the LA data in mind. Yet, dashboard B and E showed great potential for discussion, exploration, sense-making, and assessment. Even dashboard A triggered group discussions when projected in the classroom.

*Guideline: While designing LDs, keep in mind that collaborative exploration can support discussion, exploration, and assessment, and can enhance reflection and awareness. Existing research in the fields of Collaborative Visualisation* [18] *and Computer-Supported Cooperative Learning* [43] *offers a promising approach to improve collaborative exploration of LA data.*

## 4 Conclusion

The intent of this paper was to formulate the lessons learnt that the authors consider important for future development of LA dashboards. In this paper we have outlined guidelines on how to visualise relevant data (RQ1) and how to promote active exploration by students (RQ2) based on results of our user studies. We believe that it is highly important to empower students to reason about their efforts and outcomes. This paper therefore discussed how to create dashboards that support students in actively exploring their efforts and outcomes: by providing data beyond personal analytics, through visualisation techniques to make the abundance of data accessible, multi-user interaction to facilitate collaborative sense-making, and integration of dashboards into student workflow.

The guidelines are derived from a series of user studies with five LDs, but are based on first indicators only. Nevertheless we believe they present important steps towards the design of LDs that support important needs of students and teachers. We will explore how to improve on our current designs, evaluate further these choices, and deploy in other classroom settings to validate our findings.

**Acknowledgements.** The research leading to these results has received funding from the European Community's 7th Framework Programme (FP7/2007–2013) under grant agreement No. 318499 - weSPOT project and the Erasmus+ programme, Key Action 2 Strategic Partnerships, of the European Union under grant agreement 2015-1-UK01-KA203-013767 ABLE project.

## References

1. Arnold, K.E., Pistilli, M.D.: Course signals at purdue: using LA to increase student success. In: Proceedings of LAK 2012, pp. 267–270. ACM, New York (2012)
2. Black, P., Wiliam, D.: Assessment and classroom learning. Assess. Educ. Principle Policy Pract. **5**(1), 7–74 (1998)
3. Brennan, K., Resnick, M.: New frameworks for studying and assessing the development of computational thinking. In: Proceedings of the Annual Meeting of the American Educational Research Association (2012)
4. Brusilovsky, P.: Adaptive hypermedia: from intelligent tutoring systems to web-based education. In: Gauthier, G., VanLehn, K., Frasson, C. (eds.) ITS 2000. LNCS, vol. 1839, pp. 1–7. Springer, Heidelberg (2000)
5. Bull, S., Johnson, M.D., Alotaibi, M., Byrne, W., Cierniak, G.: Visualising multiple data sources in an independent open learner model. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS, vol. 7926, pp. 199–208. Springer, Heidelberg (2013)
6. Charleer, S., Klerkx, J., Duval, E.: Exploring inquiry-based learning analytics through interactive surfaces. In: Proceedings of 1st International Workshop on Visual Aspects of Learning Analytics. CEUR-WS (2015)
7. Charleer, S., Klerkx, J., Santos, J.L., Duval, E.: Improving awareness and reflection through collaborative, interactive visualizations of badges. In: CEUR Workshop Proceeding of ARTEL@EC-TEL, vol. 1103, pp. 69–81. CEUR-WS.org (2013)

8. Charleer, S., Odriozola, S., Luis, J., Klerkx, J., Duval, E.: Larae: learning analytics reflection and awareness environment. In: CEUR Workshop Proceedings, vol. 1238, pp. 85–87. CEUR-WS (2014)
9. Charleer, S., Santos, J.L., Klerkx, J., Duval, E.: Improving teacher awareness through activity, badge and content visualizations. In: Cao, Y., Väljataga, T., Tang, J.K.T., Leung, H., Laanpere, M. (eds.) ICWL 2014 Workshops. LNCS, vol. 8699, pp. 143–152. Springer, Heidelberg (2014)
10. Clow, D., Makriyannis, E.: iSpot analysed: participatory learning and reputation. In: Proceedings of LAK 2011, pp. 34–43. ACM, New York (2011)
11. Conde, M.A., Garcia-Penalvo, F.J., Gomez-Aguilar, D.A., Theron, R.: Visual learning analytics techniques applied in software engineering subjects. In: Frontiers in Education Conference (FIE 2014), pp. 1–9. IEEE (2014)
12. Dillenbourg, P., Zufferey, G., Alavi, H.S., Jermann, P., Do, L.H.S., Bonnard, Q., Cuendet, S., Kaplan, F.: Classroom orchestration: the third circle of usability. In: CSCL 2011 Conference Proceedings, vol. 1, pp. 510–517. International Society of the Learning Sciences (2011)
13. Fields, D.A., Quirke, L., Amely, J., Maughan, J.: Combining big data and thickdata analyses for understanding youth learning trajectories in a summer coding camp. In: Proceedings of SIGCSE 2016, pp. 150–155. ACM, New York (2016)
14. Fulantelli, G., Taibi, D., Arrigo, M.: A framework to support educational decision making in mobile learning. Comput. Hum. Behav. **47**, 50–59 (2015)
15. Govaerts, S., Verbert, K., Duval, E., Pardo, A.: The student activity meter for awareness and self-reflection. In: Proceedings of ACM Annual Conference Extended Abstracts on Human Factors in Computing Systems, pp. 869–884. ACM (2012)
16. Hu, Y.-H., Lo, C.-L., Shih, S.-P.: Developing early warning systems to predict students online learning performance. Comput. Hum. Behav. **36**, 469–478 (2014)
17. Huang, D., Tory, M., Aseniero, B.A., Bartram, L., Bateman, S., Carpendale, S., Tang, A., Woodbury, R.: Personal visualization and personal visual analytics. IEEE Trans. Visual. Comput. Graph. **21**(3), 420–433 (2015)
18. Isenberg, P., Elmqvist, N., Scholtz, J., Cernea, D., Ma, K.-L., Hagen, H.: Collaborative visualization: definition, challenges, and research agenda. Inf. Vis. **10**(4), 310–326 (2011)
19. Ji, M., Michel, C., Lavoué, É., George, S.: DDART, a dynamic dashboard for collection, analysis and visualization of activity and reporting traces. In: Rensing, C., de Freitas, S., Ley, T., Muñoz-Merino, P.J. (eds.) EC-TEL 2014. LNCS, vol. 8719, pp. 440–445. Springer, Heidelberg (2014)
20. Kapros, E., Peirce, N.: Empowering L&D managers through customisation of inline learning analytics. In: Zaphiris, P., Ioannou, A. (eds.) LCT 2014. LNCS, vol. 8523, pp. 282–291. Springer, Heidelberg (2012)
21. Klamma, R.: Emerging research topics in social learning. In: Proceedings of the 7th International Conference on Networked Learning, vol. 2010, pp. 224–231 (2010)
22. Klerkx, J., Verbert, K., Duval, E.: Learning analytics dashboards. In: Lang, C., Siemens, G. (eds.) Handbook of Learning Analytics & Educational Data Mining (Accepted 2016)
23. Lin, W.-J., Liu, Y.-L., Kakusho, K., Yueh, H.-P., Murakami, M., Minoh, M.: Blog as a tool to develop e-learning experience in an international distance course. In: 6th International Conference on Advanced Learning Technologies, pp. 290–292. IEEE (2006)
24. Marques, A.M., Krejci, R., Siqueira, S.W., Pimentel, M., Braz, M.H.L.: Structuring the discourse on social networks for learning: case studies on blogs and microblogs. Comput. Hum. Behav. **29**(2), 395–400 (2013)

25. Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., Clayphan, A.: Latux: an iterative workflow for designing, validating and deploying learning analytics visualisations. J. Learn. Anal. **2**(3), 9–39 (2016)
26. Martinez-Maldonado, R., Yacef, K., Dimitriadis, Y., Edbauer, M., Kay, J.: MT-Classroom and MTDashboard: supporting analysis of teacher attention in an orchestrated multi-tabletop classroom. In: International Conference on CSCL 2013, pp. 119–128 (2013)
27. Mikroyannidis, A., Okada, A., Scott, P., Rusman, E., Specht, M., Stefanov, K., Boytchev, P., Protopsaltis, A., Held, P., Hetzner, S., Kikis-Papadakis, K., Chaimala, F.: weSPOT: a personal and social approach to inquiry-based learning. J. Univ. Comput. Sci. **19**(14), 2093–2111 (2013)
28. Monroy, C., Rangel, V.S., Whitaker, R.: Stemscopes: contextualizing learning analytics in a k-12 science curriculum. In: Proceedings of LAK 2013, pp. 210–219. ACM (2013)
29. Nakahara, J., Hisamatsu, S., Yaegashi, K., Yamauchi, Y.: iTree: does the mobile phone encourage learners to be more involved in collaborative learning? In: Proceedings of CSCL 2005, pp. 470–478. International Society of the Learning Sciences (2005)
30. Ogata, H., Mouri, K.: Connecting dots for ubiquitous learning analytics. In: Cheung, S.K.S., Kwok, L.-F., Yang, H., Fong, J., Kwan, R. (eds.) ICHL 2015. LNCS, vol. 9167, pp. 46–56. Springer, Heidelberg (2015)
31. Pham, M.C., Derntl, M., Cao, Y., Klamma, R.: Learning analytics for learning blogospheres. In: Popescu, E., Li, Q., Klamma, R., Leung, H., Specht, M. (eds.) ICWL 2012. LNCS, vol. 7558, pp. 258–267. Springer, Heidelberg (2012)
32. Poulos, A., Mahony, M.J.: Effectiveness of feedback: the students perspective. Assess. Eval. High. Educ. **33**(2), 143–154 (2008)
33. Protopsaltis, A., Seitlinger, P., Chaimala, F., Firssova, O., Hetzner, S., Kikis-Papadakis, K., Boytchev, P.: Working environment with social and personal open tools for inquiry based learning: pedagogic and diagnostic frameworks. Int. J. Sci. Math. Technol. Learn. **20**(4), 51–63 (2014)
34. Rivera-Pelayo, V., Munk, J., Zacharias, V., Braun, S.: Live interest meter: learning from quantified feedback in mass lectures. In: Proceedings of LAK 2013, pp. 23–27. ACM, New York (2013)
35. Santos, J.L., Charleer, S., Parra, G., Klerkx, J., Duval, E., Verbert, K.: Evaluating the use of open badges in an open learning environment. In: Hernández-Leo, D., Ley, T., Klamma, R., Harrer, A. (eds.) EC-TEL 2013. LNCS, vol. 8095, pp. 314–327. Springer, Heidelberg (2013)
36. Santos, J.L., Govaerts, S., Verbert, K., Duval, E.: Goal-oriented visualizations of activity tracking: a case study with engineering students. In: Proceedings of LAK 2012, pp. 143–152. ACM, New York (2012)
37. Santos, J.L., Verbert, K., Govaerts, S., Duval, E.: Addressing learner issues with StepUp!: an evaluation. In: Proceedings of LAK 2013, pp. 14–22. ACM (2013)
38. Santos, J.L., Verbert, K., Klerkx, J., Ternier, S., Charleer, S., Duval, E.: Tracking data in open learning environments. J. Univ. Comput. Sci. **21**(7), 976–996 (2015)
39. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: IEEE Symposium on Visual Languages, pp. 336–343. IEEE (1996)
40. Siemens, G., Long, P.: Penetrating the Fog: Analytics in Learning and Education, vol. 46, pp. 30–32. EDUCAUSE, Boulder (2011)

41. Silius, K., Tervakari, A.-M., Kailanto, M.: Visualizations of user data in a social media enhanced web-based environment in higher education. In: Global Engineering Education Conference, pp. 893–899. IEEE (2013)
42. Buckingham Shum, S., Ferguson, R.: Social learning analytics. J. Educ. Technol. Soc. **15**(3), 3–26 (2012)
43. Suthers, D.D.: Computer-supported collaborative learning. In: Seel, N.M. (ed.) Encyclopedia of the Sciences of Learning, pp. 719–722. Springer, Heidelberg (2012)
44. Verbert, K., Duval, E., Klerkx, J., Govaerts, S., Santos, J.L.: Learning analytics dashboard applications. Am. Behav. Sci. **57**(10), 1500–1509 (2013)
45. Verbert, K., Govaerts, S., Duval, E., Santos, J., Van Assche, F., Parra, G., Klerkx, J.: Learning dashboards: an overview and future research opportunities. Pers. Ubiquit. Comput. **18**(6), 1499–1514 (2014)
46. Waters, S.: The Current State of Educational Blogging (2016). https://www.theedublogger.com/2016/01/15/educational-blogging-2015/. Accessed 29 Mar 2016
47. Wise, A.F.: Designing pedagogical interventions to support student use of learning analytics. In: Proceedings of LAK 2014, pp. 203–211. ACM (2014)

# Retrieval Practice and Study Planning in MOOCs: Exploring Classroom-Based Self-regulated Learning Strategies at Scale

Dan Davis[1(✉)], Guanliang Chen[1], Tim van der Zee[2], Claudia Hauff[1], and Geert-Jan Houben[1]

[1] Web Information Systems, Delft University of Technology, Delft, The Netherlands
{d.davis,guanliang.chen,c.hauff,g.j.p.m.houben}@tudelft.nl

[2] Graduate School of Teaching (ICLON), Leiden University, Leiden, The Netherlands
t.van.der.zee@iclon.leidenuniv.nl

**Abstract.** Massive Open Online Courses (MOOCs) are successful in *delivering* educational resources to the masses, however, the current retention rates—well below 10 %—indicate that they fall short in helping their audience become effective MOOC learners. In this paper, we report two MOOC studies we conducted in order to test the effectiveness of pedagogical strategies found to be beneficial in the traditional classroom setting: *retrieval practice* (i.e. strengthening course knowledge through actively recalling information) and *study planning* (elaborating on weekly study plans). In contrast to the classroom-based results, we do not confirm our hypothesis, that small changes to the standard MOOC design can teach MOOC learners valuable self-regulated learning strategies.

**Keywords:** MOOC · Self-regulated learning

## 1 Introduction

Open, informal learning environments, such as MOOCs, provide learners with an unprecedented level of autonomy in the learning process. While certainly empowering in one sense, this new paradigm also places the onus on the individual learner to both conceive and follow a learning process on their own.

Given that one target audience of MOOCs is disadvantaged people without experience in or access to formal educational settings [6], one cannot assume that all learners have the skill set to independently direct their own learning process. Moreover, current MOOCs are frequently designed without any of these

This work is co-funded by the Erasmus+ Programme of the European Union. Project: STELA 62167-EPP-1-2015-BE-EPPKA3-PI-FORWARD.

D. Davis and T. van der Zee—Research is supported by the *Leiden-Delft-Erasmus Centre for Education and Learning*.

G. Chen—Research is supported by the *Extension School* of the Delft University of Technology.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 57–71, 2016.
DOI: 10.1007/978-3-319-45153-4_5

considerations; they simply deliver the content to the learner without concern for fostering effective learning strategies.

The analysis of learning strategies (their benefits and effectiveness) has been a long-standing focus of classroom-based learning research. Some of the learning strategies most popular with learners, such as note-taking and rereading, have been found to be outperformed by what is known as *retrieval practice* (or the *testing effect*) [2,8]: a study strategy which focuses on the active recalling of information from memory (as opposed to rereading a passage or rewatching a video), which has a substantial, positive effect on future recall attempts [16].

A second study strategy found to be particularly effective in classroom-based learning is that of *study planning*. Research in this area has found students who spend time thinking about, explicitly stating, and reflecting on their goals on a daily, weekly, or even yearly level show increases in both engagement and academic performance [13,18,21].

Both retrieval practice and study planning are aspects of *Self-Regulated Learning* (SRL). SRL is defined as a learner's proactive engagement with the learning process through various personal management strategies in order to control & monitor cognitive and behavioral processes towards a learning outcome [20,22]. By making learners more adept at regulating their own learning process, MOOCs can also act as resources for not only domain-specific knowledge, but also as a tool for teaching people how to learn.

In this paper we investigate to what extent SRL strategies beneficial in the classroom can be successfully transferred to the MOOC setting. We implemented retrieval practice and study planning prompts aimed at promoting SRL in two edX MOOCs. Our work is guided by the following **R**esearch **Q**uestions:

**RQ1** Do learners engage with SRL interventions as much as they do with course content (videos, quizzes, etc.)?
**RQ2** Does inserting retrieval cues after MOOC lecture videos increase test performance?
**RQ3** Does providing a scaffolded means of study planning promote learner engagement and self-regulation?

Based on our experimental results, we conclude that such interventions are not beneficial enough to increase MOOC learners' success (in terms of grades earned) or engagement (in terms of activity levels observed in the course environment).

## 2 Related Work

In this section, we separately explore previous work in classroom-based and MOOC-based SRL interventions.

### 2.1 Classroom-Based SRL

**Retrieval Practice.** A study in [1] focused on metacognition, or the ability to regulate one's own cognition, such as learning processes. The study found that providing metacognitive prompts to the sample of undergraduate students resulted in improved learning outcomes. Similar to our work, [1] also observed high levels of non-compliance with the metacognitive prompts/interventions (instructional events intended to improve student learning performance [1]), thus raising the challenge of motivating students to engage with such activities.

On the topic of the "testing effect", in the context of video watching, Johnson and Mayer [8] found that, compared to only re-watching, students remember more about the content when they are asked to respond to questions about the video's content after viewing it. This lab study with 282 participants found high support for the testing effect, with subjects exposed to this condition showing higher rates of both learning transfer and retention of knowledge a week after the lesson [8].

Roediger and Butler [16] offer a review of the existing literature on retrieval practice and outline five key findings: (i) retrieval practice increases long-term learning compared to passive re-visiting, (ii) repeated retrieval practice is more effective than a single instance, (iii) retrieval practice with feedback increases the testing effect, (iv) some lag between study and testing is required for retrieval practice to work, and (v) retrieval practice not only benefits a specific response/finite body of knowledge; it allows for transfer of knowledge to different contexts [16].

Research has also been done to determine how to best implement retrieval practice; with a study including fifty middle-school students, Davey and McBride [5] found that, compared to rereading, actively retrieving and elaborating on knowledge from memory leads to better long-term learning.

The most notable difference between these works and our research is the learner population. MOOCs have an unprecedented level of heterogeneity, with learners coming from all corners of the globe with profoundly diverse backgrounds.

**Study Planning.** The goal setting intervention by Schippers et al. [18] was introduced to an entire class of students in an undergraduate business school. This intervention, which required four hours of student engagement at the beginning of their program, had a positive impact across a prolonged period of time. The reported results include a 98 % reduction in the gender and a 38 % reduction in the ethnicity gap after one year (compared to the previous year's cohort of students).

Palmer and Wehmeyer [14] implemented the "Self-Determined Learning Model of Instruction" [21] to a sample of students ages six through ten and found that even students of this age range were able to both successfully learn and effectively practice self-determined goal setting strategies.

In the context of a high school social studies class Zimmerman et al. [23] found that students perform better (in this case measured by final grade) when they are able to set their own goals and benchmarks, rather than having to adapt to those imposed upon them by parents or teachers. The findings of [23] suggest that setting one's own goals works in tandem with increases in academic efficacy, thus improving performance.

Through a "self-monitoring" course intervention, Sagotsky et al. [17] found that (elementary-middle school) students who actively monitored and reflected on their learning progress and behavior on a regular basis exhibited higher academic achievement (grades) and "more appropriate study behavior" (such as being on-task and engaged) [17]. This self-monitoring group also performed significantly better in both measures (achievement and behavior) than the control group which was only prompted to *set* study goals and not to reflect on them.

In a study including 27 undergraduate students preparing for an exam, Mahoney et al. [11] exposed students to one of three interventions: (i) continuous self-monitoring, (ii) intermittent self-monitoring, and (iii) receiving instructor feedback. The results showed that students who engaged in self-monitoring (especially continuous) exhibited higher levels of engagement and better scores on quantitative problems in the exam.

As mentioned with the literature on retrieval practice, the MOOC learner population is infinitely variant. So while the above findings on study planning may hold true in the classroom or laboratory (with required attendance and homogeneous samples), there is a chance that the findings do not translate directly to MOOCs.

### 2.2 Learning Strategy Research in MOOCs

MOOC-based research is beginning to recognize & address the current instructional design shortcomings of MOOCs. Nawrot and Doucet [12] suggest that MOOCs are in need of increased learner support, based on survey results which queried MOOC learners' experiences. They found time management in particular to be a major hindrance to learners. They propose to augment existing MOOC designs with time management support/guidance in order to curtail the dismal retention rates that MOOCs so frequently see.

By gathering information about MOOC learners' study habits through a post-course survey, Hood et al. [7] observed that learners coming from different professional backgrounds demonstrate different levels of SRL strategies—with higher-educated learners reporting higher levels of SRL. Our research aims to address this discrepancy in SRL skills and provide a scaffolded method to develop SRL strategies in our MOOC learners from all backgrounds and contexts. Instead of self-reported engagement data, however, our study views SRL and engagement in terms of log traces generated by the learning platform.

In one computer science MOOC, Kizilcec et al. [10] tested the effectiveness of sending out different types of encouraging emails to students and found them to be ineffective in increasing learner engagement with the course discussion forum. In a pre-course intervention Kizilcec et al. [9], introduced a subset of MOOC learners to a SRL support module in which seven SRL strategies were explained. Included as part of the pre-course survey, the study found this intervention to elicit no significant differences in course engagement or persistence (in terms of the number of lectures watched). As a consequence the authors proposed that such recommendations/interventions should be more robustly implemented into the structure of the MOOC. In our research we expand upon this and provide a venue (study planning advice & text input) for students to actively plan their learning strategies for the week.

In both [3,19] increasing learners' engagement with MOOC discussion forum was targeted in order to increase the overall retention rate. Coetzee et al. [3] introduced a voting/reputation system which allowed learners to vote on which posts are more or less valuable. Their main findings were that (i) the reputation system increases the response time and number of responses in the forum and (ii) forum activity is positively correlated with higher retention and final grades as compared to the learners who were exposed to the standard discussion forum design. The experiment by Tomkin and Charlevoix [19] aimed to discover the effect of having the course team (instructor & teaching assistants) engage with the forum. For one condition, the course team did not engage at all, and for the other, they were highly engaged, providing feedback to questions, comments, and compiling weekly summaries of the key discussion points. In contrast to the formulated hypotheses, the course team intervention resulted in no significant impact on completion rates, learner engagement, or course satisfaction.

To conclude, existing MOOC research has, so far, focused largely on *observing* the learning strategies employed by learners (without actively intervening), and a small-but-growing number of studies have tried to actively intervene and encourage SRL. Our research aims to expand on this existing work by designing and testing SRL interventions in MOOCs based on a theoretical foundation of teaching strategies found to be effective in traditional classroom settings.

## 3 Approach

In this section, we first describe the research hypotheses we developed based on **RQ1**, **RQ2**, and **RQ3**. Since our interventions were designed for two specific MOOCs, we first introduce them before outlining our implementation of the two interventions (retrieval practice and study planning).

### 3.1 Research Hypotheses

Regarding **RQ1**, and taking into consideration prior attempts at learning strategy research in MOOCs, we draw the following hypothesis:

**H1** *Learners do not engage with the SRL interventions as much as they engage with the main course content, such as videos and quizzes* [1,9].

Based on prior work in the area of retrieval practice we developed the following hypotheses related to **RQ2**:

**H2** *Actively retrieving/producing knowledge leads to better exam scores than passive rereading* [4,5,16].

With respect to **RQ3**, we draw the following two hypotheses from the existing literature on study planning:

**H3** *Encouraging learners to actively plan and reflect on their study habits will increase their engagement with the course* [11,17].

**H4** *Learners who actually plan and reflect on their course of study will exhibit higher engagement and achieve higher grades* [17,23].

### 3.2 MOOCs Studied

We implemented our interventions in two edX[1] MOOCs (Table 1), which were developed at the Delft University of Technology and ran in 2015. Although the choice of MOOCs was opportunistic, we consider them to be representative of a wide variety of MOOCs offered on platforms such as edX.

We deployed the retrieval practice intervention in `Functional Programming`, a 13-week MOOC which introduces basic functional programming language constructs. Nearly 28,000 learners enrolled, and 5 % eventually passed the course. The effectiveness of study planning was evaluated in `Industrial Biotechnology`, a 7-week MOOC that introduced learners to basic biotechnology concepts. Enrollment into this MOOC was lower, while the pass rate was similar to `Functional Programming`.

**Table 1.** Overview of the two courses included in the present study.

| Course | #Enrolled | Pass rate | #Learners in cohorts | #Cohorts |
|---|---|---|---|---|
| `Functional Programming` | 27,884 | 5.05 % | 9,836 | 3 |
| `Industrial Biotechnology` | 11,042 | 4.08 % | 1,963 | 2 |

On the edX platform, A/B testing (i.e. providing a different view of a MOOC to a randomly chosen subset of learners) is readily available. Upon enrolling, learners are randomly assigned into one of the provided `Cohorts`, which is either the control group (no intervention) or one of the experimental groups (an intervention of some form). One practical limitation of edX's `Cohorts` feature is that learners cannot be assigned retroactively to a group—any learner who registered to a MOOC *before* the `Cohorts` feature was activated will not be assigned to any group. This aspect is reflected in Table 1: 9, 836 (or 35 %) of the `Functional Programming` learners and 1, 963 (or 18 %) of the `Industrial Biotechnology` learners are assigned to either the control or one of the experimental groups. Although in our analysis we could have considered all non-assigned learners as part of the control group (as those learners were not exposed to any intervention), we opted not to do so to keep the groups balanced.

### 3.3 Retrieval Practice

In the original course design (i.e. no intervention) of `Functional Programming`, each week's video lecture is broken up into two or three segments. And although

[1] https://www.edx.org/.

the students must navigate themselves from one segment to the next, there are no other learning materials or activities between. In order to activate the learning process, we inserted retrieval practice cues designed to make learners stop and process the information presented in the video lecture.

In each course week, we inserted a retrieval cue directly after the final lecture video, thus prompting the learners to stop and think before moving on to the weekly quiz. The only exception to this design was one particular course week[2] where we inserted retrieval practice cues after each of the three segments of the weekly lecture, as in the previous edition of the course learners had perceived that week's material as the most challenging.

This experiment had three groups (or conditions): (1) the control group without an intervention, (2) the "cued" group, and (3) the "given" group. The "cued" group was shown the following prompt along with a blank text input box:

> *Please respond in 3–5 sentences to the following question: "In your opinion, what are the most important points from the previous video?"*

Note that these responses were not seen, graded, or given any feedback from the instructor—serving strictly as an activity for learners to exercise and improve memory recall. The "given" group, instead of being asked to create a summary themselves, was provided with a 3–5 sentence summary of the video as generated by one of the authors highly familiar with the functional programming paradigm. We included the "given" group in our study to determine the effect of actively retrieving information from memory versus being provided a summarizing text for passive reading.

### 3.4 Study Planning

The study conducted within `Industrial Biotechnology` introduced learners to SRL techniques by inserting a study planning module in each week's *Introduction* section. In order to stimulate learners to actively think about and plan their course of study and learning objectives for the week, we inserted the following question &examples followed by a blank text input box:

> *In the space below, please describe, in detail, your study plan and desired learning objectives for the week regarding your progress:*
> *e.g.*
> – *I plan to watch all of the lecture videos.*
> – *I will write down questions I have about the videos or assignments and discuss them in the forum.*

The initial prompts were bookended by a reflection prompt at the end of each week in which learners were instructed to reflect on their planning and execution:

[2] "Week 7: Functional Parsers and Monads".

*How closely did you follow your study plan from the beginning of the week? Did you successfully meet all of your learning objectives? In the space below, explain how you can improve upon your study habits in the following weeks in order to meet your goals.*

## 4 Findings

In this section, we describe our findings in line with our research hypotheses described in Sect. 3.1. Across both courses we find support for **H1** (learners engage less with interventions than course content items). Of the 3,262 learners in the "cued" condition in `Functional Programming`, 2,166 (66.4 %) logged at least one video-watching event in the course. Among these same learners only 719 (22 %) clicked on any of the retrieval practice interventions. Of the 998 learners exposed to the study planning modules in `Industrial Biotechnology`, 759 (76.1 %) logged at least one video-watching event. Among these same learners, only 147 (14.7 %) clicked on any of the study planning modules.

### 4.1 Retrieval Practice

We first tested whether the learners of the cued, given, and control groups score differently in the weekly quizzes. To this end we performed a MANOVA test with the highly engaged learners (characterized by having spent more than the group's mean time watching videos in Week 1 which is ≈22 min) in each of the three conditions as a fixed factor and the grades on the weekly quizzes as a dependent variable. The MANOVA test followed by the post hoc Games-Howell (equal variances not assumed) test yielded no significant differences between each group's weekly quiz grade.

In the previous analysis all highly engaged students from each condition were included. However, as many students did not engage with the intervention, this can give a distorted view of its effects. Therefore, we next isolated those learners

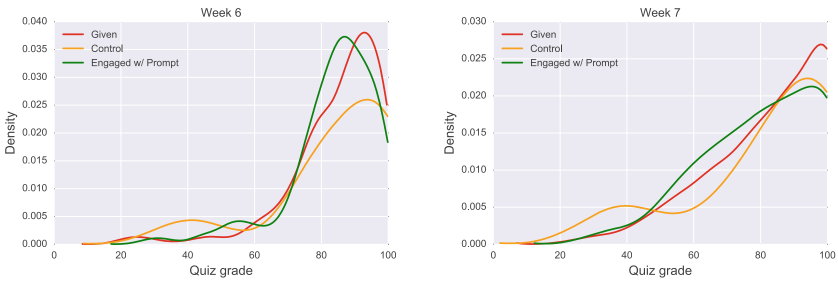



**Fig. 1.** KDE plot showing the distribution of weekly quiz grades across the groups of highly engaged learners. All lines were fit using a Gaussian kernel function. None of the differences between groups are statistically significant at the $\alpha = 0.01$ level.

who actively engaged (characterized by viewing the intervention for at least 10 s) with an intervention prompt at least once.

Using these new group definitions, we still observe no statistically significant differences between the groups as a result of a MANOVA (to test the difference between weekly quiz scores), and a one-way ANOVA (to test the difference between course final scores). In turn, we fail to reject the null hypothesis of **H2** in terms of both weekly quizzes and final course grades. Figures 1 and 2 illustrate these null findings via Kernel Density Estimation (KDE) plots.

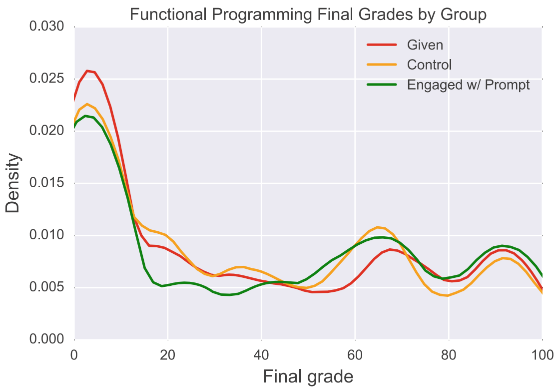



**Fig. 2.** KDE plot showing the distribution of final course grades across the groups of highly engaged learners. All lines were fit using a cosine kernel function. None of the differences between groups are statistically significant at the $\alpha = 0.01$ level.

### 4.2 Study Planning

We analyzed the differences between the two experimental groups in `Industrial Biotechnology`—those who were exposed to a study planning module intervention (condition) and those who were not (control)—and found no significant differences in their final grades, course persistence, and many engagement metrics, thus leading us to fail to reject the null hypothesis of **H3**. However, in support of **H4** at the 99 % confidence level, we do find the following statistically significant results when narrowing the sample to compare highly engaged learners (characterized by having spent more time watching Week 1 videos than the average learner, ≈33 min) in the control group and the learners in the condition group who engaged with a study planning module at least once (referred to as "Study Planners").

**Comparing Engagement Between Groups.** To determine whether there is a significant difference in the engagement levels between the highly engaged learners in the control group (N = 329) and the conditioned group (those who clicked on at least one study planning module, N = 146). In Table 2 we employ two Mann-Whitney $U$ tests, as the data is not normally distributed, showing that the study planners have a higher session count than the highly engaged learns ($U = 20{,}070$, $p = 0.003$), as well as a higher total amount of time spent in the course in hours ($U = 19{,}983$, $p = 0.002$).

The results suggest that students who engaged with the study planning intervention are significantly more engaged with other aspects of the course as well. An alternative interpretation, however, could be that students who are highly engaged with the course also tend to engage more with the planning intervention.

**Table 2.** Results of a Mann-Whitney $U$ test comparing the two conditions (study planners vs. highly engaged learners in the control group) in terms of two learner engagement metrics: total amount of logged sessions in the course and total amount of time spent in the course in hours.

| Variable | Study planning | Control |
|---|---|---|
| | Median | Median |
| Session count | 25.0 | 19.0 |
| Time in course (hours) | 18.6 | 13.1 |

**Comparing Course Persistence Between Groups.** With regard to **H4** (engagement-related) we operationalize learners' persistence as the corresponding week of a learner's latest quiz submitted or video watched (slightly different from that used in [9], where persistence measured the overall amount of course materials accessed). Whereas the analyses in Table 2 included activity throughout the entire course, irrespective of the course week, one symptom of SRL is a learners' persistence through the course, or how many weeks the learner makes it through. We define a learner's "Final Week Reached" as the latest week in the course in which the learner either watched a video or submitted a quiz question. We ran an ANOVA to compare how far into the course learners in each group reached.

The ANOVA yielded significant results, $F(2{,}734) = 21.66$, $p < 0.001$. Post hoc Games-Howell tests show that the group who engaged with the study planning module (N = 146, M = 4.60) persisted deeper into the course than highly engaged learners in the control group (N = 329, M = 3.84, $p < 0.001$) and highly engaged learners who were exposed to, but did not engage with, the study planning module (N = 262, M = 3.28, $p < 0.001$). Figure 3 presents a kernel density estimation plot in order to visualize the differences between groups.

**Comparing Final Grades Between Groups.** To answer the second aspect of **H4** (grade-related), we conducted an ANOVA to determine whether there was a significant difference in final grade between the three groups of highly engaged learners listed above. The univariate test was significant, $F(2{,}735) = 17.147$, $p < 0.001$. The results are presented in Table 3.

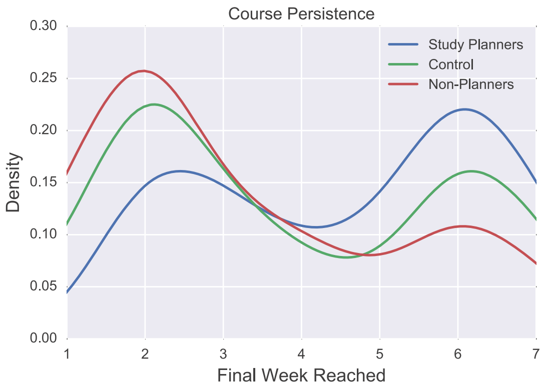



**Fig. 3.** KDE plot showing the course persistence of the three groups of learners. All lines were fit using a Gaussian kernel function.

**Table 3.** Results of the ANOVA comparing final course grades among learners who engaged with the study planning module (Mean = 46.42) against those of the two other groups. A final score of 100 would indicate a perfect score.

| | Group | Mean | Mean difference |
|---|---|---|---|
| | Study planners | 46.42 | — |
| i | Control | 36.44 | 9.98 |
| ii | Non-planners | 29.10 | 17.32 |

The follow-up Games-Howell test revealed that learners who engaged with the study planning module (M = 46.42) earned higher grades than the highly engaged learners in the control group (M = 36.44, $p = 0.003$) and highly engaged learners who did not engage with the intervention (Non-Planners, M = 29.10, $p < 0.001$). These results are visualized in Fig. 4 and illustrate how Study Planners' final grades are higher than the others'.

**Study Planners Engagement Correlations.** Focusing specifically on the learners who interacted with the study planning module intervention, we analyze the relationship between the extent to which they engaged with the intervention and their behavior elsewhere in the course. To do so, we computed a Pearson correlation coefficient to assess the relationship between a learner's average planning module response length (in text characters) and engagement-related variables such as: (i) total amount of time spent in the course, (ii) number of unique sessions logged, (iii) average length (in seconds) of learners' sessions, (iv) total amount of time spent watching videos, and (v) number of discussion forum sessions. The results are shown in Table 4. Two example correlations (unique sessions logged and time watching videos) are illustrated in the scatter plots in Fig. 5 to show the slope and overall fit of the regression line. Consistent

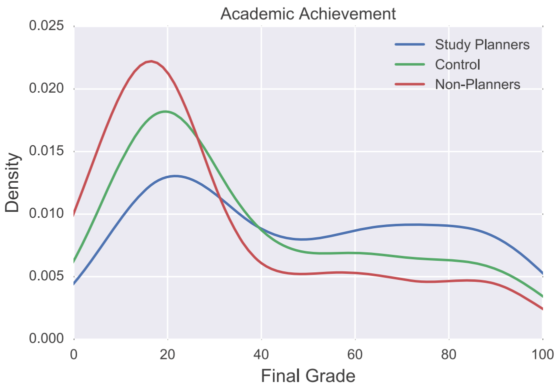



**Fig. 4.** KDE plot showing the distribution of final grades earned between the three groups of highly engaged learners. All lines were fit using a Gaussian kernel function.

with the Pearson correlation coefficients of 0.268 and 0.346, the plots indicate positive, small-to-moderate correlations.

**Table 4.** Pearson correlation coefficient test results reporting the relationship between learners' average planning module response length and five course engagement metrics. All correlations shown are significant at the $\alpha = 0.01$ level.

| Variable | Pearson correlation | N |
|---|---|---|
| Total time in course | 0.361 | 176 |
| Session count | 0.268 | 176 |
| Avg. session length | 0.346 | 176 |
| Time spent watching videos | 0.346 | 170 |
| Forum sessions | 0.305 | 154 |

The results suggest that increases in the amount of text learners write in the study planning module are correlated with small-to-moderate increases in a number of key course engagement metrics.

Overall, we find that mere *exposure* to study planning and retrieval practice interventions is not sufficient to significantly increase learner engagement or final grades. Only when narrowing the samples to learners who actually engaged with the study planning intervention do we see significant results. However, the same does not apply for learners who engaged with the retrieval practice cues, where even learners who engaged with the retrieval cues show no significant difference in any measure of performance.

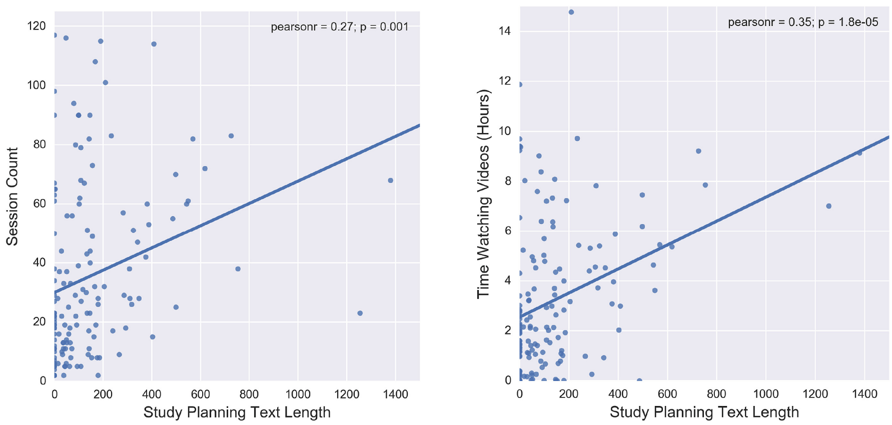



**Fig. 5.** Scatterplots illustrating two example results of the five Pearson correlation coefficient tests run in order to characterize the relationship between the amount of text characters entered in the study planning module and two key course engagement metrics: session count (left) and time spent watching video lectures (right).

## 5 Conclusions

In this work, we empirically investigated two types of instructional interventions found to be effective in traditional educational environments (study planning and retrieval practice) in the MOOC setting. In contrast to our hypotheses, we found both to be largely ineffective in boosting learner success and engagement. Only when accounting specifically for learners who engaged with the intervention directly do we observe significant increases in final grades and engagement in one of the two MOOCs studied. However, given that between 14 %–22 % of the learners meet this criteria in our studies, we too note the "non-compliance" described in [1] as a problem.

Another point of improvement for future studies is that of the frequency and chronology of the interventions. For example, future testing of retrieval practice should be made more continuous and incorporate more lag time [4,16].

Taking both the existing literature and the present study into account, we will design future theory-based (this research focuses on Zimmerman's [22] model, but future work should also investigate the effectiveness of other models, such as that of Pintrich [15]) interventions to be much more appealing and prominent in the context of the course—be it visually or perhaps making them compulsory.

## References

1. Bannert, M., Mengelkamp, C.: Scaffolding hypermedia learning through metacognitive prompts. In: Azevedo, R., Aleven, V. (eds.) International Handbook of Metacognition and Learning Technologies, pp. 171–186. Springer, Heidelberg (2013)

2. Carpenter, S.K., Pashler, H., Wixted, J.T., Vul, E.: The effects of tests on learning and forgetting. Mem. Cognit. **36**(2), 438–448 (2008)
3. Coetzee, D., Fox, A., Hearst, M.A., Hartmann, B.: Should your MOOC forum use a reputation system? In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 1176–1187. ACM (2014)
4. Cull, W.L., et al.: Untangling the benefits of multiple study opportunities and repeated testing for cued recall. Appl. Cognit. Psychol. **14**(3), 215–235 (2000)
5. Davey, B., McBride, S.: Generating self-questions after reading: a comprehension assist for elementary students. J. Educ. Res. **80**(1), 43–46 (1986)
6. Dillahunt, T.R., Wang, B.Z., Teasley, S.: Democratizing higher education: exploring MOOC use among those who cannot afford a formal education. Int. Rev. Res. Open Distrib. Learn. **15**(5), 177–196 (2014)
7. Hood, N., Littlejohn, A., Milligan, C.: Context counts: how learners' contexts influence learning in a MOOC. Comput. Educ. **91**, 83–91 (2015)
8. Johnson, C.I., Mayer, R.E.: A testing effect with multimedia learning. J. Educ. Psychol. **101**(3), 621 (2009)
9. Kizilcec, R.F., Pérez-Sanagustín, M., Maldonado, J.J.: Recommending self-regulated learning strategies does not improve performance in a MOOC. In: Proceedings of the Third ACM Conference on Learning @ Scale, L@S 2016 (2016)
10. Kizilcec, R.F., Schneider, E., Cohen, G., McFarland, D.: Encouraging forum participation in online courses with collectivist, individualist, and neutral motivational framings. eLearning Papers **37**, 13–22 (2014)
11. Mahoney, M.J., Moore, B.S., Wade, T.C., Moura, N.G.: Effects of continuous and intermittent self-monitoring on academic behavior. J. Consult. Clin. Psychol. **41**(1), 65 (1973)
12. Nawrot, I., Doucet, A.: Building engagement for MOOC students: introducing support for time management on online learning platforms. In: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, pp. 1077–1082. International World Wide Web Conferences Steering Committee (2014)
13. Nota, L., Soresi, S., Zimmerman, B.J.: Self-regulation and academic achievement and resilience: a longitudinal study. Int. J. Educ. Res. **41**(3), 198–215 (2004)
14. Palmer, S.B., Wehmeyer, M.L.: Promoting self-determination in early elementary school teaching self-regulated problem-solving and goal-setting skills. Remedial Special Educ. **24**(2), 115–126 (2003)
15. Pintrich, P.R.: The Role of Goal Orientation in Self-regulated Learning. Academic Press, San Diego (2000)
16. Roediger, H.L., Butler, A.C.: The critical role of retrieval practice in long-term retention. Trends Cogn. Sci. **15**(1), 20–27 (2011)
17. Sagotsky, G., Patterson, C.J., Lepper, M.R.: Training children's self-control: a field experiment in self-monitoring and goal-setting in the classroom. J. Exp. Child Psychol. **25**(2), 242–253 (1978)
18. Schippers, M.C., Scheepers, A.W., Peterson, J.B.: A scalable goal-setting intervention closes both the gender and ethnic minority achievement gap. Palgrave Commun. **1**, 1–12 (2015)
19. Tomkin, J.H., Charlevoix, D.: Do professors matter?: Using an A/B test to evaluate the impact of instructor involvement on MOOC student outcomes. In: Proceedings of the First ACM Conference on Learning @ Scale Conference, pp. 71–78. ACM (2014)
20. Vassallo, S.: Implications of institutionalizing self-regulated learning: an analysis from four sociological perspectives. Educ. Stud. **47**(1), 26–49 (2011)

21. Wehmeyer, M.L., Palmer, S.B., Agran, M., Mithaug, D.E., Martin, J.E.: Promoting causal agency: the self-determined learning model of instruction. Except. Child. **66**(4), 439–453 (2000)
22. Zimmerman, B.J.: A social cognitive view of self-regulated academic learning. J. Educ. Psychol. **81**(3), 329 (1989)
23. Zimmerman, B.J., Bandura, A., Martinez-Pons, M.: Self-motivation for academic attainment: the role of self-efficacy beliefs and personal goal setting. Am. Educ. Res. J. **29**(3), 663–676 (1992)

# "Keep Your Eyes on 'em all!": A Mobile Eye-Tracking Analysis of Teachers' Sensitivity to Students

Philippe Dessus[1(✉)], Olivier Cosnefroy[1,2], and Vanda Luengo[3]

[1] Univ. Grenoble Alpes, LSE (EA 602), 38000 Grenoble, France
philippe.dessus@univ-grenoble-alpes.fr
[2] DEPP, French Ministry of Education, Paris, France
olivier.cosnefroy@education.gouv.fr
[3] Sorbonne Universités, UPMC Univ Paris 06, LIP6 (UMR CNRS 7606), Paris, France
vanda.luengo@lip6.fr

**Abstract.** This study aims at investigating which cues teachers detect and process from their students during instruction. This information capturing process depends on teachers' sensitivity, or awareness, to students' needs, which has been recognized as crucial for classroom management. We recorded the gaze behaviors of two pre-service teachers and two experienced teachers during a whole math lesson in primary classrooms. Thanks to a simple Learning Analytics interface, the data analysis reports, firstly, which were the most often tracked students, in relation with their classroom behavior and performance; secondly, which relationships exist between teachers' attentional frequency distribution and lability, and the overall classroom climate they promote, measured by the Classroom Assessment Scoring System. Results show that participants' gaze patterns are mainly related to their experience. Learning Analytics use cases are eventually presented, enabling researchers or teacher trainers to further explore the eye-tracking data.

**Keywords:** Mobile eye-tracking · Learning analytics · Classroom supervision · Teacher information taking · Classroom observation system · Visualization techniques

## 1 Introduction

Maintaining some of the main variables of the classroom in adequate limits is one of the most crucial goals of every teacher, this activity being performed by continuous visual information takes. Teacher's situational awareness [1] is an important skill and is needed for supervising (i.e., taking information from the classroom environment) and controlling (i.e., acting on this environment in turn) the diverse events occurring in the classroom, often at fast pace. This skill has been shown to be directly related to learners' achievement [2].

Teachers' attentional resources are limited, so they cannot equally draw their attention to every event occurring during instruction, or on every learner. Two main concepts from the educational sciences literature have been derived from this


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 72–84, 2016.
DOI: 10.1007/978-3-319-45153-4_6

assumption. Firstly, the concept of "*withitness*" [3], which refers to the ability of teachers to proactively manage disruptive events, letting their students imagine that "they have eyes in the back of their heads". Secondly, the concept of *steering group* [4], which refers to a group of learners more or less consciously selected, and frequently supervised by the teacher in order to take on-the-fly instructional decisions. It is worth noting that the two concepts are hardly compatible with each other: a "withitness-able" teacher takes the classroom as a whole, whereas a "steering group"-focused teacher selects and a priori targets a small subset of students.

Some concerns have been raised on these two concepts. The operationalization of the "withitness" [5], as well as its empirical support [6], have been subject to difficulties. Whereas the very existence of a steering group is hardly debatable, the literature on this concept does not agree with the main features of this group. For instance, Lundgren [4] argued that the steering group is composed of students between the 10$^{th}$ and the 25$^{th}$ percentile of their cognitive abilities. Wanlin [7] reported two kinds of steering groups, comprehension-centered and behavior-centered, and showed that teachers mostly focus on medium and highly proficient students. Since these scholars did not have the same observation tool, we assume that a finer-grained observation tool may shed some light on the actual features of the steering group.

The main goal of this paper is to bring empirical support about the existence of either of these two concepts. We used a mobile eye-tracker to determine the continuous teacher's eye-fixation behavior during a whole lesson, accounting for their selective visual attention. We related this information to the cues (both behavioral and related to students' achievement) that lead a teacher to focus his or her interest on a given student. Novice and experienced teachers participated to this study in order to seek likely differences of behavior. Eventually, thanks to a Learning Analytics (LA) system, we will argue that we can unveil teacher–students interactional patterns during instruction, which in turn would be useful in some real-life contexts (use cases).

## 2 Eye-Tracking Devices and Teacher Decision Making

A well-established fact is that every teacher has to keep an overall awareness on the instructional situation [3]. However, the kind of cognitive processes undertaken to maintain this awareness has been studied so far mostly from verbalization procedures (either current or posterior to the activity), which are known to offer an incomplete access to the action and decision cognitive processes [8], because of their partly implicit nature.

Eye-tracking devices have become a reliable way to overcome this problem [9]. They enable the capture of eye fixations and saccades so that two pieces of information can be inferred [10]: which kind of information is extracted from a scene (static or dynamic); how much a scene is complex (the more complex a scene, the longer are the eye fixations). Moreover, the amount of gathered direct information is far larger than with other ways to observe teachers' behaviors, and makes possible LA-based procedures. All in all, they allow the processing of a large amount of "low inference" measures, which can be seen as more objective than measures that rely on the interpretation of a scarcer set of information.

Eye-tracking devices have seldom been used in educational contexts, but they have mainly been used in very constrained environments, like text reading or information seeking on screens. However, a few recent researches used eye-tracking devices for dynamic classroom scenes, either for analyzing student's gaze [11], teacher's cognitive load [12], or the whole classroom [13]. So far, two studies have investigated teachers' selective visual attention through the use of eye-tracking devices.

The study from van den Bogert et al. [14] analyzed teacher's (20 novices and 20 experts) fixations when viewing two videotaped lessons on a TV screen equipped with an eye-tracker. The expert–novice (E–N) paradigm predicted that, firstly, the fixations would be longer and more variable for novices (i.e., more complex) and, secondly, the number of targeted students would be larger for experts than for novices. Three kinds of video segments were identified: "blank segments" (containing no event, as identified by neither novice nor experienced teachers), "low contrast segments" (containing events identified only by experienced teachers), "high contrast segments" (with events identified by both). The results showed that novice teachers devoted more time in observing a disruptive student than experienced teachers did, the latter having a wider observation scheme. In low contrast segments, the experienced teachers exhibited shorter fixation times and a wider sampling across students than novices did. No differences were shown for the high contrast segments. There were no significant differences on the observation of blank segments between N and E teachers; no differences on the homogeneity of variance were found either. Since this study captures eye movements on a TV screen displaying a video footage, based on predetermined scenes, its proximity with authentic conditions is weak.

Cortina et al. [15] used a mobile eye-tracker to study the gaze behavior of 24 teachers (12 novices and 12 experienced). They analyzed the relationship between the quality of the classroom climate (using the CLASS, see below for more information), and the level of attention teachers devoted to each student of the classroom, computed with the Gini coefficient (ranging from 0: all students have the same number of fixations, to 1: only one student gets all the fixations). Results showed that the Gini coefficient of experienced teachers was significantly lower than this of novice teachers. Correlations between each CLASS dimension and the Gini coefficient were computed: quality of feedback score correlated positively and significantly with Gini scores ($r = .46$, $p < .05$), showing that the more teachers support learning in delivering feedback, the more their attention is equally drawn towards all the students.

These studies did not attempt to uncover steering groups, nor did they make any assumptions about the actual level of the students. We set up the following study to investigate these questions.

## 3 Research Questions

The main purpose of this research is to study the strategies of teachers' information gathering through a mobile eye-tracking device and in an ecological context. The use of such a device suits the highly dynamic nature of the classroom environment [16], where the diversity of the potential sources of change are difficult to capture with indirect observation tools. Our research questions are threefold:

- *Classroom awareness*: How can we characterize teachers' attention distribution among students? Is this attention related to some students' characteristics (like performance or behavior)? Does any "steering group" exist? If so, which are its features (number and level of students, number of groups)?
- *Relationship between classroom awareness and teacher–students interaction*: A teacher can be fully aware of what happens in his or her classroom without being reactive to any event. We thus have to check to what extent the teachers' awareness is related to the quality of his or her interactions with students. In other words, we sought to determine the relationship between teachers' visual cues in the classroom environment and their level and quality of the interactions they promote with students.
- *Learning Analytics-based visualization reuse*: Can the large dataset of this study, as well as its LA-related procedures, be spread to every researcher, or even teacher, who wants to investigate gaze teachers' behavior? Can we come up with some use cases of this database for teacher training or educational research purposes?

A novice–experienced comparison was undertaken for the first two research questions, supposing that more experienced teachers would be more aware of students' participation and achievement [17]. A specific LA-based procedure was undertaken to answer the third question.

# 4 Method

## 4.1 Participants

Four teachers (100 % female) volunteered to participate to this study. Table 1 below shows teachers' main characteristics.

**Table 1.** Basic information about teachers

| ID | Grade | Nb students | Experience (nb years) |
|---|---|---|---|
| 1 | 1st | 22 | High (20) |
| 2 | 3rd | 24 | Novice (0.5) |
| 3 | 2nd | 23 | Novice (0.5) |
| 4 | 1st | 24 | High (25) |

## 4.2 Measures

First of all, information about the students was gathered: age, gender, quartile level of performance in French and mathematics, special needs, and a 11-item questionnaire assessing the students' behavioral self-regulation abilities [18]. The following abilities were assessed: attention, tiredness, integration into the classroom, work speed, effectiveness, organizational capacity in performing a task, autonomy, and mastery of gestures. Teachers responded on a 4-point Likert scale, ranging from 1 for a behavior never noticed or not learned yet, to 4 for a behavior usually noticed or learned. A maximum likelihood exploratory factor analysis identified one factor as in a previous research [18]. The reliability was satisfactory (α = .77). In order to estimate each pupil's behavioural self-regulation perceived by the teacher, each student was given a score taking into account the factor weight of each item.

Then, we had to represent the occurrence of pedagogical events throughout the teaching sessions. We adapted the Teaching Dimensions Observation Protocol (TDOP) [19], which is a reliable observation tool that captures a large variety of pedagogical practices and events. We used the TDOP to characterize the diversity of pedagogical events that occur in classrooms (e.g., the teacher gives an explanation then the students are doing a guided exercise). This information was coded independently by two researchers from the video footages, and disagreements were resolved by a discussion to reach a consensus.

Eventually, we assessed the level of the teacher–students interactions in the classrooms with the Classroom Assessment Scoring System (CLASS) [2], one of the most used and valid classroom observation systems. The quality of the interactions was assessed upon three main domains: emotional support, classroom organization, and instructional support, derived into ten dimensions (see Table 2 for more information). This judgment of quality of the teacher–students interactions is related to the observation of four 30-minute sessions, hence lasting a whole morning session for each observed teacher.

**Table 2.** Gini coefficients, behavioral- and performance-related gazing time ratio, CLASS scores, per teacher

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Gini Coeff. Overall | 0.35 | 0.33 | 0.32 | 0.29 |
| Behav. Gazing Time Ratio Overall | 2.07 | 1.49 | 1.05 | 2.14 |
| Overall Attentional Lability | 36.9 | 53.8 | 49.8 | 36.8 |
| Gini Coeff. Interactive Exercise | 0.33 | 0.32 | 0.45 | 0.29 |
| Behav. Gazing Time Ratio Int. Ex. | 2.10 | 1.14 | 1.09 | 2.16 |
| Int. Exercise Attentional Lability | 37.6 | 65.2 | 48.1 | 42.2 |
| CLASS Positive Climate | 6.0 | 6.3 | 5.9 | 4.5 |
| CLASS Negative Climate | 1.0 | 1.1 | 1.1 | 1.7 |
| CLASS Teacher Sensitivity | 5.4 | 6.1 | 5.6 | 4.8 |
| CLASS Regard for Student Persp. | 4.5 | 5.6 | 5.3 | 4.2 |
| CLASS Behavior Management | 5.7 | 6.3 | 5.9 | 5.2 |
| CLASS Productivity | 5.7 | 5.9 | 5.9 | 5.2 |
| CLASS Instr. Learning Formats | 5.4 | 5.3 | 4.9 | 4.6 |
| CLASS Concept Development | 2.7 | 5.0 | 3.8 | 2.9 |
| CLASS Quality of Feedback | 4.5 | 4.9 | 4.3 | 3.4 |
| CLASS Language Modeling | 3.7 | 4.8 | 3.9 | 3.3 |

### 4.3 Procedure and Data Analysis

After a pre-experiment with a university teacher to rehearse the whole experimental scheme, we undertook a lesson recording in the four teachers' classrooms. The four participants taught a regular lesson of mathematics (numeracy or multiplication) lasting about 45 min in the morning, wearing a mobile eye-tracker (ASL Mobile Eye-GX). An additional video camera and an ambient microphone captured the whole classroom activity. Two trained observers gathered CLASS-related information during the whole morning class.

The set of lessons was then transcribed by two trained coders using ELAN [20]. The whole dataset was afterwards exported onto *UnderTracks* [21], a web-based Learning Analytics platform that enables the gathering, analysis, and sharing of a wide range of traces. *UnderTracks* is composed of a web platform to share traces and operators (processing algorithms in Python or R) and a client-side software (*Undertracks-Orange*) to build, share and reuse analysis processes (combination of operators and traces) thanks to the open source *Orange* data mining toolbox (http://orange.biolab.si). A given dataset, as well as its related analysis procedures, can thus be shared, reused, and modified by the *UnderTracks* researchers' community. Once shared, the processes can be applied to other traces. We created a specific data space in *UnderTracks*, called "SuperViseur", for storing raw data of this study, as well as displaying operators and processes used in its analysis. Raw data stored comprises gaze behaviors, students' characteristics, and pedagogical episodes.

The design and processing of the analyses of this study reused processes from within the *UnderTracks-Orange* client application. Figure 1 shows an *Orange* process that builds several interactive visualizations displaying teachers' gaze behavior under several considerations (pedagogical episode, students characteristics). For example, one of the interactive visualizations shows, for each teacher, his or her gaze behavior by pedagogical episode. Its interactivity lies in the possibility to have more information about a given student when the mouse hovers each gaze target representation.

Any visualization can be saved onto the *UnderTracks* web platform; visualization results can be uploaded in other web sites as well. A dedicated website (http://superviseur.lip6.fr) proposes several visualizations to furthering the exploration of our

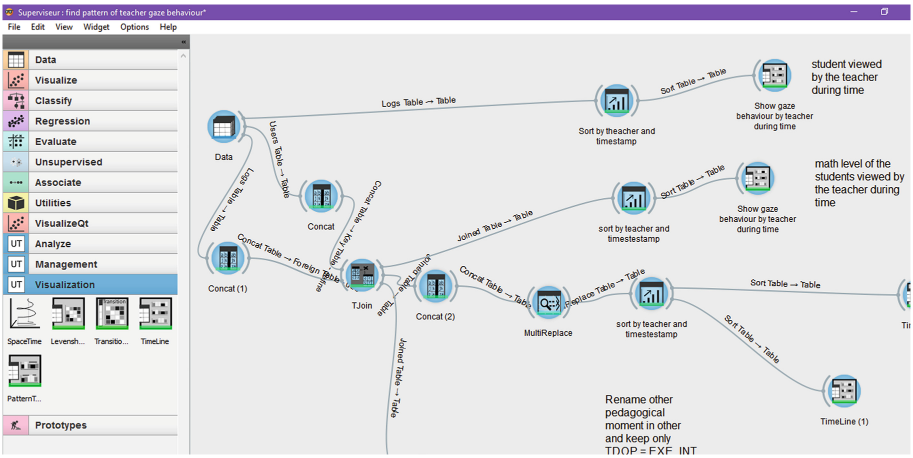



**Fig. 1.** An *Undertracks-Orange* process producing visualizations from teachers' gaze behaviors

data, likely following use cases (see Sect. 6). Moreover, any researcher can conveniently perform, upon registration, some of the analyses described in the paper, as well as new ones.

Specific attention was drawn to privacy concerns, and we sought agreements from our university data protection and ethical committees, mostly because the mobile eye-tracker necessarily captures the whole attentional stream of teachers, which makes it difficult to isolate students whose parents decided they would not be videotaped. The parents were given a description of the project and had to confirm their agreement. All of them agreed. The dataset available from within *UnderTracks* delivers fully anonymized data only, thus the video shots of the study are not viewable.

# 5 Results

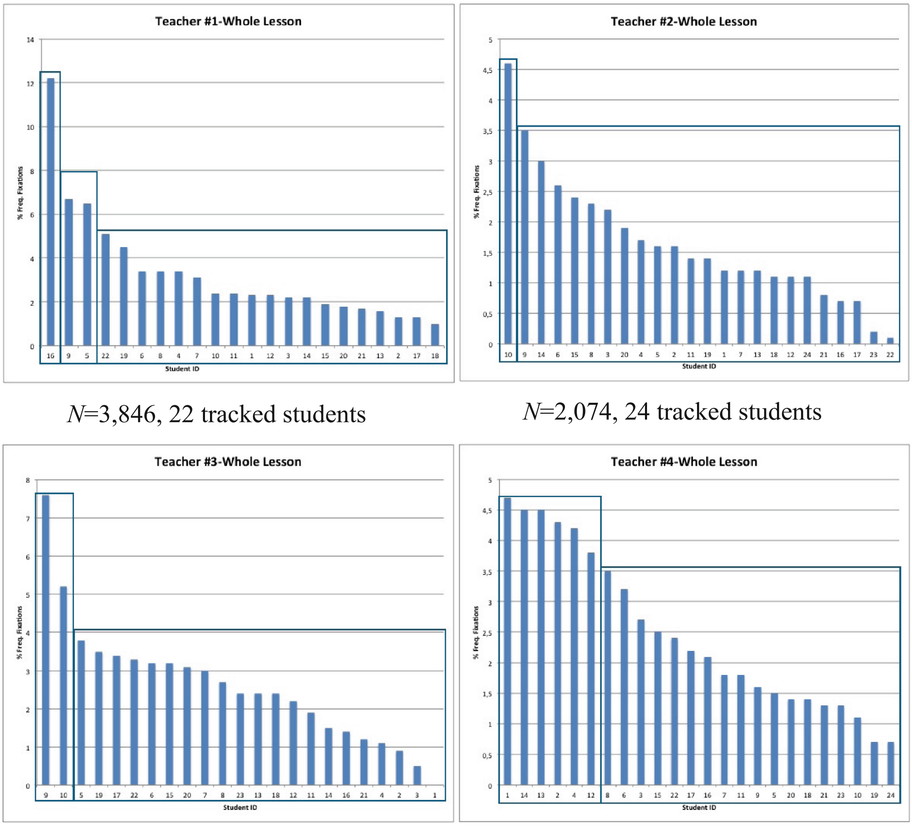
## 5.1 Gazing Time: Whole Lesson Analysis

We first selected the same video time range (44 min) to control for time, corresponding to 5,280 eye fixations of 500 ms duration each, per video footage. We then extracted those targeting a student and computed the percentage of time a teacher is focusing a given student (see Fig. 2). Whereas every student was targeted during the whole lesson session, the distribution of the gazing time differed among teachers: the first three devoted most time (about 10 %) on a reduced number of student while the fourth distributed her attention between students more evenly.

We then tried to compose "steering groups" (rectangles in Fig. 2) in function of the attention distribution of the teachers, using the following rule of thumb: We empirically set the cut-off value of eye fixations as 200 fixations (100 s of gazing time) for determining groups. Then, we separated the distribution by tiers every 200 fixations. Results show that Teacher #1 focused her attention towards three distinct "steering groups", a unique student (#16), a second group composed of two students (#9 and #5), and a third composed of the rest of the students. Teacher #4 exhibited a similar behavior, essentially focusing on a group of six students, the rest of the classroom being almost equally scrutinized. Teacher #2 and #3 focused more often their attention on a more reduced set of students (Student #10 for Teacher #3; Students #9 and #10 for Teacher #3), the others being far less attentionally sampled.

## 5.2 Teachers Gazing Time in Function of Pupils Behavior and Level

We then wondered if the teachers' gazing time would depend on some salient characteristics of their students, like their behavior or their level in mathematics. Figure 3 orders the students as in the previous figure (descending percentage of gazing time per student), in addition with categorical data about their level in mathematics (bars, the higher the better), as well as score data about their behavioral self-regulation (dotted lines, the higher the less dysfunctional). We expected that the more often a student is sampled over the lesson, the lower his or her level is (either in mathematics or related to his or her behavior).




*N*=3,846, 22 tracked students

*N*=2,074, 24 tracked students

*N*=3,176, 23 tracked students

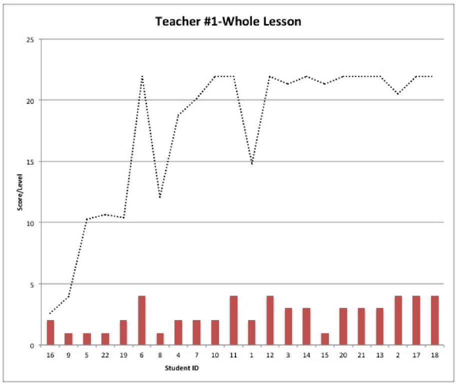
*N*=3129, 24 tracked students

**Fig. 2.** Distribution of the percentage of teachers' gazing time per student, sorted by descending rank, whole lesson (time range: 44 min)

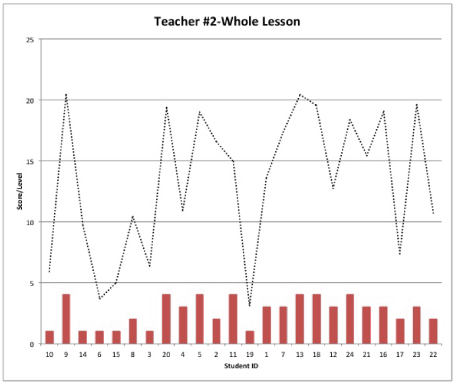
Figure 3 depicts this relationship, showing that Teachers #1 and #4, again, had similar ocular behaviors: their students' level curves are globally ascending, even if some irregularities occur (e.g., Students #8 and #1 were less observed by Teacher #1 than their behavioral level would let us think; likewise for Students #22, #5, and #21, Teacher #4). In comparison, the students' level curves of Teachers #2 and #3 are not ascending and much more erratic, showing no relationship between the percentage of eye fixations and students' level.

### 5.3 Analysis of a Specific Episode: Interactive Exercises

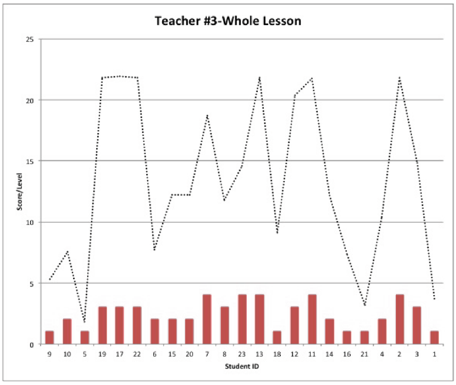
The above analyses made the assumption that teachers behave uniformly during the whole lesson in terms of information takes. In order to control for the kind of pedagogical event, we have now to analyze the participants' gaze behavior on the same kind of event. We chose to focus on the *interactive exercise* derived from the TDOP



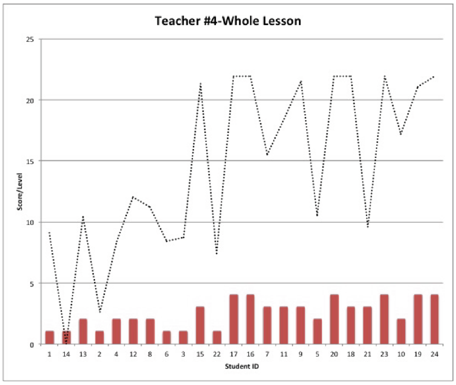

22 tracked students




24 tracked students




23 tracked students




24 tracked students

**Fig. 3.** Descending rank of the students in function of their gazing time, with information about their levels of behavior (dotted lines) and mathematics (bars) (time range: 44 min)

taxonomy (a mix of "interactive lecture" and "deskwork", frequently undertaken in French classrooms, enabling students to do exercises under teacher's guidance), which lasted sufficiently long in each lesson, and necessitated a larger amount of information from students than others. All in all, the pattern of results related to these episodes was very similar to the overall results (see Sect. 5.1 and Table 2). For the sake of brevity, the interactive exercise-based results are available at http://superviseur.lip6.fr.

### 5.4 Relationship Between the Attention Focus and the Classroom Climate

We computed Gini coefficients to measure teachers' attention distribution in interactive exercises [15], appropriate when the variable (in our case, attentional focus, or gazing time) is not independently distributed among students: if a given student is subject to focus, there leaves less chances of attentional focus to the others). The Gini coefficient ranges from 0 (all students get the same number of fixations) to 1 (one student gets all the fixations). Results show that Teacher #4 was the most "egalitarian".

Table 2 also shows the gazing time ratio between the amount of fixations towards less able students (in terms of behavior) and towards more able students, the cut-off between the two groups being the median. For instance, Teacher #1 had an overall behavioral gazing time ratio of about 2, meaning that she gathered two times more information from less able students than from more able students. The pattern of results regarding CLASS shows that, firstly, the smaller their behavior management CLASS-based scores are, the more teachers are "egalitarian", needing to scan a larger sample of students to manage their classroom. We obtained similar results with the gazing time ratio related to performance in mathematics.

### 5.5 Relationship Between Attentional Lability and Classroom Climate

The previous sub-Section considered teachers' gazing time as a whole. However, two teachers may differently distribute their overall – and equivalent – amount of attention over time, one being focused on the same student for many contiguous saccades, the other being constantly changing his or her attention across students. We computed (see Table 2) the percentage of gaze changes, namely "attentional lability," for the whole lesson and for the Interactive Exercises episodes (100 % stands for a change every saccade; 50 % stands for a change every two saccades).

The results on attentional lability are the only ones that are both related to the experience difference between teachers (the more experienced have the lower percentages), and to their CLASS scores. The ranking order of the teachers' overall attentional lability is the same as their respective CLASS scores for Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Student Perspectives, Conceptual Development, and Quality of Feedback. There thus might be a relationship between the teachers' attentional change over the students (his or her sensibility to students needs), and the quality of the teacher–students relationships (briefly put, the classroom climate); at the same time, experienced teachers exhibited lower attentional lability than novices did.

## 6 *UnderTracks* Use Cases

We can now sketch three uses cases, showing situations where researchers, teacher trainers, and even teachers, would take benefit from the analysis of eye-tracking data with *UnderTracks-Orange*. These use cases enable to foresee advances in the novel research domain of "Teaching Analytics" [22].

*Use Case #1: Studying teachers' cognition from classroom management patterns*. As argued in the first two Sections of this paper, there are numerous hypotheses on teacher cognition that would take advantage from being more objectively validated through LA-based eye-tracking data analyses. Researchers connected to large datasets of teachers' behaviors would uncover novel fine-grained classroom management patterns.

*Use Case #2: Studying teachers' efficacy in relation with students learning*. Evidence-based research has recently spread from medicine to educational research [23]. Given that perspective, researchers would use the kind of data we gathered, extended by students' indicators of performance. This would enable the study of the causality between raw behavioral indicators and learning.

*Use Case #3: Uncovering behavior patterns for teacher training purposes*. Teacher training sessions would also benefit from the device tested in this study. Pre-service teachers would be given access to videotaped lessons and their *UnderTracks*-based data; they would investigate some hypotheses about the teacher's awareness, his or her information takes, and their relationships with the students' behavior and performance, as well as with the classroom climate. Eventually, some instructional strategies would be derived from their conclusions.

## 7 Discussion

This paper considered the combined use of eye-tracking data together with Learning Analytics procedures leading to open and interactive visualizations of teachers' strategies. Our main results are summarized as follows. Firstly, every student of the four classrooms was looked at by his or her teacher, even a few times. This brings some support to the "withitness" hypothesis. Secondly, steering groups composition differed across teachers: very small groups of students were particularly subject to focus by the teachers, and thus can be considered as more complex in terms of decision-making. The size of the gazed groups seems to be related to the amount of experience of the teachers, as found in [14]. Thirdly, very little variability was observed across different kinds of pedagogical activity. Fourthly, the criterion for choosing a steering group is not clear-cut across teachers: again, teachers' amount of experience better predicted their steering groups-related behavior than the characteristics of their students, in terms of behavior or performance. Eventually, we found a small relationship between teachers' gazing time and the quality of the classroom climate, replicating Cortina et al.'s [15] results, as well as a more obvious relationship between the teachers' attentional lability and many of their CLASS scores.

During their activity, novice teachers engage a larger amount of cognitive load than more experienced do. The way the latter scan a larger "steering group" would make them able to perceive more fine-grained events [24], since they are less overwhelmed than novices are. This "steering group" is action-oriented, so it likely contains students whose behavioral changes may have effects on teachers' strategies (activity change, feedback, etc.) [25]. Novice–expert comparison studies in many fields (aviation, chess, sport, surgery) showed that experts, compared to novices, have fewer fixations of longer duration on nodal points of the situation [26, 27], while novices exhibit more variability. This line of results complies with our paradoxical result, at least at first sight, showing that an experienced teacher might be either egalitarian (i.e., with a smaller gazing time variability across students), and focused on a small set of specific students (i.e., with a restricted "steering group"). Focusing on this group of students allows expert teachers to make sound decisions, grounded on a representative set of students.

Further research will engage a larger set of participants, and consider the actual teachers' location in the classroom to test more ecological hypotheses, as well as finer-grained analyses of more complex episodes, like those involving teacher feedback. The implementation of some use cases in real-life contexts will be considered as well. They are paths to understand how teachers adapt themselves, with sensitivity, to their classroom environment and their students' needs.

**Acknowledgments.** This research was partly funded by the Pôle Grenoble Cognition, Univ. Grenoble Alpes, France. This research was approved by the CERNI (University's local ethical committee, n° 2013-09-24-25). We would like to thank the four teachers for having accepted to wear so weird a device and nevertheless doing good teaching; Brigitte Meillon for her invaluable help in calibrating, capturing, and post-producing the video footages; Michèle Arnoux and Mathieu Louvart for CLASS and videos' coding; Luc Sindirian and Pascal Bilau for making this research possible; and Andrea Doos for checking the English of a previous version of this paper.

## References

1. Rittenbruch, M., McEwan, G.: An historical reflection of awareness in collaboration. In: Markopoulos, P., Mackay, W., de Ruyter, B. (eds.) Awareness Systems: Advances in Theory, Methodology and Design, pp. 3–48. Springer, London (2009)
2. Pianta, R.C., La Paro, K.M., Hamre, B.K.: Classroom Assessment Scoring System: Manual K-3. Brookes, Baltimore (2008)
3. Kounin, J.S.: Discipline and Group Management in Classrooms. Holt, Rinehart & Winston, New York (1970)
4. Lundgren, U.P.: Frame Factors and the Teaching Process. Almqvist & Wiksell, Stockholm (1972)
5. Hastie, P.A., Sinelnikov, O.A., Brock, S.J., Sharpe, T.L., Eiler, K., Mowling, C.: Kounin revisited: tentative postulates for an expanded examination of classroom ecologies. J. Teach. Phys. Educ. **26**, 298–309 (2007)
6. Irving, O., Martin, J.: Withitness: the confusing variable. Am. Educ. Res. J. **19**(2), 313–319 (1982)
7. Wanlin, P.: Elèves forts ou faibles: qui donne le tempo? Université de Liège, Liège (2011)
8. Ericsson, K.A., Simon, H.A.: Verbal reports as data. Psychol. Rev. **87**, 215–251 (1980)
9. Salvucci, D.D., Anderson, J.R.: Automated eye-movement protocol analysis. Hum. Comput. Interact. **16**, 39–86 (2001)
10. Mele, M.L., Federici, S.: A psychotechnological review on eye-tracking systems: towards user experience. Disabil. Rehabil. Assist. Technol. **7**(4), 261–281 (2012)
11. Yang, F.-Y., Chang, C.-Y., Chien, W.-R., Chien, Y.-T., Tseng, Y.-H.: Tracking learners' visual attention during a multimedia presentation in a real classroom. Comput. Educ. **62**, 208–220 (2013)
12. Prieto, L.P., Sharma, K., Wen, Y., Dillenbourg, P.: The burden of facilitating collaboration: towards estimation of teacher orchestration load using eye-tracking measures. In: 11th International Conference on CSCL 2015, vol. 1, pp. 212–219. ISLS, Gothenburg (2015)
13. Wolff, C., Van den Bogert, N., Jarodzka, H., Boshuizen, H.P.A.: Differences between experienced and student teachers' perceptions and interpretations of classroom management events. In: Inter-University Center for Educational Sciences Fall School, Girona, Spain (2012)

14. van den Bogert, N., van Bruggen, J., Kostons, D., Jochems, W.M.G.: First steps into understanding teachers' visual perception of classroom events. Teaching Teach. Educ. **37**, 208–216 (2014)
15. Cortina, K.S., Miller, K.F., McKenzie, R., Epstein, A.: Where low and high inference data converge: validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. Int. J. Sci. Math. Educ. **13**(2), 389–403 (2015)
16. Doyle, W.: Classroom organization and management. In: Wittrock, M.C. (ed.) Handbook of Research on Teaching, pp. 392–431. McMillan, New York (1986)
17. Gettinger, M., Kohler, K.M.: Process-outcome approaches to classroom management and effective teaching. In: Evertson, C.M., Weinstein, C.S. (eds.) Handbook of Classroom Management: Research, Practice, and Contemporary Issues, pp. 73–96. Erlbaum, Mahwah (2006)
18. Guimard, P., Cosnefroy, O., Florin, A.: Evaluation des comportements et des compétences scolaires par les enseignants et prédiction des performances et des parcours à l'école élémentaire et au collège. Orientation Scolaire et Professionnelle **36**(2), 179–202 (2007)
19. Hora, M.T., Oleson, A., Ferrare, J.J.: Teaching Dimensions Observation Protocol (TDOP) User's Manual. Wisconsin Center for Education Research, University of Wisconsin-Madison, Madison (2013)
20. Sloetjes, H., Wittenburg, P.: Annotation by category – ELAN and ISO DCR. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008). ELRA, Marrakech (2008)
21. Bouhineau, D., Lalle, S., Luengo, V., Mandran, N., Ortega, M., Wajeman, C.: Share data treatment and analysis processes in technology enhanced learning. In: Data Analysis and Interpretation for Learning Environments Workshop, Held in Conjunction with the Alpine Rendez-Vous '2013, Villard-de-Lans, France (2013)
22. Vatrapu, R., Reimann, P., Halb, W., Bull, S.: Second international workshop on teaching analytics. In: Proceedings of Third International Conference on Learning Analytics and Knowledge (LAK 2013), pp. 287–289. ACM, New York (2013)
23. Hattie, J.: Visible Learning: A Synthesis of over 800 Meta-analyses Relating to Achievement. Routledge, New York (2009)
24. Hogan, T., Rabinowitz, M., Craven, J.A.: Representation in teaching: inferences from research of expert and novice teachers. Educ. Psychol. **38**(4), 235–247 (2003)
25. Doyle, W.: Making managerial decisions in classrooms. In: Luke, D.L. (ed.) Classroom Management, pp. 42–74. University of Chicago Press, Chicago (1979)
26. Ward, P., Williams, M., Hancock, P.A.: Simulation for performance and training. In: Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (eds.) The Cambridge Handbook of Expertise and Expert Performance, pp. 243–262. Cambridge University Press, Cambridge (2006)
27. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., van de Weijer, J.: Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press, Oxford (2011)

# Flipped Classroom Model: Effects on Performance, Attitudes and Perceptions in High School Algebra

Peter Esperanza[1], Khristin Fabian[2(✉)], and Criselda Toto[3]

[1] Barstow High School, California, USA
Peter_esperanza@busdk12.com
[2] University of Dundee, Dundee, Scotland
m.k.fabian@dundee.ac.uk
[3] Chapman University, Orange, USA
toto@chapman.edu

**Abstract.** In this study, we evaluated student perceptions of the flipped classroom model and its effects to students' performance and attitudes to mathematics. A randomized controlled trial with 91 high school algebra students was conducted. The experimental group participated in a year-long intervention of the flipped classroom model while the control group followed the traditional lesson delivery. Results of the year-end evaluation of this model showed positive student perceptions. An analysis of covariance of the algebra post-test score with learning model as treatment factor and pre-test as covariate resulted in a significant treatment effect at .05 level of significance. A paired-sample t-test by treatment group to compare pre-test and post-test math attitude scores resulted in a significant decrease in the control groups' value of mathematics while the experimental group had a significant positive change in their confidence and enjoyment of mathematics.

**Keywords:** Flipped classroom · Mathematics education · Attitudes · Student perceptions

## 1 Introduction

Past research has indicated that a strong grounding in algebra correlates to successful post-secondary education [1] but research has also shown that algebra students need more support to succeed as even students taking post-secondary level algebra classes are still inadequately prepared [2]. Among the strategies suggested to better prepare students include: provision of supplementary learning [3], promotion of conceptual understanding and procedural fluency in algebra [4] and the use of solved problems to engage students in analyzing algebraic reasoning and strategies [5]. These strategies appear to be a good match to the flipped classroom model.

Abeysekera and Dawson [6] characterizes the flipped classroom model as a change in use of class time and out-of-class time. Sometimes called the inverted classroom [7], this model utilizes a setup where previous homework activities are now done in class in the forms of active learning, peer learning and problem solving. Typical class lectures


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 85–97, 2016.
DOI: 10.1007/978-3-319-45153-4_7

are then delivered via videos for out-of-class viewing. With this setup, less time is dedicated by the teacher to repeat information thus making it possible to provide students with more exercises and activities that promote conceptual and procedural fluency.

Reported benefits of flipped learning model include an increased student satisfaction, improved communication skills and consequently, an enhanced learning experience [8]. These findings, however are for higher education and evidence of positive effects of flipped learning in high school particularly those that examine student performance are limited [9]. To fill the gap in research, this study is conducted with high school students and focuses evaluation in student perception and performance.

This study aims to answer the following research questions:

RQ1. Is there an effect to students' performance in an algebra test when flipped classroom is adopted as a teaching model?

RQ2. Is there a change in students' attitudes to mathematics when flipped classroom model is used?

RQ3. How do students perceive the use of the flipped classroom model in terms of its usefulness in learning mathematics?

## 2 Review of Literature

There is a considerable amount of literature that showcases positive student perceptions towards flipped learning. Some students feel that the use of lecture videos as preparatory material before class helped them understand the concepts better [10–14] and that the ability to pause and replay sections of the video allowed students to learn at their own pace [15–18]. The class activities, on the other hand, were more enjoyable, engaging and useful, [11, 15, 19–21]. In addition, the teacher in a flipped classroom model appears to be more available to provide guidance on difficult topics [12, 22]. Furthermore, this model has also fostered improved communication skills among students particularly their skills in communicating mathematical ideas [19].

Not all reports about flipped learning are positive. One of the frequently cited advantages of flipped learning is its ability to support students to follow their own pace through the use of the media controls available on the video lectures but some studies report that this utility is not fully utilized [10, 11]. Some studies note that students had difficulties in adapting this model [23–25]. Issues with flipped learning include: the lack of access to an expert while viewing the videos out of class [17, 23, 25]; that it required more effort and organization; and gives one the feeling of being left out when videos are not viewed prior to class [23, 25]. In fact, some students prefer the traditional model over the flipped classroom approach [18, 23].

Limited information is known about the effects of the flipped classroom model to students' attitudes towards mathematics. Only two studies [18, 23] used pre and post intervention data to measure change in mathematics attitude. Guerrero et al. [23] found that this model led to significant student gains in enjoyment and value of mathematics. In contrast, Young [21], found that students in the flipped classroom had more negative attitudes towards mathematics after the intervention. The rest of the studies that covered students' attitudes were self-reports students provided at the end of the intervention. Weng [16] reported that students feel less anxious about mathematics as a result of using this model. Love et al. [14] found that using the flipped classroom format led to students having reasonably more positive outlook about the importance of mathematics to future careers. Similarly, Touchton [26] found that more students in the flipped classroom expressed an increased interest to take more advanced statistics courses. Lape et al. [17], on the other hand, found that students lacked the motivation to attend class because of the model. Given this gap in literature, it is the goal of this study to investigate the effects of using the flipped classroom model to student attitudes towards mathematics using before and after intervention data.

A literature search of flipped classroom implementations in mathematics and its effect to student performance yielded a limited number of results. A summary of these studies is listed in Table 1. There were studies that showed students in the flipped group outperformed their comparison groups [11, 12, 20, 22, 26]. Two studies had mixed results. Love et al. [14] found that while students in the flipped group initially outperformed the control group midway of the study, the control group was able to catch up towards the end of the intervention. Overmyer [27] found that students taught using the flipped classroom model by a lecturer with an experience in inquiry-based learning and cooperative learning performed better than the non-flipped group and those who were taught using the flipped model but with an inexperienced teacher. There were also studies that found student performance did not vary by teaching model [17, 18, 23, 28].

Student perceptions of a flipped classroom was not found to be an indicator of performance [11]. In general, however, studies that reported an improvement in students' performance also reported positive student perceptions and studies that reported no difference in student performance between the control and experimental group are the same studies that reported negative student perceptions.

All studies mentioned in this section were conducted at university level mathematics except for Muir and Geiger [13] and Kirvan et al.'s [28] work. Muir et al. reported positive student perceptions towards flipped learning while Kirvan et al.'s work found no difference in the performance of students who were taught using the traditional model of mathematics and students taught with the flipped classroom model. It is thus, another goal of this study to focus the evaluation of the flipped classroom to students' performance in high school mathematics, where student expectation and classroom setup is very much different to undergraduate level mathematics.

## 3 Methodology

### 3.1 Research Design and Nature of the Intervention

The study adopted a randomized controlled trial to evaluate the effectiveness of the flipped classroom model. It took place in a public high school in a high desert area in California, USA. The school population is about 1380 students comprised of 26 % Caucasian, 3 % Asian, 55 % Hispanic, and 16 % African-American students where 70.6 % of students are qualified in free or reduced price lunch.

**Table 1.** Summary of findings related to student performance in mathematics

| Study | Math topic | Performance measure | Results |
|---|---|---|---|
| [11] | Statistics | Course grade and final exam | There was an improvement of course grades of EG ($p < .001$). Their final exam scores were also better than CG ($p < .001$). |
| [12] | Calculus | Exam | Students from EG performed better in their exams in comparison to the non-flipped class. |
| [14] | Linear algebra | Exam | EG outperformed CG in the two midterm exams but by the final exam, the two groups' performance was not significantly different. |
| [17] | Differential equations | Homework, criterion referenced test (CRT), exam | There was no difference between EG and CG's pre and post CRT scores ($p > .05$). The composite homework and exam scores of the two groups also showed no difference. |
| [18] | Calculus and finite mathematics | Exam and course grade | There was no statistically significant difference found between the experimental and comparison group. |
| [20] | Statistics | Exam, grade and standard test | EG outperformed students in CG in their course grade ($p < .01$), exam grades ($p < .05$), and standard test ($p < .05$). |
| [22] | Algebra | Final exam scores | EG performed better than the CG ($p < .05$). |
| [23] | Finite mathematics | CRT | There was no statistically significant difference between EG and CG at pre-test nor at post-test. |
| [26] | Statistics | Project | EG performed better than CG but the magnitude of this difference is small. |
| [27] | Algebra | CRT | EG taught by an experienced teacher in inquiry-based learning performed better than CG as well as those in the EG but taught by an inexperienced teacher. |
| [28] | Algebra | Standard algebra test | The similar magnitudes of the pre- to posttest effect sizes for the EG and CG suggest that the degree of difference in instructional focus had less of an effect in student performance. |

Note: Students in the experimental group (EG) are students in the flipped classroom model. CG refers to the comparison group.

Students were randomly assigned into two groups: flipped classroom model (experimental group) and traditional model (control group). Both groups participated in the study for the whole academic year. For the duration of the study, the experimental group received an average of three videos per week as part of the flipped classroom model whereas the traditional group received an average of three homework/practice exercises per week. All learning activities carried out in the experimental group was also carried out in the control group. For example, if the lesson includes 10 practice exercises, then the experimental group will work on these exercises within class hours. The control group will work on half of the exercises within class and the other half as assigned homework. A typical 50-minutes lesson structure and how it varies between groups is illustrated in Fig. 1.

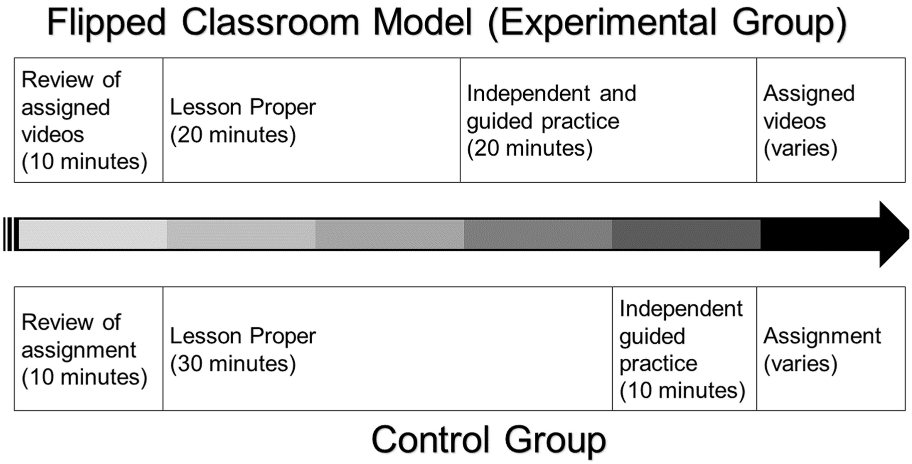



**Fig. 1.** Comparison of lesson structure in control and experimental group.

Learning activities for both groups include collaborative activities, problem solving, guided and independent practice, however, the videos are additional resources for the experimental group. For example, in one of the student projects, students took photos of the four different conic sections that they see outside the school and were asked to generate the equation of each conic section. Their task is to construct a poster that relates the photos they have taken to conic sections. This activity is a two-day paired activity and students worked together to finish all the work in class. The nature of the activity is the same for traditional and flipped classroom. The differences lie in the amount of time dedicated to classroom based learning activities and student implementation of the activities. For example, students in the flipped classroom accessed the flipped videos with their mobile devices to remind them of the concepts they need for the project. The questions they raised to the teacher was more about the output expected (i.e. how big should the poster or in what format). They were also able to finish the activity within the allocated time. Students in the traditional classroom, on the other hand, used the notes they took during class, their textbook and asked the teacher to review the concepts that they forgot. They also worked on the activity for two days but some students had to carry on doing the work at home as they weren't able to finish on time.

The videos used in the experimental group follows production using the Fizz Method [29]. Using this method, the videos have the following characteristic: minimal post-production, and usually completed in a single attempt; the teacher appears in the video and the notes are handwritten. The minimal post-production contributes to the simplicity of the video and easiness of video production. The talking head provides the non-verbal cues that might aid students and is also proven to be more engaging in online video formats [30]. The handwritten notes, as McCammon explains, is a form of modeling that allows students to see their thought processes and supports understanding.

### 3.2 Participants

A total of 91 second and third year high school students were randomly allocated into experimental and control group. There were 46 students (23 male, 23 female) in the control group and 45 students (24 male, 21 female) in the experimental group. The teacher participant taught both groups and more than 10 years of experience teaching high school mathematics and two years of teaching using the flipped classroom model.

### 3.3 Measures and Instruments

**Attitudes Towards Mathematics Inventory (ATMI).** Tapia and Marsh's [31] attitude inventory for mathematics consists of four subscales with a test-retest reliability of .89. The subscales (with the corresponding number of questions and reliability scores are as follows): value of mathematics (10 questions, .70), enjoyment (10 questions, .84), self-confidence (15 questions, 88) and motivation (5 questions, .78). Responses were scored using a five-point Likert-scale with 1 being strongly disagree and 5 being strongly agree. Negatively phrased items were reversed-scored. Scores on the ATMI subscale were computed for each student by adding the corresponding numerical score for each of the item on that subscale.

**Algebra Test.** To measure student performance, twenty-five questions from the released California Standards Test [32] were randomly selected to be included in the study. The resulting test consists of multiple-choice questions with the following topic distribution: polynomial and rational expressions (9 items), quadratics, conics, and complex numbers (5 items); exponents and logarithms (5 items), series, combinatorics, and probability (6 items).

**End Activity Evaluation.** The end activity evaluation consists of five questions relating to student perception about the usefulness of the flipped classroom model. Questions were arranged in a 5-centimeter line marking scale with labeled endpoints (0 = strongly disagree; 5 = strongly disagree). Students rated their agreement with the statement by placing a dot on the line. The score was measured by measuring the placement of the dot from the left-hand side of the scale using a ruler. The higher the score, the higher the agreement with the statement. Students were also asked, in the form of open-ended questions, what they liked/disliked about the flipped classroom and suggestions on how to improve the current model.

### 3.4 Procedure

At the start of the term, students in the experimental group were given an orientation on the nature of the course. In the orientation, the experimental group were made aware that the purpose of the videos is to help them prepare for the next lesson and to cut down the time that they are allocated for note-taking in class. Their obligations, as such, is to watch the videos beforehand and summarize the video content and list down questions that they might have. An ATMI pre-test, followed by the algebra test the day after, was completed by both groups during the first two days of the semester. Students in the experimental group followed the flipped classroom model and the control group the traditional model as was illustrated in Fig. 1. At the end of semester 2, students completed the ATMI and algebra post-test. An end activity evaluation was also completed by the experimental group.

## 4 Results and Discussion

### 4.1 Student Performance

To compare the groups before and after the intervention, an independent t-test of the CST score was conducted. There was no significant difference between the experimental group ($M = 5.93$, $SD = 2.50$) and control group ($M = 5.96$; $SD = 2.18$), $t(89) = -.047$, p-value = .962, ES = 0.01 but there was a significant difference in the groups test score at post-test, $t(89) = 2.029$, p-value = .045, ES = –0.43. The experimental group ($M = 10.36$, $SD = 3.10$) performed better than the control group ($M = 9.02$, $SD = 3.173$). An independent t-test of the gains score, however, resulted in no significant difference, $t(89) = 1.710$, p-value = 0.09, ES = 0.59 but with a moderate effect size. This change is illustrated in Fig. 2. To address the question whether the learning method had an effect in the post-test scores, an analysis of covariance (ANCOVA) was conducted. An ANCOVA of the post-test score with learning model as treatment factor and pre-test as covariate resulted in a significant treatment effect, $F(88) = 3.23$, $p = .04$.

The findings of this study are in keeping with Van Sickel's [22] work on Algebra as opposed to those studies that found no difference in the performance of experimental and control group [17, 18, 23, 28]. The length of the intervention of this study, however, is arguably longer than the previously cited studies so it is possible that the length of the intervention might have been a factor in the improved scores. It can be assumed that students over time became more familiar with the flipped classroom model and consequently was able to make better use of it to fit their learning styles. It is also worth noting that the instructor for this module had 11 years of teaching experience and has been using the flipped classroom model in the past 2 years. This supports

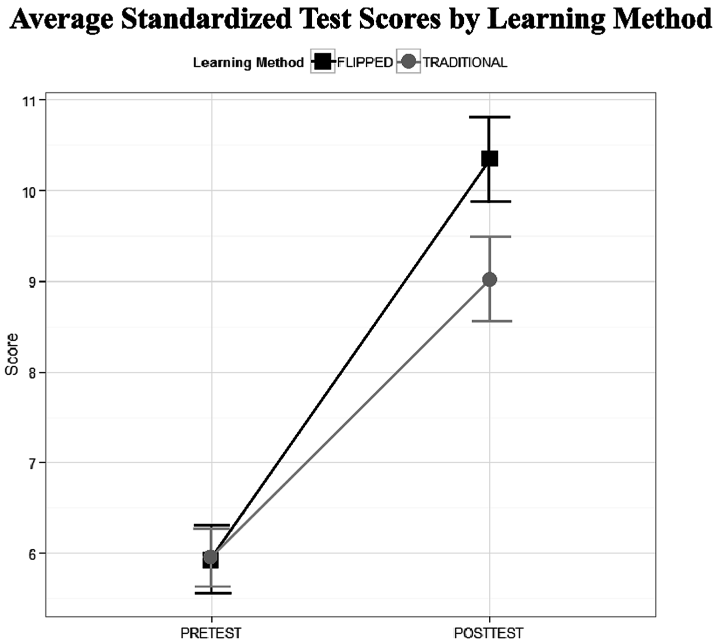



**Fig. 2.** Pretest and posttest scores of experimental and control groups

Overmyer's [27] findings that the experience of the teacher is a factor in running successful flipped classrooms.

### 4.2 Attitudes Towards Mathematics

Table 2 shows the means and standard deviations of the ATMI scores of the experimental and control group. The gain scores were computed by subtracting the pre-test score from the post-test score. A positive difference in the score means an increase in students' attitude whereas a negative difference means otherwise. To test whether this change in score was significant a paired sample t-test was conducted for each subscale. To test whether the gains of the experimental and control group were statistically different, an independent t-test of the gains score was also computed.

For the subscale value of mathematics, the control group, had a significant change in their pre and post-test scores, $t(45) = -2.74$, $p = .008$, $ES = -.21$. In contrast, while the experimental group also had a decrease in score, this change was not significant, $t(44) = -1.90$, $p = .064$, $ES = -.11$. No other significant change was found in the control group. The experimental group, on the other hand had a significant positive change in their enjoyment of mathematics, $t(44) = 3.15$, $p = .003$, $ES = .47$ and self-confidence $t(44) = 2.88$, $p = .006$, $ES = .43$. The findings of this study has some similarities with Guerrero et al. [18] which also found significant gains in student enjoyment of mathematics although in this instance, students actually had lower value of mathematics at post-test in comparison to their pre-test score. This decline, however,

**Table 2.** ATMI scores of control and experimental group

| | Value of mathematics | Enjoyment | Self-confidence | Motivation |
|---|---|---|---|---|
| Control | | | | |
| Pre M(SD) | 4.05(.51) | 3.52(.75) | 3.55(.82) | 3.54(.79) |
| Post M(SD) | 3.94(.62) | 3.56(.84) | 3.71(.80) | 3.55(.94) |
| Gains M(SD) | −.10(.50) | .04(.56) | .16(.70) | .004(.87) |
| p-value | .008 | .632 | .119 | .251 |
| Effect size d | −.21 | .07 | .23 | .01 |
| Experimental group | | | | |
| Pre M(SD) | 4.02 (.45) | 3.36 (.66) | 3.44 (.77) | 3.50 (.72) |
| Post M(SD) | 3.96 (.65) | 3.62 (.65) | 3.79 (.78) | 3.61 (.82) |
| Gains M(SD) | −.06 (.58) | .26 (.54) | .34 (.80) | .12 (.67) |
| p-value | .064 | .003 | .006 | .973 |
| Effect size d | −.11 | .47 | .43 | .17 |
| Independent t-test on gains between groups | | | | |
| p-value | .73 | .06 | .25 | .50 |
| Effect size d | .08 | .31 | .23 | .15 |

was not as severe as Young's [16] study which resulted in more negative attitudes towards mathematics.

### 4.3 Student Evaluation of the Flipped Classroom Model

Student evaluation of the flipped classroom model has been positive (see Table 3). The benefits of flipped learning as covered by previous studies were also observed in this study. This includes support to pace one's learning [15–18], improved communication channels [19], and improved understanding of mathematics concept [10–14]. Furthermore, students from this study also reported that they became more motivated to study math because of the flipped classroom model contrary to the findings of Lape et al. [17]. It is worth noting that Lape et al.'s study had different conditions—the study duration was shorter, on a different and more advanced math topic, with older students and with teachers who are relatively new to this approach. These factors may explain the differences in results.

Relationship between students' perception of the flipped classroom model against their gains in the algebra test and ATMI was also examined (see Table 4). There was a moderate positive correlation between students' gains in motivation and students' perception of the flipped classroom models' support for pacing ones' learning, r = .334. There was also a positive correlation between students' perception of the utility of the flipped classroom model to improve communication channels and their gains in value of mathematics, r = .348 and gains in motivation, r = .295. Contrary to Cilli-Turner's result [11], students' perception of the usefulness of flipped classroom to improve performance was found to be positively correlated to gains in the ATMI subscales and gains in the algebra test. Items relating to student motivation and

**Table 3.** Student evaluation of the flipped classroom model

| | Mean | SD |
|---|---|---|
| Q1. The flipped classroom allowed me to pace my own learning. | 3.94 | 1.28 |
| Q2. I feel that this model helped me communicate with my teachers and classmates. | 3.34 | 1.48 |
| Q3. I became more motivated to study maths as a result of the flipped classroom model. | 3.80 | 1.39 |
| Q4. I feel that my understanding of maths concepts has improved as a result of using this model. | 4.44 | 0.82 |
| Q5. I prefer the flipped classroom model over traditional lectures. | 4.53 | 0.98 |

**Table 4.** Correlation between student evaluation and gains

| | Gains algebra test | Gains value math | Gains enjoyment | Gains self-confidence | Gains motivation |
|---|---|---|---|---|---|
| Q1 | .201 | .180 | .145 | .002 | .334* |
| Q2 | .123 | .348* | .094 | .146 | .295* |
| Q3 | .211 | .115 | −.012 | .026 | .260 |
| Q4 | .380* | .620** | .622** | .356* | .493** |
| Q5 | .210 | .120 | .198 | −.011 | .249 |

*Correlation is significant at 0.05 level
**Correlation is significant at 0.01 level

preference to use this model had no significant correlation with the gains computed for this study.

In the open-ended questions, students from the experimental group explained what they liked/disliked about the flipped classroom model. Students appreciated the model because it allowed them to pace their own learning, go back to the videos when they have to and spend some time to reflect on the material as they take down notes for the topic covered in the video (n = 18). One student explained *"I like how if I didn't understand something I could rewind the video and listen again–something I could not do if a teacher were lecturing in class.*" This finding is consistent with the frequently quoted advantages of flipped learning model [15–18].

The lack of homework is also another thing they appreciated (n = 17). As one student explained, the flipped model allowed her to "*not have to answer math problems that I don't understand at home.*" Instead, the students feel that as they are doing their homework in school, they in-turn receive more help (n = 7). Students feel that this model have made them understand the topic better (n = 7). In addition, students mentioned that the videos used to prepare for class allowed them more time to reflect and gives them an idea of what is going to happen in-class (n = 3). Other advantages that students mentioned are: the chance to get more worked examples (n = 3), its support for anytime, anywhere learning (n = 3), and the opportunity it allows to make up for missed classes (n = 1).

When asked about what they disliked about the model most of the students replied that they have no complaints about the setup (n = 22) but a few have mentioned that they encountered difficulties accessing the video in some occasion (n = 6) which in turn gives the feeling of having to catch up in class the next day. Another recurring issue is that the videos did not allow them to ask questions (n = 5) so whatever question they have will have to wait for the next class. This ties in with their comments on how to improve the current model by having a comment section so that students can leave questions about the videos they just viewed. Overall, however, students were satisfied with the implementation and the recommendations for improvement have more to do with the interface design of the video channel rather than the content.

### 4.4 Limitations of This Study

There are several limitations of this study. The sample size is slightly lower than the recommended sample size. To counter this limitation, we reported the effect size to help us analyze the results. It is also possible that the order of the test at the end of the study might have affected the results. In the pre-test, ATMI was administered before the algebra test but in the post-test, this was not followed. Whether this had an effect to students' rating of their attitudes towards mathematics is not known. This leads to another limitation of the self-reporting nature of the two measures used in this study. For example, in the end evaluation of the study, the mean score of 3.8 for the question "I became more motivated to study because of the flipped classroom model" is a good indicator that students became more motivated but the change in ATMI-motivation was not significant. We have not addressed this limitation in this current study but for future research, we think it would be worthwhile adding qualitative data to the current design to validate these self-reports. Last, the videos used in this study are available on the web which the control group might be aware of. We had no way of monitoring whether the control group used these videos to support their learning needs or not. It is important to note, however, that the flipped classroom is not about the videos but about the structure of the course. Whether select students used these videos to help them with their assignments does not change the way control groups' classes were organized.

## 5 Conclusions and Implications for Research

The results of this study found that the use of the flipped classroom model had resulted to gains in student performance (#RQ1) and positive attitudes towards mathematics (#RQ2). We also found that students have positive perceptions about the usefulness of the flipped classroom model (#RQ3). We aimed to provide the same learning activities for the control and experimental group but admittedly the need to cover more material in class resulted to shortened learning activities in the control group and we believe that this is where the difference lies.

The videos that we used for this session were short 5 to 10-minutes videos. Keeping the videos to a minimum length is not just useful for production purposes but also for maintaining students' focus. The videos are, after all, meant to be preparatory materials for the next day's lesson and not substitutes to the actual discussion.

A lot of studies on flipped learning focused on the video element of the course but implementing the flipped classroom model required not just preparation of the videos to be used but also required planning of in-class activities. Successful implementation of a flipped classroom requires an agreement with the students that they will engage with the videos before class in place of the assignments that they are normally assigned. We believe that this preparation enables students to engage with the materials better in class and contributes to the success of the flipped classroom model.

Flipped classroom requires a lot of initial effort particularly in the preparation of video materials. For this study, the videos used were prepared and used the previous year so no further effort was required from the instructor in terms of developing new videos. We understand, however, that this is something those new to flipped learning would struggle with but it is also worth keeping in mind that the videos produced are reusable resources that teachers can build over time so this balances out the initial effort required.

## References


1. National Mathematics Advisory Panel: Foundations for Success: The Final Report of the National Mathematics Advisory Panel. National Mathematics Advisory Panel (2008)
2. Pinzon, D., Pinzon, K., Stackpole, M.: Re ‘modeling’ college algebra: an active learning approach. Primus **26**, 179–187 (2016)
3. Sorensen, N.: Supplementary learning strategies to support student success in algebra I research brief. American Institutes for Research (2014)
4. Smith, T.: Instructional coaching strategies to support student success in algebra I Research Brief. American Institutes for Research (2014)
5. Star, J.R., Caronongan, P., Foegen, A., Furgeson, J., Keating, B., Larson, M.R., Lyskawa, J., McCallum, W.G., Porath, J., Zbiek, R.M.: Teaching strategies for improving algebra knowledge in middle and high school students. Report, Institute of Education Sciences (2015)
6. Abeysekera, L., Dawson, P.: Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. High. Educ. Res. Dev. **34**, 1–14 (2015)
7. Lage, M.J., Platt, G.: The internet and the inverted classroom. J. Econ. Educ. **31**, 11 (2000)
8. O’Flaherty, J., Phillips, C.: The use of flipped classrooms in higher education: a scoping review. Internet High. Educ. **25**, 85–95 (2015)
9. Bishop, J.L., Verleger, M.: The flipped classroom : a survey of the research. In: Proceedings Annual Conference of the American Society for Engineering Education (2013)
10. Carney, D., Ormes, N., Swanson, R.: Partially flipped linear algebra: a team-based approach. Probl. Resour. Issues Math. Undergraduate Stud. **25**, 641–654 (2015)
11. Cilli-Turner, E.: Measuring learning outcomes and attitudes in a flipped introductory statistics course. Probl. Resour. Issues Math. Undergraduate Stud. **25**, 833–846 (2015)
12. McGivney-Burelle, J., Xue, F.: Flipping calculus. Probl. Resour. Issues Math. Undergraduate Stud. **23**, 477–486 (2013)
13. Muir, T., Geiger, V.: The affordances of using a flipped classroom approach in the teaching of mathematics: a case study of a grade 10 mathematics class. Math. Educ. Res. J. **28**(1), 149–171 (2015)

14. Love, B., Hodge, A., Grandgenett, N., Swift, A.W.: Student learning and perceptions in a flipped linear algebra course. Int. J. Math. Educ. Sci. Technol. **45**, 317–324 (2013)
15. McCallum, S., Schultz, J., Sellke, K., Spartz, J.: An examination of the flipped classroom approach on college student academic involvement. Int. J. Teach. Learn. High. Educ. **27**, 42–55 (2015)
16. Weng, P.: Developmental math, flipped and self-paced. Probl. Resour. Issues Math. Undergraduate Stud. **25**, 768–781 (2015)
17. Lape, N.K., Levy, R., Yong, D.H., Haushalter, K.A., Eddy, R., Hankel, N.: Probing the inverted classroom: a controlled study of teaching and learning outcomes in undergraduate engineering and mathematics. In: 121st ASEE Annual Conference and Exposition (2014)
18. Zack, L., Fuselier, J., Graham-Squire, A., Lamb, R., O'Hara, K.: Flipping freshman mathematics. Probl. Resour. Issues Math. Undergraduate Stud. **25**, 803–813 (2015)
19. Murphy, J., Chang, J.-M., Suaray, K.: Student performance and attitudes in a collaborative and flipped linear algebra course. Int. J. Math. Educ. Sci. Technol. **47**, 653–673 (2016)
20. Wilson, S.G.: The flipped class: a method to address the challenges of an undergraduate statistics course. Teach. Psychol. **40**, 193–199 (2013)
21. Young, A.: Flipping the calculus classroom: a cost-effective approach. Probl. Resour. Issues Math. Undergraduate Stud. **25**, 713–723 (2015)
22. Van Sickle, J.: Adventures in flipping college algebra. Probl. Resour. Issues Math. Undergraduate Stud. **25**, 600–613 (2015)
23. Guerrero, S., Beal, M., Lamb, C., Sonderegger, D., Baumgartel, D.: Flipping undergraduate finite mathematics: findings and implications. Probl. Resour. Issues Math. Undergraduate Stud. **25**, 814–832 (2015)
24. Strayer, J.F.: How learning in an inverted classroom influences cooperation, innovation and task orientation. Learn. Environ. Res. **15**, 171–193 (2012)
25. Chen, Y., Wang, Y., Chen, N.S.: Is FLIP enough? Or should we use the FLIPPED model instead? Comput. Educ. **79**, 16–27 (2014)
26. Touchton, M.: Flipping the classroom and student performance in advanced statistics: evidence from a quasi-experiment. J. Polit. Sci. Educ. **11**, 28–44 (2015)
27. Overmyer, J.: Research on flipping college algebra: lessons learned and practical advice for flipping multiple sections. Probl. Resour. Issues Math. Undergraduate Stud. **25**, 792–802 (2015)
28. Kirvan, R., Rakes, C.R., Zamora, R.: Flipping an algebra classroom: analyzing, modeling, and solving systems of linear equations. Comput. Schools **32**, 201–223 (2015)
29. McCammon, L.: Fizz method. http://lodgemccammon.com/flip/research/fizz-method/
30. Guo, P.J., Kim, J., Rubin, R.: How video production affects student engagement : an empirical study of MOOC videos. In: Proceedings of the first ACM Conference on Learning @ Scale Conference, pp. 41–50 (2014)
31. Tapia, M., Marsh, G.E.: An instrument to measure mathematics attitudes. Acad. Exch. Q. **8**, 16–21 (2004)
32. California Department of Education: California Standards Test. http://www.cde.ca.gov/ta/tg/sr/documents/cstrtqalgebra2.pdf

# Argumentation Identification for Academic Support in Undergraduate Writings

Jesús Miguel García Gorrostieta(✉) and Aurelio López-López

Instituto Nacional de Astrofísica, Óptica y Electrónica,
Luis Enrique Erro No. 1, Tonantzintla, Puebla, Mexico
{jesusmiguelgarcia,allopez}@inaoep.mx

**Abstract.** Argumentation in student research writings is needed to clearly communicate ideas and convince the reader of the presented claims. In this paper, we introduce a methodology to approach the analysis of argumentative writing in undergraduate research texts. We elaborate an annotation scheme to detect claims/premises and support/attack relations. An exploratory analysis was carried out to know the amount of argumentation in selected sections of theses. We analyze five types of argumentation (Authority, Example, Causal, Comparison and Analogy) in these sections. And we also explore the identification of arguments in paragraphs using machine learning techniques with lexical features, with encouraging results.

**Keywords:** Computer-assisted argument analysis · Argumentation studies · Academic writing · Corpus analysis

## 1 Introduction

Writing is a complex process that involves several stages such as planning, editing and reviewing. This process can be supported by computational tools of writing assessment which provide instructions to students in order to improve their writing skills, for example programs such as Criterion [1], Writing Pal W-Pal [2] and SWORD [3] are used in academic review of essays. As part of academic writing, we find argumentative writing which is used in student essays, scientific articles and theses to support presented claims with solid arguments. An argument is defined as a set of statements (premises) that individually or collectively provide support to a claim (conclusion). Some investigations in the literature have addressed the argumentative analysis in essays and scientific papers, such as Stab and Gurevych [4] and Kirschner et al. [5] who identify premises and conclusions, as well as their relations. However, we have not found work aimed to automatically analyze argumentation in larger academic works such as research proposals, theses or technical reports. Theses are often written at the end of college as one of the requirements for the degree, being in consequence quite important for students and academia. For this reason, a study in the analysis of argumentation in thesis is necessary to identify for example: What are the sections of the thesis containing more arguments?, What are the types of arguments used in these sections?, What are the types of argument components (i.e. premises or conclusions) used and their relations (e.g. support or attack)? How can be automatically identified


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 98–109, 2016.
DOI: 10.1007/978-3-319-45153-4_8

the arguments in thesis documents? To answer these questions is necessary a research in automatic arguments analysis for theses and research proposals, to support students in this difficult task.

In this paper, we give a first step toward the solution, presenting a model to guide the way we tackle the task. We also discuss an annotation scheme to create a corpus of academic text with recognized arguments, we annotate a sample and perform experiments to automatically identify argumentation in paragraphs, showing an efficacy similar to other approaches in other kinds of text.

The paper is structured as follows. In Sect. 2, we detail the related work for building annotation schemes and argument identification found so far. In Sect. 3, we explain our proposed methodology for automatic analysis of arguments. We discuss our argument annotation scheme used for corpus creation in Sect. 4. In Sect. 5, we present an exploratory analysis of the academic corpus. In Sect. 6, we report the result of argument identification in a part of the corpus. And finally in Sect. 7, we conclude with some final remarks and work in progress.

## 2 Related Work

The first step in the task of text analysis is to have an annotated corpus that allows us to validate the efficacy of the proposed method. As we found in the literature, the majority of researchers in the field of argument analysis create their own annotated corpus, using certain argumentative scheme. In the literature, we found few annotated argument corpus available. One of the most used among researchers to identify the presence of arguments is the Araucaria corpus [6], which has various types of documents (e.g. parliament records, newspapers, judicial summaries and forum discussions) with annotated premises, conclusions, and the argument scheme used, however the level of agreement between annotators is not reported in the study, turning it unreliable.

The creation of a corpus for each research is observed in different types of text, as well as in different domains. In Mochales and Moens [7], a corpus is built with 10 legal documents from the ECHR (European Court of Human Rights) corpus, with annotated premises and conclusions. In this corpus, the level of agreement between the two annotators is with a Kappa of 80 %, then in another study, Mochales and Moens [8] increased the number of annotators to three and the number of documents in the corpus to 47, with this, the level of agreement among annotators decreased to a Kappa of 75 %. It is important to note that dealing with legal texts with a well-defined structure facilitates the annotation process and increases the level of agreement. On the other hand, in the work of Stab and Gurevych [4], 90 persuasive essays on randomly-chosen topics are annotated by three participants. They annotated argumentative components with a level of agreement for the component of major conclusion of 83 % (position of the author), premises with 70 % and conclusions with 65 %; in addition the annotation for relations between argumentative components for attack is 80 % and support with 81 %, using Fleiss Kappa. In the research of Kirschner et al. [5], a corpus was created with 24 scientific articles in education for the sections of introduction and discussion. It was annotated by four participants with the argument components of premises and conclusions, as well as four relationships between these argumentative components (support, attack, sequence and detail), with an average in the level of agreement of Fleiss Kappa of 41 %. Therefore, it is observed that obtaining acceptable levels of annotation agreement in scientific texts is a complex task, which depends on an appropriate annotation guide and regularly monitoring of annotators during the corpus construction. For our research, the closest kind of document are scientific articles since theses or student research proposals share a similar structure, but are longer.

Once the corpus is built, it is necessary to perform the task of detecting the presence of arguments either in paragraphs, sentences or clauses. Moens et al. [9] perform an automatic identification of argumentative and non-argumentative sentences in Araucaria corpus. They represent sentences with features like combinations of pairs of words, verbs and text statistics using a naive Bayes classifier, achieving a 73.75 % of accuracy. In their investigation with legal texts in ECHR corpus [8], they reached an 80 % of accuracy. It is important to note that legal texts have a particular structure which allows lawyers to clearly identify the arguments. Another approach to identify presence of arguments in texts was reported by Florou [10], using a set of discourse markers and features based on mood and tense of verbs. They achieved an F1-Measure of 76.4 % using a decision tree classifier. Goudas et al. [11] performed identification of argumentative sentences, employing structural, lexical, contextual and grammatical features to represent each sentence. With a logistic regression classifier, they achieved an F1-Measure of 77.1 % on a corpus of 204 documents collected from social media written in Greek. They also performed identification of argument components (claim and premise). For this task, they applied a CRF (Conditional Random Field) classifier to obtain an F1-Measure of 42.37 %. Sardianos et al. [12] presented a similar approach with CRF and distributed representations of words to identify segments that correspond to argument components. For this task, they reported an F1-Measure of 32.21 %.

As mentioned before, after identifying a segment of text as argumentative, the next step is to find out the type of argumentative component (e.g. claim or premise). Stab and Gurevych [4] employed a SVM to classify propositions as non-argumentative, major claim, claim or premise, in academic essays. They used several structural, lexical, syntactic, and contextual features. They reported an accuracy of 77 %. Also Nguyen and Litman [13] performed the same argument component classification with SVM, they achieved a 79 % of accuracy, using argument and domain words extracted from unlabeled persuasive essays using LDA. Another approach to identify premises is applying techniques of sentiment analysis. Villalba and Saint-Dizier [14] identified discourse structures such as justification, elaboration and illustration that support opinions (evaluative expressions) in a corpus of hotel and restaurants reviews. They designed argument extraction rules with lexical features such as terms expressing polarity, adverbs of intensity, and domain verbs to identify discourse structures. They reported a precision of 92 % and recall of 86 % when identifying justifications.

In this paper we present our annotation scheme that includes the types of arguments most commonly observed in undergraduate academic proposals and theses, such as authority, causality, examples, analogy and comparison.

## 3 Approach to the Solution

Argumentation is based on philosophy and logic. An argument can be defined as a set of statements (e.g. sentences) that individually or collectively provide support to a claim [15]. The supported claim is called conclusion. There is only one conclusion for every argument, but there may be a series of statements of support. Statements that support a given conclusion are called premises. Among the theories of argumentation [16, 17], the consensus for the structure of an argument indicates that is composed of several argumentative components, which can be several premises and one conclusion. In this section, we present our general approach to identify argument components and relations for research writings.

Our approach to the solution relies on certain processes of the methodology used in argument mining [18]. In Fig. 1, we depict the process to analyze the student text. First, a text segmentation in minimum argumentation units is required; based on the preliminary revision applied to the corpus, the segmentation is done in terms of clauses. Secondly, it is necessary to classify each segment according to their argumentative role, either as a premise or a conclusion. Once the premises and conclusions of each paragraph are identified, we can assess the argumentative level of the text under consideration and identify the type of argumentation used in the paragraph. Finally, to identify the relations between argument components of a paragraph is necessary to pair premises and conclusions, to detect the kind of relation between them. Using such components and relations we can model the argument structure. Our challenge is to develop a method to successfully identify argumentative components (premises and

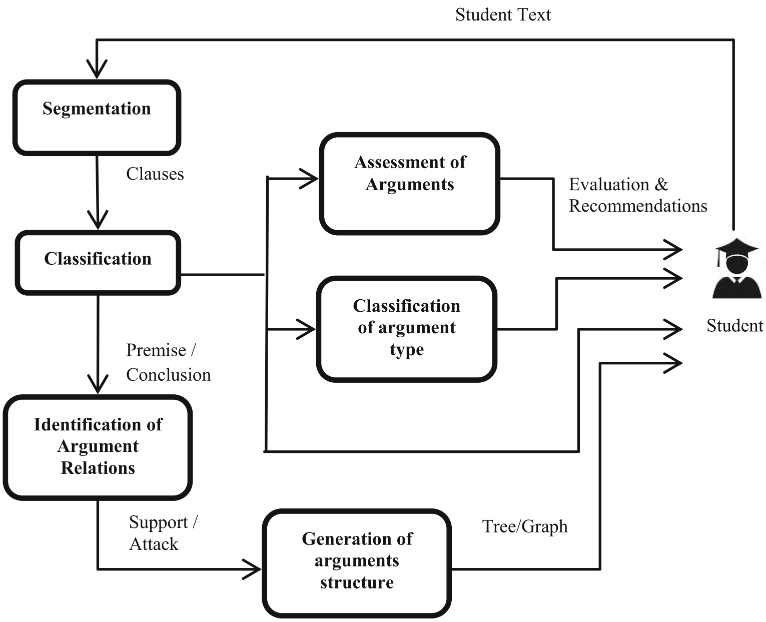



**Fig. 1.** Argument analysis model

conclusions), and their relations (attack and support), as well as the level and type of argumentation of each of the paragraphs. The ultimate aim is to provide an assessment along recommendations to support students in improving their argumentation in academic texts.

## 4 Argumentation Annotation Scheme

For our argumentation scheme, we consider two argument components: premises and conclusions, as well as two types of relation between components: support and attack. A graphical representation of the argument structure helps to understand how their components are interacting. This is done by identifying each premise and conclusion with a letter, which are associated with nodes of a graph. Then using directed arcs (arrows) to indicate a relationships between these components. The simple argument has only one premise that is used to support one conclusion [19]. For example:

*[Today educational institutions have a greater number of computers with Internet.] /P+ [**Therefore**, more students have access to the Internet.] /C*

As we can observe, the first sentence is the premise (in square brackets /P+) supporting the conclusion in the second sentence (in square brackets /C). In a simple argument, a premise provides elements to sustain the veracity of the associated conclusion. A diagram in Fig. 2 illustrates a simple argument in which the premise A supports the conclusion B. However, several other more elaborated structures can emerged from the analysis.

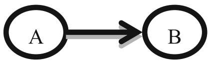



**Fig. 2.** Simple argument diagram

As we can notice in the example of argument, the word "therefore" plays an essential role in the identification of a possible conclusion; these words are called argumentative markers, and help us with the identification of elements in an argument.

To conduct the annotation study, we formulate an annotation guide. In this guide, we describe the different argumentative structures with their argument components (conclusion/premise) and their relations (attack/support). We also include the types of arguments and a score to establish the level of an argument. To guide the annotator, a set of examples taken from research proposals is included in each section. At the end of the guide, the annotation procedure is presented.

The procedure of the annotation guide includes the following steps. First the annotator needs to read the title and objective of the thesis or proposal. Then she has to identify if the text includes a conclusion or assertion. Next, she has to determine the ideas that support the conclusion and mark them as supporting premises. Also we advise the annotator to mark complete sentences or clauses as a conclusion or premise. To indicate a premise, square brackets are used by adding at the end /P, that is: [text of premise]/P. For the case of conclusions, in a similar way, the text was enclosed in square brackets ending with /C, i.e.: [text of conclusion]/C. Additionally, the annotator was asked to indicate the type of argument found in the paragraph, the types most commonly used by students in academic texts [20] are: authority, example, causal, analogy and comparison. Finally the annotator assesses paragraphs according to their level of argumentation, based on the score of Table 1.

**Table 1.** Argument assessment score

| Score | Scale | Description |
|---|---|---|
| 0 | None | There is no argument. |
| 1 | Weak | It is not a complete argument. Conclusion without a premise. |
| 2 | Medium | An argument with one reasoning. Conclusion with one premise. |
| 3 | Strong | An argument with one conclusion and two or more premises (support or attack) |

As shown in Table 1, a score of zero (0) is assigned to texts without any argumentation; this is the case for descriptions and definitions. And, when an argument is identified in the text, we use the next scale, if only the conclusion is found without any premises, the score of one (1) is assigned to the paragraph, if a conclusion and premise are located in the paragraph, a score of two (2) is assigned. If we found a conclusion and two or more premises in the paragraph, the highest score is assigned (3) that indicates the presence of strong argumentation. In this way, the paragraphs of the corpus were assessed.

The annotation will be performed by an instructor who has experience reviewing research proposals and theses. With this scheme, we show to the annotator how to identify the conclusions, premises and relations, and also, how to evaluate arguments in paragraphs. With the identification of such information, a writing support system can indicate students the weakness in argumentation or lack of conclusions or premises in their writings.

## 5 Corpus Analysis

Corpus analysis is performed to understand the argumentative characteristics in writings of undergraduates and graduate level. For this analysis, we used the corpus Coltypi [21] consisting of 468 theses and research proposals in the computer and information technologies domain, in Spanish. This corpus has undergraduate (TSU and Bachelor Degree) and graduate level (M. Sc. and Ph.D.) texts. According to [22], the sections of problem statement, justification and conclusions are considered highly argumentative so we focused the analysis on them.

By analyzing the corpus, it is observed that each section contains an average of 11 sentences. Each sentence contains 35 words on average with a total of 398 words per section. The size of sentences for the undergraduate level is 38 words per sentence which makes it difficult to read, in contrast to the doctoral level, having an average of 30 words. Based on this, we consider that doctoral writings are better and we can take them as reference.

In the work [23], the use of argumentative markers to identify argumentative text is reported with a precision of 89 % and recall of 4 %. Although with a low recall, precision is sufficient to give an idea of the presence of arguments in each section. Therefore, we use argumentative markers to identify the presence of arguments in paragraphs.

As shown in Table 2, most sections have more than half of the paragraphs with arguments. At Bachelor level, in the conclusion section, we have 446 paragraphs with arguments indicating a proportion of 61 %. Moreover, it is noted that in Ph.D. level in the justification section, there is a 61 % of paragraphs with arguments. For the TSU level, in section problem statement, 62 % of paragraphs with arguments is observed. With this analysis, a large number and proportion of arguments in our corpus is identified in the different academic levels.

**Table 2.** Paragraphs with arguments in the corpus

| | | With Argument | Total | Percentage |
|---|---|---|---|---|
| Ph.D. | Problem Stmt | 124 | 245 | 51 % |
| | **Justification** | **56** | 92 | **61 %** |
| | Conclusion | 194 | 449 | 43 % |
| Master | Problem Stmt | 206 | 392 | 53 % |
| | Justification | 203 | 375 | 54 % |
| | Conclusion | 707 | 1414 | 50 % |
| Bachelor | Problem Stmt | 150 | 269 | 56 % |
| | Justification | 180 | 313 | 58 % |
| | **Conclusion** | **446** | 731 | **61 %** |
| TSU | **Problem Stmt** | **95** | 153 | **62 %** |
| | Justification | 99 | 212 | 47 % |
| | Conclusion | 205 | 363 | 56 % |

For the analysis of the argument types, a sample of 46 theses was taken. From 224 paragraphs analyzed, we got that 127 are argumentative with a proportion of 56.7 %. Argument types identified in the sample were: authority, causality, examples, comparison, and analogy.

As presented in Table 3, the most abundant type is the causal argument with 95 paragraphs, followed by the arguments of authority with 20 paragraphs; we can observe this graphically in Fig. 3. This is because generally in academic texts, ideas are causally related and also based on citations of recognized authors in the area used as support.

By definition, authority arguments are based on the use of several references (authors, organizations) to explain what we need to know about the world [20]. These types of arguments are common in academia, because usually the author relies on publications produced by researchers in the field to support his assertions. An example of this type of argument is presented below, where we can notice a reference to an author which is used as premise (/P+) to support the conclusion (/C). For example:

**Table 3.** Argument types in sample

| | Paragraphs | Percentage |
|---|---|---|
| Authority | 20 | 14 % |
| Example | 18 | 13 % |
| **Causal** | **95** | **66 %** |
| Comparison | 3 | 2 % |
| Analogy | 6 | 4 % |

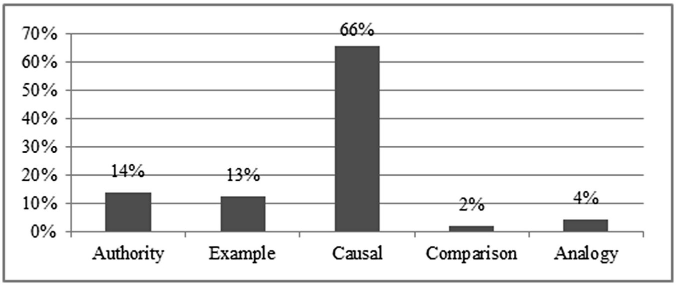



**Fig. 3.** Percentage by type of argument

*[Nowadays we are in the so-called Network Society] /C [**that according to Castells (2000)**, is a society that was generated from the technological revolution of information and from flourishing of social networks.] /P+*

## 6 Argument Identification

For experimentation, as detailed in previous section, a sample of 224 paragraphs were taken from Coltypi corpus and were manually annotated as argumentative or not argumentative by one annotator, an expert with knowledge in argument review and formulation. This classification done by the annotator is taken as our ground truth. As presented in Table 4, the proportion of paragraphs with arguments is 56.7 % and for paragraphs without arguments the remaining 43.3 %. The problem was approach as binary classification for each paragraph, i.e. to identify if it has arguments or not. To perform the validation, we took randomly 90 % of paragraphs for training and 10 % for testing. To determine the efficacy of each classifier, we use a 10-fold cross validation.

To perform the classification process, we employ the machine learning toolkit Weka [24]. Classifiers used are Support Vector Machine (SVM) [4], Naive Bayes (NB) [9] and J48 implementation of C4.5 Decision Tree (DT) [10], because these classifiers have been previously used in argument mining.

**Table 4.** Class Distribution Among Instances

| Paragraphs with arguments | Paragraphs without arguments |
|---|---|
| 127 (56.7 %) | 97 (43.3 %) |

Vector representations were used to identify paragraphs with arguments, mainly lexical with frequency of n-grams as consecutive words of length 1–3. We built the three vector representations for unigram, bigrams and trigrams which includes all words and punctuation symbols. Then we trained the classifiers with the dataset for training and applied them to the test dataset. Table 5 shows the average accuracy of each classifier for their 10 folds and for the three vector representations constructed. As we can notice, the SVM classifier produces better accuracy, precision, recall and F1-measure than NB and DT, identifying with 71.9 % accuracy and 81.28 % precision paragraphs with arguments.

**Table 5.** Classifiers results

| Unigrams | | | | |
|---|---|---|---|---|
| | **Precision** | **Recall** | **F1-Measure** | **Accuracy** |
| **Support Vector Machine** | **79.55 %** | **69.42 %** | **72.64 %** | **70.99 %** |
| Naïve Bayes | 73.27 % | 65.58 % | 68.35 % | 66.69 % |
| Decision Tree | 62.76 % | 56.15 % | 53.10 % | 56.48 % |
| +Bigrams (n-grams 1-2) | | | | |
| | **Precision** | **Recall** | **F1-Measure** | **Accuracy** |
| **Support Vector Machine** | **80.52 %** | **67.95 %** | **72.20 %** | **71.59 %** |
| Naïve Bayes | 71.99 % | 67.95 % | 68.75 % | 66.25 % |
| Decision Tree | 60.92 % | 55.38 % | 57.46 % | 55.01 % |
| +Trigrams (n-grams 1-3) | | | | |
| | **Precision** | **Recall** | **F1-Measure** | **Accuracy** |
| **Support Vector Machine** | **81.26 %** | **66.99 %** | **72.44 %** | **71.90 %** |
| Naïve Bayes | 74.67 % | 63.91 % | 67.80 % | 65.78 % |
| Decision Tree | 62.18 % | 56.86 % | 58.88 % | 56.31 % |

As we can observe in Table 6 and Fig. 4, the SVM classifier has better accuracy when using trigrams of words with 71.9 %, and a standard deviation of 6.6, which is lower compared to unigram (9.3) and bigrams (8.3).

**Table 6.** Mean and standard deviation of SVM accuracy

| | Mean | Stdev |
|---|---|---|
| Unigram | 71.0 | 9.3 |
| +Bigram | 71.6 | 8.3 |
| **+Trigram** | **71.9** | **6.6** |

Another vector representation used for this task is constructed with unigrams, bigrams (pairs of successive words with a minimum frequency of 5 instances) and trigrams (three successive words with a minimum frequency of 5 instances) with all words and punctuation symbols. Then, we trained the classifiers with the dataset for training and applied them to the test dataset. Table 7 shows the average accuracy of

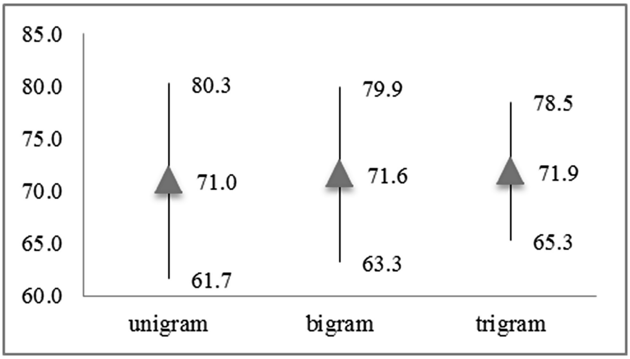



**Fig. 4.** Mean and standard deviation of SVM accuracy

**Table 7.** Classifiers results

| Unigrams + Bigrams (Freq. $\geq$ 5) + Trigrams (Freq. $\geq$ 5) | | | | |
|---|---|---|---|---|
| | Precision | Recall | F1-Measure | Accuracy |
| **Support Vector Machine** | **85.65 %** | **70.19 %** | **75.76 %** | **75.55 %** |
| Naïve Bayes | 74.69 % | 66.28 % | 69.26 % | 67.08 % |
| Decision Tree | 61.08 % | 55.32 % | 57.47 % | 55.45 % |

each classifier for their 10 folds. As we can observe, the three classifiers improve and again SVM classifier obtains the best accuracy with 75.55 and 85.65 % precision, identifying paragraphs with arguments.

## 7 Conclusion

We have described our annotation scheme for arguments in academic research proposals and we have applied it on a small sample. Our goal is to create a free available annotated corpus of research proposals in computer science to support research on argumentation. We are in the process of creating such corpus with the support of two annotators.

As we can observe, there is a sufficient amount of arguments in academic work (research proposals and theses), with the analysis of the corpus we realized that more than half of the paragraphs written by undergraduate students include arguments, so it is important to make further progress in building a system that supports the assessment of this kind of academic texts.

According to the results, the best accuracy was observed in our preliminary experiments to identify paragraphs with arguments by the SVM classifier using lexical features, i.e. n-grams of words of length up to 3 with frequency equal or higher than five. In future work, we plan to explore the use of structural and syntactical features to improve the classification task.

In addition, we will continue tackling the other subtasks in our argument analysis model, specifically we expect to achieve an adequate representation for argument components and relations in academic texts.

**Acknowledgments.** We thank the annotator Tania Maria Tequida Castillo for the assistance in the corpus creation. The first author was partially supported by CONACYT, México, under scholarship 357381. The second author was partially supported by SNI, México.

## References

1. Burstein, J., Chodorow, M., Leacock, C. CriterionSM online essay evaluation: an application for automated evaluation of student essays. In: IAAI, pp. 3–10 (2003)
2. Roscoe, R.D., Allen, L.K., Weston, J.L., Crossley, S.A., McNamara, D.S.: The Writing Pal intelligent tutoring system: usability testing and development. Comput. Compos. **34**, 39–59 (2014)
3. Cho, K., Schunn, C.D.: Scaffolded writing and rewriting in the discipline: a web-based reciprocal peer review system. Comput. Educ. **48**(3), 409–426 (2007)
4. Stab, C., Gurevych, I.: Identifying argumentative discourse structures in persuasive essays. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 46–56 (2014)
5. Kirschner, C., Eckle-Kohler, J., Gurevych, I.: Linking the thoughts: analysis of argumentation structures in scientific publications. In: Proceedings of the 2nd Workshop on Argumentation Mining, pp. 1–11 (2015)
6. Katzav, J., Reed, C.A., Rowe, G.W.: Argument research corpus. In: Huget, M.-P. (ed.) Communication in Multiagent Systems. Lecture Notes in Computer Science, pp. 269–283. Springer Verlag, Berlin (2004)
7. Mochales, R., Moens, M.F.: Study on the structure of argumentation in case law. Front. Artif. Intell. Appl. **189**(1), 11–20 (2008)
8. Mochales, R., Moens, M.F.: Argumentation mining. Artif. Intell. Law **19**(1), 1–22 (2011)
9. Moens, M.F., Boiy, E., Mochales R., Reed, C.: Automatic detection of arguments in legal texts. In: Proceedings of the 11th International Conference on Artificial Intelligence and Law, pp. 225–230. ACM (2007)
10. Florou, E., Konstantopoulos, S., Koukourikos, A., Karampiperis, P.: Argument extraction for supporting public policy formulation. In: Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, pp. 49–54 (2013)
11. Goudas, T., Louizos, C., Petasis, G., Karkaletsis, V.: Argument extraction from news, blogs, and social media. In: Likas, A., Blekas, K., Kalles, D. (eds.) SETN 2014. LNCS, vol. 8445, pp. 287–299. Springer, Heidelberg (2014)
12. Sardianos, C., Katakis, I.M., Petasis, G., Karkaletsis, V.: Argument extraction from news. In: NAACL HLT 2015, p. 56 (2015)
13. Nguyen, H., Litman, D.: Extracting argument and domain words for identifying argument components in texts. In: Proceedings of the 2nd Workshop on Argumentation Mining, pp. 22–28 (2015)
14. Villalba, M.P.G., Saint-Dizier, P.: Some facets of argument mining for opinion analysis. COMMA **245**, 23–34 (2012)
15. Capaldi, N.: Cómo Ganar una Discusión. Gedisa, Barcelona (2000)
16. Toulmin, S.E.: The uses of argument. Cambridge University Press, England (1958)

17. Walton, D., Reed, C., Macagno, F.: Argumentation Schemes. Cambridge University Press, Cambridge (2008)
18. Peldszus, A., Stede, M.: From argument diagrams to argumentation mining in texts. Int. J. Cogn. Inform. Natural Intell. **7**(1), 1–31 (2013)
19. Walton, D.: Fundamentals of Critical Argumentation. Cambridge University Press, Cambridge (2005)
20. Weston, A.: Las claves de la argumentación. Ariel, Barcelona (1994)
21. González-López, S., López-López, A.: Colección de tesis y propuesta de investigación en TICs: un recurso para su análisis y estudio. XIII Congreso Nacional de Investigación Educativa, pp. 1–15 (2015)
22. López, C.: La argumentación en los géneros académicos. In: Actas del Congreso Internacional La Argumentación, pp. 1–11. Universidad de Buenos Aires, Buenos Aires (2003)
23. Lawrence, J., Reed, C.: Combining argument mining techniques. In: Proceedings of the 2nd Workshop on Argumentation Mining, NAACL HLT 2015, pp. 127–136 (2015)
24. Hall, M., Eibe, F., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explor. **11**(1), 10–18 (2009)

# Mobile Grading Paper-Based Programming Exams: Automatic Semantic Partial Credit Assignment Approach

I-Han Hsiao(✉)

School of Computing, Informatics and Decision Systems Engineering, Arizona State University, 699 S. Mill Ave., Tempe, AZ, USA
Sharon.Hsiao@asu.edu

**Abstract.** In this paper, we report a study of an innovative mobile application to support grading paper-based programming exams. We call the app – Programming Grading Assistant (PGA). It scans pre-generated QR-codes of paper-based question-and-concepts associations and uses OCR to recognize handwritten answers. PGA provides interfaces for teachers to calibrate recognition results, as well as to adjust partial credit assignment according to conceptual incorrectness of the answers. We evaluate the mobile grading process and the quality of grading results based on the assessed semantic information. The results demonstrate that the mobile grading approach keeps persistent traces of students' performance, including semantic feedback and ultimately enhances grading consistency.

## 1 Introduction

Today, the majority of programing classes are delivered via a blended instructional method with face-to-face instruction in the classrooms supported by online tools such as intelligent tutors, self-assessment quizzes, and course management systems etc. Such a blended instructional strategy, in contrast to pure on-line learning/instruction through massive open online courses, which still has inconclusive results [1, 2], allows teachers to focus on systematically instructing complex topics in class while supplying many supplemental exercises outside the classroom. The blended instructional classrooms still mainly rely on paper-based exams as the main method to assess students' knowledge in today's lower division programming courses. It is very challenging for teachers to provide personalized feedback on each individual test. The large size of the classes makes it impractical to discuss with each individual student on his/her exam paper. Instead, typically, teachers discuss on the returned exam in the class (hopefully thorough and detailed enough to cover all the students' misconceptions).

Although teachers still point out the common mistakes and try to pinpoint the key concepts related to the such mistakes, many desired detailed learning analytics are unavailable, such as *how did s/he receive partial credits*, *was it a single concept or multiple concepts mistake*, *a careless mistake or a long-term misconception* etc. As a result, students often focus solely on the scores they earned on the returned exams, but miss several learning opportunities, such as *identification of strength and weakness*,


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 110–123, 2016.
DOI: 10.1007/978-3-319-45153-4_9

*characterization of the nature of their errors or any recurring patterns* if any, *assessment of appropriateness of their study strategies and preparation*, etc. Furthermore, from teachers' perspective, there is an increasing difficulty in managing paper-based exams. Teachers can hardly memorize all students' names, it is becoming even more challenging for teachers to manage all mistake patterns in students' exam answers. Thus, it is common for teachers to focus on common mistakes based on the history of a course rather than a specific exam. Moreover, with graders or teaching assistants being recruited to do the majority of grading in order to handle large classes, teachers could overlook the detailed course performances. In this case, varying level of training of the graders and potential inconsistency among graders are among additional factors that may further complicate students' learning.

In this work, we investigate an innovative method to capture paper-based programming exams in order to provide semantic personalized feedback in supporting large-scale auto-grading in blended instruction classrooms. We emphasize that paper-based approach is still one of the primary and preferred ways of delivering programming exams, due to the sake of easiness and other logistics or potential academic dishonesty issues to occur in online settings. The rest of the paper is organized with reviewing related work in Automatic Program Assessment and Technology Support for Blended Classroom in Computing Education. We then describe the study methodology and layout the study design. Finally, we present the evaluation results and discuss the approach with current limitations and future work.

## 2 Literature Review

### 2.1 Automatic Program Assessment

Automatic program evaluation is not a new topic. Special interest group on computer science education (SIGCSE) reports several work on automatic grading students' programming assignments in last couple decades. For instance, WEB-CAT [3], BOSS [4], ASSYST [5] and among many others. The common approach is to apply pattern-matching techniques to verify students' answers and the correct answers. Most of these systems are web-based evaluation tools; less is emphasized on automatic evaluation for paper-based programming formal assessments. There are a few relevant early innovations attempting to apply to process paper exams and hand written codes, such as tablet grading system [6, 7]. It uses tabletop scanners to digitalize the exam papers and provides a central grading interface on the tablet in assisting mass programming grading. It reports that a few benefits of digitizing paper exams (i.e. some default feedback can be kept on the digital pages; student's identity can be kept anonymous and potentially prevent graders' bias in recognizing names). There is also an adjacent related work attempted to address scaling up assessment production, it is called parameterized exercises. Parameterized questions and exercises use randomly generated parameters in every question template and produce many similar, yet sufficiently different questions. This approach not only automatically evaluates students' programs, but also dramatically reduces authoring efforts and creates a sizeable collection of questions to facilitate programming assessment. As demonstrated by a number of projects such as CAPA [8], WebAssign [9], QuizPack [10], and QuizJET [11] parameterized questions can be used effectively in a number of domains allowing to increase the number of assessment items, decrease authoring efforts and reduce cheating. Overall, the field of automatic program evaluation is less focused on grading paper-based programming problems, therefore, less support of personalization as such.

### 2.2 Technology and Instructional Support for Blended Classroom in Computing Education

In the field of Computer Supported Collaborative Learning (CSCL), researchers describe classroom orchestration as a field in transition, which defines how a teacher manages multilayered activities in real time and in a multi-constraints context. It discusses how and what research-based technologies have been adopted and should be done in classrooms [12]. We have begun to see more tabletops, wearable cameras, smart classrooms or interactive tools such as Classroom Response Systems (AKA: Clickers) etc. provide dynamic feedback and integrative students knowledge updates [13–15]. Such tools attempt to capture in-the-moment teaching pace and on–the-fly students' learning pace; however, they are usually highly customized to the content or require a large collection of content for teachers to start using the tools. Classes may still suffer from lacking comprehensive content collections due to high development or maintenance costs or run into *off-sync* issues when the students leave the interactive classrooms. Over the last decade, several educational technologies and instructional pedagogies have been proposed and studied to amend and assist large size of classrooms. For instance, flipped classroom model to promote utilization of class time for interaction [16]; peer instruction to facilitate students' conceptual reasoning [17, 18]; media computation to increase learning motivation [19, 20] etc. In the context of computing education, a dozen of research projects attempted to apply these methods in programming classes [21–24]. In spite of there were positive results reported, most of the findings were still in early stage and inconclusive. For example, instructors are supposed to utilize the class time to maximize student and teacher interactions, however, until today, it is still challenging to gather and interact with large amount of students with laptops in the lecture hall or in the computer lab [25].

## 3 Methodology

### 3.1 Mobile Grading Framework

In order to automatically evaluate programming problems on paper-based exams, we create a mobile PGA grading framework (Fig. 1) and develop an instance of android application[1] based on the framework. (1) Using a camera-enabled mobile device to scan questions, which are attached with pre-generated Quick Response codes (Fig. 2). The scanning can be done at a batch process by scanning multiple questions and

[1] The mobile grading app is currently available upon request.

multiple exams at a time before entering to auto-grading phase; (2) The grading service begins to process the scanned questions as images, which includes: (a) Converting pixel images to binary images; (b) Removing noises from the image in order to just focus on texts in the recognition step, but not to falsely remove the punctuations, which is considered an important element in code writing; (c) Defining character boundaries for later recognition to calculate word separation and alignments. (3) We deploy an open source OCR (Optical Characteristic Recognition) library[2] to recognize hand written and/or printed texts from the scanned images; (4) The app then compares the recognized answers to the correct answers; (5) The app assigns scores based on two grading schemes: (a) a binary function of correct or incorrect answer; (b) a partial credit assignment based on the proportion of recognized concepts to the overall concepts of the correct answer; (6) Finally, the app aggregates step 1 to 5 results to generate reports and updates analytics.

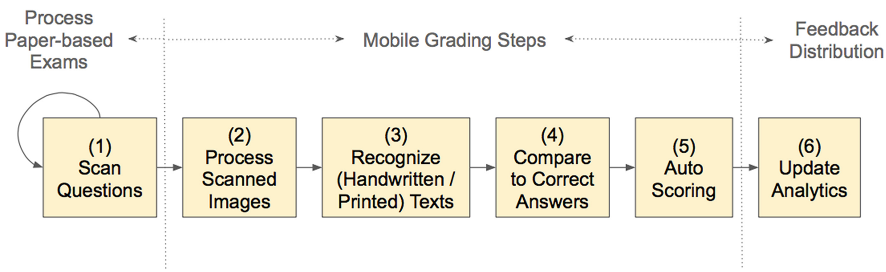



**Fig. 1.** Mobile grading framework.

### 3.2 Research Platform: Programming Grading Assistant

We develop an Android application and deploy it in a Samsung Galaxy Tab S2 8 device to support grading paper-based exams. Currently, we consider two major types of programming problems on the exams, multiple choice question (MCQ) and code writing question (CWQ). Due to the handwritten texts naturally carry various complications, such as scripts and prints mixture, word separation ambiguity, inconsistent word alignments etc. Based on the unique characteristics of these two types of questions (MCQ consists of limited handwriting texts and printed question texts; CWQ mainly includes substantial amount of handwriting texts and limited printed texts), we design two separate mobile grading modules to deal with different level of handwriting recognition complexity. In this paper, we focus on the CWQ module, which allows us to focus on the proposed technology in examining procedural knowledge of programming problem solving.

**CWQ Grading Module.** This module utilizes 2-dimensional quick response code (QR-code) to associate each question, answer, corresponding concepts and weights (the importance of the concepts). Figure 2 demonstrates the CWQ grading interface, where

[2] https://github.com/tesseract-ocr.

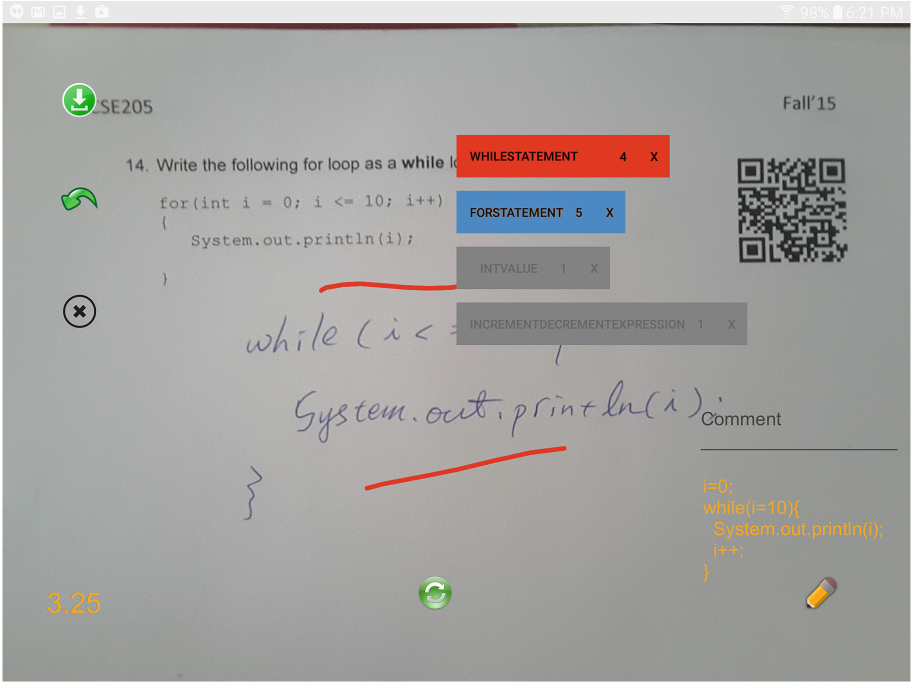



**Fig. 2.** Code writing question grading interface.

the grader can keep notes anywhere on the screen and to leave free-form feedback or just simply highlight the missing/incorrect codes with her fingertip. The pencil icon located at the lower right is the functionality to edit misconceptions. Grading is done by tapping on the concepts, which resembles the action of punishing misconceptions of incorrect codes. The grades are shown at the lower left corner. The graders can also manually adjust the grades as appropriately. Once graders are done with editing, they can press the top-left corner save icon. All the graded questions will be recorded into database and be exported as an xml format file to be ready to feed in learning analytics dashboard tool, such as EduAnalysis [26].

**Semantic Feedback.** All questions' associated concepts are parsed and indexed by Java Programming ontology[3] and weights are automatically indexed by EduAnalysis [26]. On toggling on and off the pencil icon, the associated concepts of the question will be brought out to the screen. The grader can tap to edit the weights to indicate the missing concepts and/or misconceptions. In Fig. 2, the red highlighted concept indicates the *misconception*, blue shows the *gained concept*, and grey shows *missing concepts*. In this CWQ example, the student has clearly missed to initialize a counter variable *i* (`int i=0`) and the *increment statement* (`i++`) in the *while* loop. Therefore, the *IntVariable* and *IncrementDecrementExpression* concepts are greyed out.

[3] Source of Java Ontology: http://www.pitt.edu/~paws//ont/java.owl.

Such grading process not only leaves conceptual feedback for each question, but also allows automatic partial credit computation (will be discussed in next sub-section).

### 3.3 Recognition Optimization

Students' handwritings are heterogeneous. Typically, OCR requires a training process to calibrate recognition. However, due to the target corpus in this experiment is programming language domain, we anticipate students will only be writing code syntactical texts. In other words, all recognized texts should be identified from Java glossary. For instance, Fig. 3(a) shows the OCR recognized results from the same example in Fig. 2. However, without proper training, the recognition fails to identify variable *i* or the *print* method. In order to minimize the training effort, we adopt spelling correction logic and implement *recognition correction* algorithm. We use *Damerau–Levenshtein* distance algorithm [27] to iteratively transpose, replace and insert the recognized characters from simultaneously referencing to Java glossary dictionary. In Fig. 3(b) illustrates a corrected recognition codes, which improves the readability of original recognition. Note that the punctuations and variables are still not yet optimized.

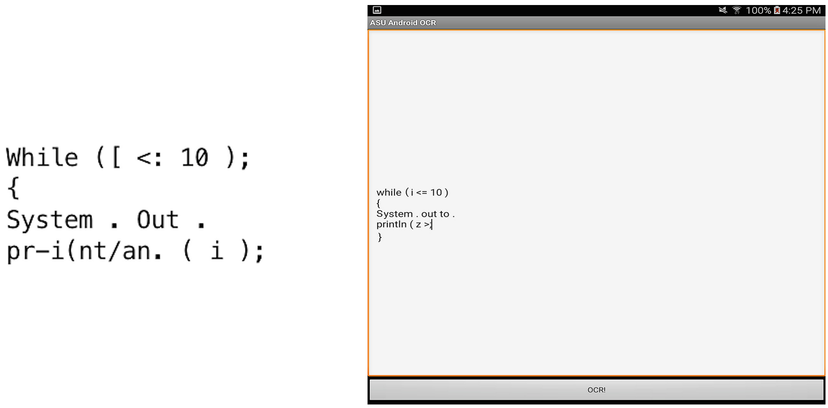



**Fig. 3.** Before (a) and after (b) optimized hand writing recognition.

### 3.4 Semantic Partial Credit Algorithm

Assigning grades on the programming exams is not a trivia task. Each code-writing question can potentially have multiple solutions; each solution can have miscellaneous variations, such as local variable differences, utility methods, interfaces etc. Therefore, it may not be fair to assign scores by judging codes similarity between students' and teacher's codes. Typically, teachers will evaluate the code solutions and assign partial credits to award the logic soundness instead of code completeness. For instance, a common strategy is to give points to conceptual integrity and deduct points by punishing conceptual mistakes. The question is, *how many points are appropriate as partial credits?*

We discover there often exists the inconsistency in assigning partial credits. Figure 4 illustrates some of the inconsistent scenarios: (a-left) The student clearly implemented key concepts *ArrayList* and *Foreach-loop*, but missed to aggregate values from each iteration and to print out the final results. However, this student was being punished by missing minor concepts and suffered from major credits loss (5 out of 7). In case (a-right) shows the same grader gave same partial credits to another student, while s/he not only missed out the *sum* variable declaration and initialization, but actually wrongly implemented the *Foreach-loop*. In case (b), it shows a different grader gave different amount of partial credits on the exactly the same question. These examples demonstrate a series of grading inconsistency issues: *oversight on key concept misconception*, *over-emphasized on minor concepts*, *limited feedback* etc.

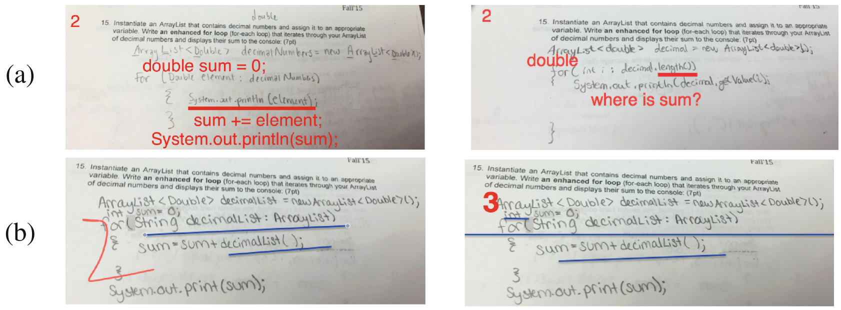



**Fig. 4.** (a) left & right are two different students' answers that were graded by the same grader; (b) left & right are the same student's answer that were graded by two different graders.

In order to enforce consistency in assigning partial credits, we design a semantic partial credit algorithm to calculate students' proportional conceptual errors of a question (Table 1). There are 3 parameters to determine the partial credits: *Concept Similarity*, *Concept Saliency*, and *Miscellaneous*. We assume partial credit would be given based on conceptual inconsistency with the correct answer, therefore, *Concept Similarity* is calculated by the cosine similarity between student's and the correct one. We use *Concept Saliency* coefficient (Eq. 1) to highlight the importance of key concept and to demote peripheral conceptual mistakes. For instance, the question in Fig. 4 (a-left) deserves more credits when the key concepts are intact, and the peripheral concepts are missed; vice versa in Fig. 4(a-right). Finally, we reserve a *Miscellaneous* coefficient to capture all other mistakes that are not conceptual, such as careless mistakes.

## 3.5 Study Setup

We design a study to investigate paper-based programming exams grading process, specifically focus on the semantic partial credit assignment effects. We randomly sampled 20 students' exams from an Object-Oriented Programming & Data Structure

**Table 1.** Semantic partial credit algorithm

**function** pc = **partial_credit**(question){
**if** concepts incorrect **then**
  pc = ConceptSimilarity$_{Tester\&Correct}$ * ConceptSaliency;
  $\text{ConceptSaliency} = w_{c,q} / \sum_i w_{c_i,q}$ (Eq.1)
  **if** pc<0 **then** return pc = 0;
**else**
  pc = 1 – $\varepsilon_{Miscellaneous}$;
  **if** pc<0 **then** return pc = 0;
}
  **return pc;**

class in 2015 Fall semester offered in Arizona State University. We recruited six graders who have been graders or teaching assistants for the same course at least once. Among the recruited 6 graders, there are 3 graduate students and 3 undergraduate students, 1 female and 5 male. All of them are either Computer Science major or Information Science major. They have 1 ~ 5 years Java programming experiences and have taken 3 ~ 8 programming courses. In addition, they all code multiple all-purposes programming languages on a daily basis (mainly, C, C++, and PHP).

**Data Collection.** We scan two questions and answers from the sampled 20 paper-based exams and use photo editing software to remove the original grading remarks from the scans. Thus, there are 40 questions (2 questions × 20 different students' exams) in total. Exam questions are presented in Table 2.

**Table 2.** Sampled exam questions & answer keys/grading scheme

| Question | Answers & Grading Scheme |
|---|---|
| 1. Write the following for loop as a **while** loop: (5pt)<br>`for(int i = 0; i <= 10; i++)`<br>`{`<br>`System.out.println(i);`<br>`}` | `// full score must be exactly the same codes`<br>`// as the correct answer (except variable names)`<br>`int i = 0;          // -1 incorrect initialization`<br>`while (i<=10){      //-3 incorrect while statement`<br>`System.out.println(i);`<br>`i++;          // -1 incorrect increment statement`<br>`}`<br>`// -0.5 other errors` |
| 2. Instantiate an ArrayList that contains decimal numbers and assign it to an appropriate variable. Write **an enhanced for loop** (for-each loop) that iterates through your ArrayList of decimal numbers and displays their sum to the console: (7pt) | `// -2 incorrect ArrayList`<br>`ArrayList<Double> numList = new ArrayList<Double>();`<br>`double sum = 0;`<br>`// -4 incorrect enhanced for-loop statement`<br>`for(Double d: numList){`<br>`sum += d;`<br>`}`<br>`System.out.println(sum);`<br>`// -0.5 other errors` |

**Study Procedure.** In the lab study, we instructed graders to refer to the provided solution keys and grading schemes (Table 2) and assign grades based on their best judgment to all 40 questions. Noted that graders graded on the same 40 questions. The grading scheme was solicited from the same teacher who designed the exam. Graders were also instructed to mark or leave feedback as appropriately. They spent 10 ∼ 33 min to finish grading all the questions.

## 4 Evaluation Results

### 4.1 Semantic Partial Credit Accuracy

We compared all graders' grading and automatic partial credit algorithm calculation results to original teacher's grading. We set the threshold at 0.1 marks; where we consider it is a correct answer when the grades differences are less than the threshold. We found that given the grading schemes for graders (per Table 2), they could still make considerably inaccurate grading outcomes. The inaccuracy effect is especially noticeable in Q2, which is a more complex question than Q1. Q2 involves more key concepts and more overall concepts. Recall that the case of Fig. 4-(a), the student clearly knew how to implement *ArrayList* and *Foreach Loop*, but the grader penalized this answer as missing a lot codes, and neglected the grading scheme. Figure 5 illustrates all graders CWQ grading accuracy distribution by question compared to automatic method. Overall, we found that automatic partial credit algorithm improved 20% of the accuracy.

Additionally, We found that the complex question (Q2) consistently being taken more than 20 points off in total, and averagely each question was mis-graded respectively 0.658 and 1.491 in Q1 and Q2 (Table 3). Meanwhile, the automatic partial credit algorithm achieved low grades discrepancy in both Q1 and Q2. On average, they were graded only 0.122 and 0.370 off compared to teacher's grades. Such results suggested that if an exam consists of 10 code-writing questions, the variance between grader and teacher could be as large as one letter grade.

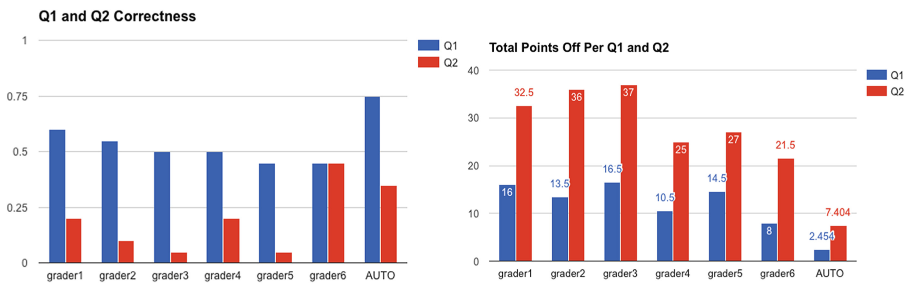



**Fig. 5.** Partial Credit Accuracy (left); Grades Discrency by Question Complexities (right).

**Table 3.** Grading discrepancy magnitude

| *Average* | Grader | Auto PC Algorithm |
|---|---|---|
| Q1 | 0.658 | 0.122 |
| Q2 | 1.491 | 0.370 |

### 4.2 Mobile Grading Enhances Grading Coherence

To gauge graders' coherence, we evaluated how did the graders grade questions; *do they give extra points than they should have?* or *do they penalize the students in deducting more points than they should have?* Figure 6 showed each grader's inconsistency. We found that grader 1, 3, 5, and 6 tend to give more credits; grader 2 and 4 tended to be stricter and give fewer credits than they should have. The gaps among graders are evident. We found that automatic partial credit algorithm achieved higher grading coherence (smaller gap between + & −).

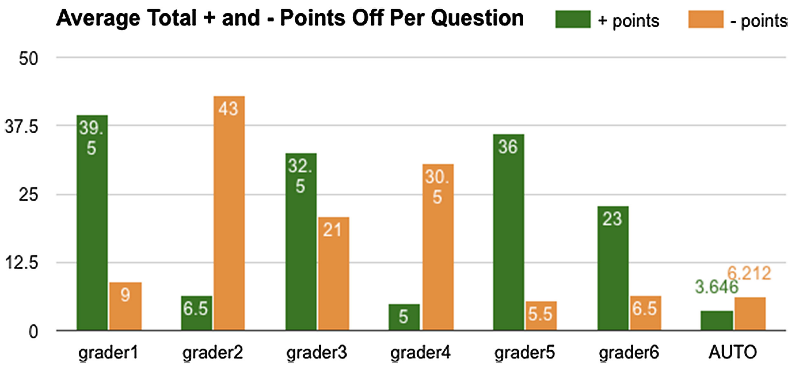



**Fig. 6.** Inconsistencies Among Graders, some graders graded loosely, some strictly.

In addition, noted that currently the auto-grading algorithm tends to give fewer points to students (-0.155 points on average), which means PGA is rather to grade slightly harshly than mercifully. On the other hand, graders give mixed signals in grading; either half point more or half point less (Table 4). Such results show prominent news for teachers, who may consider partial incorrectness as incorrect rather than giving false positive grading and mislead students.

**Table 4.** Grading discrepancy magnitude

| *Average* | Grader | Auto PC Algorithm |
|---|---|---|
| + | 0.594 | 0.091 |
| − | 0.481 | 0.155 |

### 4.3 Feedback Quality

We analyze graders' feedback on total 180 questions not-entirely-correct questions (30 out of 40 sampled questions per grader). We categorize six types of graders' feedback: *No feedback at all*, *Highlights on students' errors*, *(Partial) correct answers*, *Justifications on penalty*, *Conceptual feedback*, and *Wrong feedback* (Fig. 7).

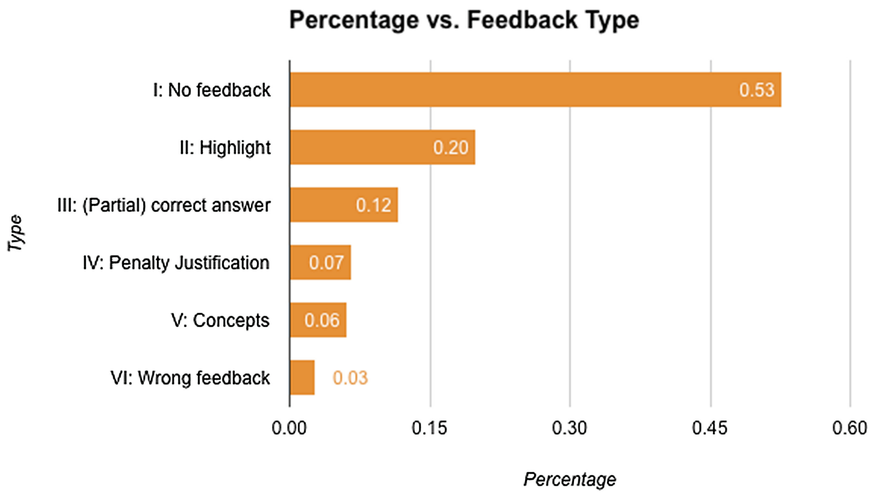



**Fig. 7.** Feedback types and percentage.

We found that the majority of the questions (52.8%) did not receive feedback from the graders at all. The cases are often occurred when the students had completely wrong implementation. They usually receive *a big red cross* and *zero* along with the question and nothing else (Fig. 8 Left). These students are usually the ones who have none or incomplete knowledge and demand for more support. However, they tend not to get any feedback at all on the paper-based exams. Nonetheless, the second type of feedback

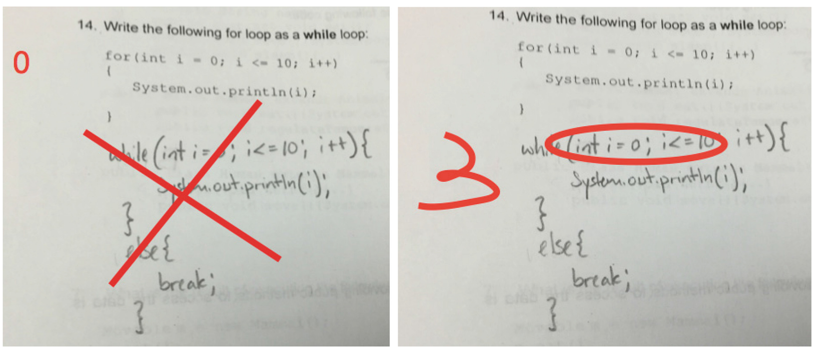



**Fig. 8.** Left: No feedback at all; Right: Highlights on errors.

is to highlight students' errors (20.0%) (Fig. 8 Right). In such scenario, students could potentially obtain point-of-interests to focus on mistakes, but no further guidance. Unfortunately, these two types of feedback are not only shallow, but also very common strategies for grading paper-based exams.

Type III feedback is that graders directly write down the answers or partial answers on students' exams (11.7%). For example, Fig. 3(a). Type IV feedback lists the reasons why there are points of deduction (6.7%), for instance, graders left comments on the exams *"your sum is not computed", "results is not displayed", "your output is misplaced", "where is sum?"* Type V explains the misconceptions semantics (6.1%), for example, *"Wrong while condition", and "no initialization"*. These three types (Type III ~ V) are considered more substantial feedback. However, in the context of learning, the correct solutions may not necessarily the best next steps for all learners. It is harder to provide personalized feedback on paper-based exams, due to the lack of understanding on students other learning performances. Finally, Type VI shows a 2.8% of messages are actually wrong feedback to the students.

## 5 Conclusions

### 5.1 Summary

In this work, we design and evaluate an innovative mobile application to investigate automatic grading paper-based programming exams. We call it Programming Grading Assistant (PGA), it utilizes mobile's inbuilt-camera to scan question and answers. We use OCR technology to recognize students' handwriting answers and design interfaces to calibrate recognition or log misconceptions. We use 2-dimensional quick response code (QR-Code) to associate each code-writing question, answer, and their corresponding concepts and importance. Based on semantics associations with the exam content, a partial credit assignment algorithm is constructed to leverage grading inconsistency and to be utilized to provide semantic feedback.

Study results show that human graders exhibit multiple grading inconsistencies and provide insufficient and shallow feedback. Meanwhile, PGA not only elevates the grading consistency, but also systematically assigns partial credits and improves the grading coherence. It also reveals human graders provide insufficient feedback, while the proposed approach provides consistent semantics remarks as feedback. In addition, handwriting recognition is currently not optimized, but can be improved with recognition correction logic. Overall, PGA's auto-grading framework via mobile devices shows promising results in capturing paper-based programming exams for advanced learning analytics.

### 5.2 Limitations and Future Work

In spite of several promising findings, there are a few limitations in current study. First of all, the handwriting recognition requires *good* lighting (i.e. natural sun light) and ball pen writing. In our experiment, we found that indoor lighting often resulted in recognition failure, which was also one of the reasons that took slightly longer time than we expected. In addition, penciled texts of students' answers also resulted in recognition failure. However, programming exams typically require iterative problem solving and trial-and-error, thus, students usually prefer to use pencils than pens. We recognize these challenges with OCR technology, and have begun to instruct students to write their code carefully by following sound Object-Oriented Programming principles and coding conventions.

Secondly, we did not measure codes recognition accuracy yet, since the recognition is not yet optimized. It is in our research agenda, to expedite fully automatic grading process and to reach reliable consistent grading outcome. We have begun training to use designated underscores for uppercases and whitespaces to increase word separation recognition.

In the near future, we anticipate providing personalized feedback to students from their paper-based exams. We are currently developing APIs to synchronize grading results to the learning analytics dashboards [26]. We plan to conduct more user studies and larger scale of field studies to explore PGA with graders in UI related issues and measure the effects for grading entire class exams.

## References


1. Zhenghao, C., et al.: Who's Benefiting from MOOCs, and Why. Harvard Business Review (2015)
2. Kolowich, S.: Puts MOOC Project with Udacity on Hold, in The Chronicle of Higher Education (2013)
3. Edwards, S.H., Perez-Quinones, M.A.: Web-CAT: automatically grading programming assignments. In: ACM SIGCSE Bulletin. ACM (2008)
4. Joy, M., Griffiths, N., Boyatt, R.: The boss online submission and assessment system. J. Educ. Res. Comput. (JERIC) **5**(3), 2 (2005)
5. Jackson, D., Usher, M.: Grading student programs using ASSYST. In: ACM SIGCSE Bulletin. ACM (1997)
6. Bloomfield, A., Groves, J.F.: A tablet-based paper exam grading system. In: ACM SIGCSE Bulletin. ACM (2008)
7. Bloomfield, A.: Evolution of a digital paper exam grading system. In: 2010 IEEE Frontiers in Education Conference (FIE). IEEE (2010)
8. Kashy, E., et al.: Using networked tools to enhance student success rates in large classes. In: Proceedings of the 27th Annual Conference Frontiers in Education Conference, 1997. Teaching and Learning in an Era of Change. IEEE (1997)
9. Titus, A.P., Martin, L.W., Beichner, R.J.: Web-based testing in physics education: methods and opportunities. Comput. Phys. **12**(2), 117–123 (1998)
10. Brusilovsky, P., Sosnovsky, S.: Individualized exercises for self-assessment of programming knowledge: an evaluation of QuizPACK. J. Educ. Res. Comput. (JERIC) **5**(3), 6 (2005)
11. Hsiao, I.-H., Sosnovsky, S., Brusilovsky, P.: Guiding students to the right questions: adaptive navigation support in an E-Learning system for Java programming. J. Comput. Assist. Learn. **26**(4), 270–283 (2010)
12. Dillenbourg, P.: Design for classroom orchestration. Comput. Educ. **69**, 485–492 (2013)

13. Martinez-Maldonado, R., et al.: Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. Int. J. Comput. Support. Collaborative Learn. **8**(4), 455–485 (2013)
14. Roschelle, J., Penuel, W.R., Abrahamson, L.: Classroom response and communication systems: research review and theory. In: Annual Meeting of the American Educational Research Association (AERA). San Diego, CA, pp. 1–8 (2004)
15. Slotta, J.D., Tissenbaum, M., Lui, M.: Orchestrating of complex inquiry: three roles for learning analytics in a smart classroom infrastructure. In: Proceedings of the Third International Conference on Learning Analytics and Knowledge. ACM (2013)
16. Bishop, J., Verleger, M.: The flipped classroom: a survey of the research. In: 120th ASEE Annual Conference & Exposition. Atlanta, GA (2013)
17. Crouch, C.H., Mazur, E.: Peer instruction: ten years of experience and results. Am. J. Phys. **69**(9), 970–977 (2001)
18. Fagen, A.P., Crouch, C.H., Mazur, E.: Peer instruction: results from a range of classrooms. Phys. Teach. **40**(4), 206–209 (2002)
19. Guzdial, M.: Exploring hypotheses about media computation. In: Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research. ACM (2013)
20. Porter, L., et al.: Success in introductory programming: what works? Commun. ACM **56**(8), 34–36 (2013)
21. Simon, B., et al.: Experience report: peer instruction in introductory computing. In: Proceedings of the 41st ACM Technical Symposium on Computer Science Education. ACM (2010)
22. Simon, B., et al.: Experience report: CS1 for majors with media computation. In: Proceedings of the Fifteenth Annual Conference on Innovation and Technology in Computer Science Education. ACM (2010)
23. Sarawagi, N.: Flipping an introductory programming course: yes you can! J. Comput. Sci. Coll. **28**(6), 186–188 (2013)
24. Amresh, A., Carberry, A.R., Femiani, J.: Evaluating the effectiveness of flipped classrooms for teaching CS1. In: 2013 IEEE Frontiers in Education Conference. IEEE (2013)
25. Rosiene, C., Rosiene, J.: Flipping a programming course: the good, the bad, and the ugly. In: Frontiers in Education Conference. IEEE (2015)
26. Hsiao, I.-H., Govindarajan, S.K.P., Lin, Y.-L.: Semantic visual analytics for today's programming classrooms. In: The 6th International Learning Analytics and Knowledge Conference. ACM, Edinburgh, UK (2016)
27. Brill, E., Moore, R.C.: An improved error model for noisy channel spelling correction. In: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics (2000)

# Which Algorithms Suit Which Learning Environments? A Comparative Study of Recommender Systems in TEL

Simone Kopeinik(✉), Dominik Kowald, and Elisabeth Lex

Knowledge Technologies Institute, Graz University of Technology, Graz, Austria
{simone.kopeinik,elisabeth.lex}@tugraz.at, dkowald@know-center.at

**Abstract.** In recent years, a number of recommendation algorithms have been proposed to help learners find suitable learning resources online. Next to user-centered evaluations, offline-datasets have been used to investigate new recommendation algorithms or variations of collaborative filtering approaches. However, a more extensive study comparing a variety of recommendation strategies on multiple TEL datasets is missing. In this work, we contribute with a data-driven study of recommendation strategies in TEL to shed light on their suitability for TEL datasets. To that end, we evaluate six state-of-the-art recommendation algorithms for tag and resource recommendations on six empirical datasets: a dataset from European Schoolnets TravelWell, a dataset from the MACE portal, which features access to meta-data-enriched learning resources from the field of architecture, two datasets from the social bookmarking systems BibSonomy and CiteULike, a MOOC dataset from the KDD challenge 2015, and Aposdle, a small-scale workplace learning dataset. We highlight strengths and shortcomings of the discussed recommendation algorithms and their applicability to the TEL datasets. Our results demonstrate that the performance of the algorithms strongly depends on the properties and characteristics of the particular dataset. However, we also find a strong correlation between the average number of users per resource and the algorithm performance. A tag recommender evaluation experiment reveals that a hybrid combination of a cognitive-inspired and a popularity-based approach consistently performs best on all TEL datasets we utilized in our study.

**Keywords:** Offline study · Tag recommendation · Resource recommendation · Recommender systems · ACT-R · SUSTAIN · Technology enhanced learning · TEL

## 1 Introduction

Recommender systems have grown to become one of the most popular research fields in personalized e-learning. A tremendous amount of contributions has been presented and investigated over its last fifteen years of existence [1]. However, up to now there are no generally suggested or commonly applied recommender


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 124–138, 2016.
DOI: 10.1007/978-3-319-45153-4_10

system implementations for TEL environments. In fact, the majority of holistic educational recommender systems remain within research labs [2]. This may be partly attributed to the fact that proposed recommendation approaches often require either runtime-intensive computations or unavailable, expensive information about learning domains, resources and learner preferences. Furthermore, in informal learning settings, information like ontologies, learning object meta-data and even user ratings are very limited [3]. The spectrum of commonly available tracked learner activities varies greatly, but typically includes implicit usage data like learner-ids, some general information on learning resources, timestamps and indications of a user's interest in learning resources (e.g. opening, downloading or bookmarking) [4]. While existing research investigates the application of implicit usage data-based algorithms (e.g., [5–7]) on selected datasets, a more extensive comparative study directly opposing state-of-the-art recommendation algorithms is still missing. We believe such a study would benefit the community since we hypothesize that recommendation algorithms show different performance results depending on learning context and dataset properties as also suggested in [5,8]. This motivates our main research question: ***RQ1:*** *How accurate do state-of-the-art resource recommendation algorithms, using only implicit usage data, perform on different TEL datasets?*

To this end, we collected six datasets from different TEL domains such as social bookmarking, social learning environments, Massive Open Online Courses (MOOCs) and workplace learning to evaluate accuracy and ranking of six state-of-the-art recommendation algorithms. Results show a strong correlation between the average number of users per resource and the performance of most investigated algorithms. Further, we believe that content-based algorithms that match user characteristics with resource properties, could present an alternative for informal environments with sparse user-resource matrices. However, a prominent factor that hampers the finding and recommending of learning resources is the lack of learning object meta-data, which is resource-intensive to generate. Bateman et al. [9] proposed the application of tagging mechanisms to shift this task to the crowd. Furthermore, tag recommendations can assist and motivate the user in providing such semantic meta-data. Also, tagging supports the learning process, as it is known to foster reflection and deep learning [10]. Yet, so far, tag recommender investigations have been widely unattended in the TEL research community [11]. To this strand, we want to contribute with our second research question: ***RQ2:*** *Which computationally inexpensive state-of-the-art tag recommendation algorithm performs best on TEL datasets?*

The evaluation of three recommendation algorithms, implemented as six variations based on usage data and hybrid combinations, identifies a cognitive-inspired recommendation algorithm combined with a popularity-based approach as most successful.

## 2 Related Work

In general, there already exists a large body of research on recommender systems in the context of TEL, see e.g., [1,3,12]. Surveys like for example [11] additionally discuss the potential of collaborative tagging environments and tag recommender systems for TEL. From the wide range of existing contributions, we identify two lines of research that are most related to our work, (i) data driven studies of tag recommendations and (ii) learning resource recommendations in the field of TEL.

### 2.1 Learning Resource Recommendations

Verbert et al. [5] studied the influence of different similarity measures on user- and item-based collaborative filtering for the prediction of user ratings. Additionally, they compared user-based collaborative filtering on implicit learner data, among four different datasets and first used and analyzed the prominent TEL datasets, TravelWell and MACE. Fazeli et al. [6] showed that the integration of social interaction can improve collaborative filtering approaches in TEL environments. Niemann and Wolpers [7] investigated the usage context of learning objects as a similarity measure to predict and fill in missing user ratings and subsequently improve the database for other recommendation algorithms such as collaborative filtering. The approach is evaluated in a rating prediction setting. The suggested approach does not require any content information of learning objects and thus could also be applied to cold start users, but not cold start items. For further research examples, a broad overview on data-driven learning recommender studies is given in [13]. In contrast to previous work, we do not focus on a specific algorithm or dataset but we study the performance of a range of recommendation algorithms on various TEL datasets.

### 2.2 Tag Recommendations

Considerable experiments exploring learning resource annotation through tags are presented in [14], in which generally the suitability of tagging within the learning context was investigated. Results claim guidance to be an important factor for the success of tagging. Diaz et al. [15] investigated automated tagging of learning objects utilizing a computationally expensive variant of Latent Dirichlet Allocation [16] and evaluated the tagging predictions in a user study. In [17], an approach to automatically tag learning objects based on their usage context was introduced, which builds on [7]. It shows promising results towards the retrospective enhancement of learning object meta-data. However, their approach cannot be used in online settings as it is based on context information of resources that is extracted from user sessions. In this work, we concentrate on tag recommendation algorithms that are applicable also in online settings.

## 3 Evaluation

In this work, we evaluate six recommendation algorithms in terms of performance on six TEL datasets from different application areas such as social bookmarking systems (BibSonomy, CiteULike), MOOCs (KDD15), open social learning

(MACE, TravelWell) and workplace learning (Aposdle). We evaluate two recommender application cases, (i) the recommendation of learning resources to support finding relevant information and (ii) the recommendation of tags to support the annotation of learning resources.

### 3.1 Methodology

For evaluation, we split each dataset into a training and a test set, following a common evaluation protocol used in recommender systems research [18,19]. To predict the future based on the past, each user's activities are sorted in chronological order by the timestamp the activities were traced in the systems. For the tag recommender evaluation, we put the latest post of a user (i.e. all tags assigned by a user to a resource) into the test set and the remaining posts of this user into the training set (see [18]). When evaluating resource recommendations, this process slightly differs. We select 20 % of most recent activities of a user for testing and the remains for training (see [19]). Also, to ensure that there is enough training data available per user, we only consider users with at least five available activities. For the tag recommender test sets, we only consider users with at least two available posts. This procedure avoids a biased evaluation as no data is deleted from the original datasets.

### 3.2 Algorithms

For the purpose of this study, we selected well-established, computationally inexpensive tag and resource recommendation strategies (for a more substantial review on complexity please see [20]) as well as approaches that have been proposed and discussed in the context of TEL. All algorithms of this study as well as the evaluation methods are implemented in Java as part of our *TagRec* recommender benchmarking framework [21], which is freely available via GitHub[1].

**Most Popular (MP).** MP is a simple approach to rank items according to their frequency of occurrence [22]. The algorithm can be implemented on user-based, resource-based or group-based occurrences and is labeled respectively, as $MP_U$, $MP_R$ and MP. $MP_{U,R}$ describes a linear combination of $MP_U$ and $MP_R$.

**Collaborative Filtering (CF).** This approach calculates the neighborhood of users ($CF_U$) or resources ($CF_R$) to find items that are new to a user by either considering items that similar users engaged with or items that are similar to resources the target user engaged with in the past [23]. The neighborhood is defined by the $k$ most similar users or resources, calculated by the cosine-similarity measure on the binary user-resource matrix. Tag recommendations require the triple: (user, resource, tag). Therefore, we implemented an adaptation of $CF_U$ for tag recommendations [24]. Accordingly, the neighborhood of a user is determined through a user's tag assignments instead of resource engagements. As suggested by literature [25], we set $k$ to 20 for all CF implementations.

[1] https://github.com/learning-layers/TagRec/.

**Content-Based Filtering (CB).** CB recommendation algorithms rate the usefulness of items by determining the similarity between an item's content with the target user profile [26]. In this study, we either use topics (if available) or otherwise tags to describe the item content. The similarity between the item vector and the user vector is calculated by the cosine-similarity measure.

**Usage Context-Based Similarity (UCbSim).** This algorithm was introduced by [27] and further discussed in the TEL context by [7,28]. The approach is inspired by paradigmatic relations known in lexicology, where the usage context of a word is defined by the sequence of words occurring before or after it in the context of a sentence. The equivalent to a sentence in online activities is defined as a user session, which describes the usage context. In line with literature [7], we calculate the significant co-occurrence of two items $i$ and $j$ by the mutual information (MI):

$$MI_{i,j} = log_2 \frac{O}{E} \tag{1}$$

where $O$ is the number of observed co-occurrences and $E$ the number of expected co-occurrences. The similarity ($sim_{i,j}$) between two objects is given by their cosine-similarity, where each object is described as a vector of its 25 highest ranked co-occurrences. For this study, we recommend resources that are most similar to the resources a user engaged with in her last session. Further, we conclude a session if no user interaction is observed for 180 min.

**Base Level Learning Equation with Associative Component (BLL$_{AC}$).** This cognitive-inspired tag recommendation algorithm, mimics retrieval from human semantic memory. A detailed description and evaluation can be found in [29]. It is based on equations from the ACT-R architecture [30] that model the availability of elements in a person's declarative memory as activation levels $A_i$. Equation 2 comprises the base-level activation $B_i$ and an associative component that represents semantic context. To model the semantic context we look at the tags other users have assigned to a given resource, with $W_j$ representing the frequency of appearance of a $tag_j$ and with $S_{ji}$ representing the normalized co-occurrence of $tag_i$ and $tag_j$, as an estimate of the tags' strength of association.

$$A_i = B_i + \sum_j W_j S_{ji} \tag{2}$$

With $B_i = ln(\sum_{j=1}^{n} t_j^{-d})$, we estimate how useful an item (tag) has been in an individual person's past, with $n$ determining the frequency of tag use in the past, and $t_j$ representing recency, i.e., the time since a tag has been used for the $j^{th}$ time. The parameter $d$ models the power law of forgetting and is in line with [30] set to .5. We select the most relevant tags according to the highest activation values. As BLL$_{AC}$+MP$_R$, we denote a linear combination of this approach with MP$_R$.

**SUSTAIN.** SUSTAIN [31] is a cognitive model aiming to mimic humans' category learning behavior. In line with [19], which suggested and analyzed the model to boost collaborative filtering, we implemented the first two layers, which depict an unsupervised clustering mechanism that maps inputs (e.g., resource features) to outputs (e.g., activation values that decide to select or leave a resource).

In the initial training phase, each user's personal attentional tunings and cluster representations are created. The number of clusters per user evolves incrementally through the training process (i.e., a new cluster is only recruited if a new resource cannot be assimilated with the already existing clusters). As input features describing a resource, we select either topics (if available) or tags. The total number of possible input features determines the clusters' dimension. Further, the clustering algorithm has three tunable parameters, which we set in line with [31] as follows: attentional focus $r = 9.998$, learning rate $\eta = .096$ and threshold $\tau = .5$, where the threshold specifies the sensitivity to new cluster creation. The resulting user model is then applied to predict new resources from a candidate set that is given by the 100 highest ranked resources according to $\mathrm{CF}_U$. For the prediction, we calculate and rank an activation value for each resource given by the highest activated cluster in the user model and select the most relevant items accordingly. As SUSTAIN+$\mathrm{CF}_U$, we denote a linear normalized combination of SUSTAIN and $\mathrm{CF}_U$.

### 3.3 Datasets

Table 1 summarizes the dataset properties such as posts, users, resources, tags, topics and their relations, as descriptive statistics. For the purpose of this study, we use *sparsity* to designate the percentage of resources that are not described by topics or tags. A more elaborate presentation of the datasets follows.

**BibSonomy.** The university of Kassel provides SQL dumps[2] of the open social bookmarking and resource sharing system *BibSonomy*, in which users can share and tag bookmarks and bibliographic references. Available are four log data files that report users' tag assignments, bookmark data, bibliographic entries and tag to tag relations. Since topics are not allocated [32], we used the tag assignment data, which was retrieved in 2015.

**CiteULike.** CiteULike is a social bookmarking system for managing and discovering scholarly articles. Since 2007, CiteULike datasets[3] are published on a regular basis. The dataset for this study was retrieved in 2013 (resource recommendation dataset) and 2015 (tag recommendation dataset). Three log data files report on users' posting of articles, bibliographic references, and group membership of users. Activation data of user posts, including tags, have been used for this study. Topics are not available.

[2] http://www.kde.cs.uni-kassel.de/bibsonomy/dumps/.

[3] http://www.citeulike.org/faq/data.adp.

**KDD15.** This dataset origins from the KDD Cup 2015[4], where the challenge was to predict dropouts in Massive Open Online Courses (MOOCs). The MOOC learning platform was founded in 2013 by Tsinghua University and hosts more than 360 Chinese and international courses. Data encompasses course dates and structures (courses are segmented into modules and categories), student enrollments and dropouts and student events. For the purpose of this study, we filtered the event types *problem*, *video* and *access* that indicate a student's learning resource engagement. There are no tags in this dataset but we classify categories as topics.

**Table 1. Properties of the six datasets that were used in our study.** $|P|$ depicts the number of posts, $|U|$ the number of users, $|R|$ the number of resources, $|T|$ the number of tags, $|Tp|$ the number of topics, $|AT_r|$ the average number of tags a user assigned to one resource, $|ATp_r|$ the average number of topics describing one resource, $|AR_u|$ the average number of resources a user interacted with, $|AU_r|$ the average number of users that interacted with a specific resource. The last two parameters $\text{SP}_t$ and $\text{SP}_{tp}$ describe the sparsity of tags and topics, respectively.

| | $\|P\|$ | $\|U\|$ | $\|R\|$ | $\|T\|$ | $\|Tp\|$ | $\|AT_r\|$ | $\|ATp_r\|$ | $\|AR_u\|$ | $\|AU_r\|$ | $\text{SP}_t$ | $\text{SP}_{tp}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BibSonomy | 82539 | 2437 | 28000 | 30889 | 0 | 4.1 | 0 | 33.8 | 3 | 0 | 100 |
| CiteULike | 105333 | 7182 | 42320 | 46060 | 0 | 3.5 | 0 | 14.7 | 2.5 | 0 | 100 |
| KDD15 | 262330 | 15236 | 5315 | 0 | 3160 | 0 | 1.8 | 17.2 | 49.4 | 100 | 1.1 |
| TravelWell | 2572 | 97 | 1890 | 4156 | 153 | 3.5 | 1.7 | 26.5 | 1.4 | 3.2 | 28.7 |
| MACE | 23017 | 627 | 12360 | 15249 | 0[a] | 2.4 | 0 | 36.7 | 1.9 | 31.2 | 100 |
| Aposdle | 449 | 6 | 430 | 0 | 98 | 0 | 1.1 | 74.8 | 1 | 100 | 0 |

[a]Generally the dataset contains topics but unfortunately, at this point, we do not have them available.

**MACE.** In the MACE project an informal learning platform was created that links different repositories from all over Europe to provide access to meta-data-enriched learning resources from the architecture domain. The dataset encompasses user activities like the accessing and tagging of learning resources and additional learning resource descriptions such as topics and competences [33]. At this point, unfortunately, we do not posses access to competence and topic data. However, user's accessing of learning resources and tagging behavior were used in our study.

**TravelWell.** Originating from the Learning Resource Exchange platform[5], the dataset captures teachers search for and access of open educational resources from a variety of providers all over Europe. Thus, it covers multiple languages and subject domains. Activities in the dataset are supplied in two files with

[4] http://kddcup2015.com/information.html.
[5] http://lreforschools.eun.org.

either bookmarks or ratings which both include additional information about the learning resource [34]. Relevant information to our study encompasses user names, resource names, timestamps, tags and categories.

**Aposdle.** An adaptive work integrated learning system that origins from the Aposdle EU project. The target user group are workers from the innovation and knowledge management domain. The dataset origins from a workplace evaluation that also included a context-aware resource recommender. Three files with user activities, learning resource descriptions with topics but no tags and a domain ontology were published [35]. The very small dataset has only six users. For the purpose of our evaluation study, we considered the user actions *VIEW_RESOURCE* and *EDIT_ANNOTATION* as indications for learning resource engagements.

### 3.4 Metrics

For the performance evaluation of the selected recommendation algorithms (MP, CF, CB, UCbSim, BLL, Sustain) we use the further described metrics recall, precision and f-measure, which are commonly used in recommender system research [5,36]. Additionally, we look at nDCG, which was reported to be the most suitable metric for evaluations of item ranking [37].

When calculating recall and precision, we determine the relation of recommended items $\hat{I}_u$ for a user $u$, to items that are of a user's interest $I_u$. Items relevant to a user are determined by the test set. All metrics are averaged over the number of considered users in the test set.

**Recall.** Recall (R) indicates the proportion of the $k$ recommended items that are relevant to a user (i.e., correctly recommended items), to all items relevant to a user.

$$R@k = \frac{|I_u \cap \hat{I}_u|}{|I_u|} \tag{3}$$

**Precision.** The precision (P) metric indicates the proportion of the $k$ recommended items that are relevant to a user.

$$P@k = \frac{|I_u \cap \hat{I}_u|}{|\hat{I}_u|} \tag{4}$$

**F-measure.** The F-measure (F) calculates the harmonic mean of recall and precision. This is relevant as recall and precision normally do not develop symmetrically.

$$F@k = 2 \cdot \frac{(P@k \cdot R@k)}{(P@k + R@k)} \tag{5}$$

**nDCG.** Discounted Cumulative Gain (DCG) is a ranking quality metric that calculates usefulness scores (gains) of items based on their relevance and position in a list of $k$ recommended items and is calculated by

$$DCG@k = \sum_{i=1}^{k}\left(\frac{2^{B(i)}-1}{log_2(1+i)}\right) \tag{6}$$

where $B(i)$ is 1 if the $i^{th}$ recommended item is relevant and 0 if not. To allow comparability of recommended lists with different item counts, the metric is normalized. nDCG is calculated as DCG divided by the ideal DCG value iDCG, which is the highest possible DCG value that can be achieved if all relevant items are recommended in the correct order, formulated as $nDCG@k = \frac{DCG@k}{iDCG@k}$.

# 4 Results and Discussion

This section presents our results in terms of prediction accuracy (R, P, F) and ranking (nDCG). Six algorithms with a total of thirteen variations were applied on six TEL datasets from different learning settings. We consider metrics @5 as most relevant, as this seems to be a reasonable number of items to confront a learner with. Additionally, we report F@10 and nDCG@10. To best simulate real-life settings, we conducted the study on the unfiltered datasets.

## 4.1 Learning Resource Recommendations (*RQ1*)

In line with [5] who compared the performance of CF on different TEL datasets, we observe that the algorithms' performance values strongly depend on the dataset and its characteristics. Solely $CF_U$ shows a stable behavior over all datasets. As expected, the performance of $CF_U$ is related to the average number of resources a user interacted with. The SUSTAIN algorithm, which re-ranks the 100 best rated $CF_U$ values, uses categories of a user's resources to construct learning clusters. Hence, the extent of the resource's descriptive features (we either use topics or tags, if topics are not available) is crucial to the success of the algorithm. Comparing our results of Table 2 with the dataset statistics of Table 1, we find that an average of at least three features per resource is needed to improve the performance of $CF_U$. Similarly, a poor performance of $CF_R$ is reported for MACE, TravelWell and Aposdle, where the average number of users per resource is lower than two. MP as the simplest approach performs widely poor, except for MACE, where it almost competes with the more complex $CF_U$. This may relate to the number of learning domains covered by a learning environment. MACE is the only learning environment that is restricted to one subject, namely *architecture*.

The importance of a dense user resource matrix is underlined by our results. In fact, we find a strong correlation of .958 (t = 19.5502, df = 34, p-value $<$ 2.2e$-$16) between the average number of users per resource ($|AU_r|$) (see Table 1)

**Table 2. Results of our resource recommender evaluation.** The accuracy estimates are organized per dataset and algorithm (*RQ1*). The datasets BibSonomy, CiteULike and MACE did not include topic information, thus for those three, we calculated $CB_T$ and SUSTAIN on tags instead of topics. *Note:* the highest accuracy values per dataset are highlighted in bold.

| Dataset | Metric | MP | $CF_R$ | $CB_T$ | $CF_U$ | UCbSim | SUSTAIN | SUSTAIN + $CF_U$ |
|---|---|---|---|---|---|---|---|---|
| BibSonomy | R@5 | .0073 | .0447 | .0300 | .0444 | .0404 | .0396 | **.0530** |
| | P@5 | .0154 | .0336 | .0197 | .0410 | .0336 | .0336 | **.0467** |
| | F@5 | .0099 | .0383 | .0238 | .0426 | .0367 | .0363 | **.0496** |
| | F@10 | .0102 | .0380 | .0226 | .0420 | .0351 | .0374 | **.0497** |
| | nDCG@5 | .0088 | .0416 | .0270 | .0440 | .0371 | .0392 | **.0541** |
| | nDCG@10 | .0103 | .0490 | .0313 | .0509 | .0440 | .0469 | **.0629** |
| CiteULike | R@5 | .0051 | **.0839** | .0472 | .0567 | .0716 | .0734 | .0786 |
| | P@5 | .0048 | **.0592** | .0353 | .0412 | .0558 | .0503 | .0553 |
| | F@5 | .0050 | **.0694** | .0404 | .0477 | .0627 | .0597 | .0650 |
| | F@10 | .0042 | .0601 | .0362 | .0488 | .0573 | .0530 | **.0618** |
| | nDCG@5 | .0048 | **.0792** | .0427 | .0511 | .0686 | .0704 | .0717 |
| | nDCG@10 | .0054 | **.0901** | .0504 | .0635 | .0802 | .0815 | .0863 |
| KDD15 | R@5 | .0067 | **.4774** | .1885 | .4325 | .4663 | .3992 | .4289 |
| | P@5 | .0018 | .2488 | .1409 | .2355 | **.2570** | .2436 | .2377 |
| | F@5 | .0029 | .3074 | .1612 | .3050 | **.3314** | .3025 | .3059 |
| | F@10 | .0034 | .2581 | .1244 | .2773 | **.3195** | .2756 | .2769 |
| | nDCG@5 | .0053 | **.3897** | .1927 | .3618 | .3529 | .3227 | .3608 |
| | nDCG@10 | .0081 | **.4740** | .2090 | .4281 | .4465 | .3939 | .4284 |
| TravelWell | R@5 | .0035 | .0257 | .0174 | .0404 | .0471 | **.0483** | .0139 |
| | P@5 | .0127 | .0212 | .0382 | **.0425** | .0297 | .0382 | .0382 |
| | F@5 | .0056 | .0232 | .0240 | .0414 | .0365 | **.0427** | .0204 |
| | F@10 | .0078 | .0194 | .0304 | .0456 | .0459 | **.0481** | .0429 |
| | nDCG@5 | .0072 | .0220 | .0275 | .0305 | **.0491** | .0446 | .0220 |
| | nDCG@10 | .0092 | .0239 | .0353 | .0461 | **.0631** | .0544 | .0405 |
| MACE | R@5 | .0253 | .0080 | .0016 | **.0283** | .0151 | .0093 | .0222 |
| | P@5 | .0167 | .0079 | .0023 | **.0251** | .0213 | .0065 | .0190 |
| | F@5 | .0201 | .0079 | .0019 | **.0266** | .0177 | .0076 | .0205 |
| | F@10 | .0169 | .0116 | .0031 | **.0286** | .0189 | .0155 | .0241 |
| | nDCG@5 | .0248 | .0082 | .0014 | **.0264** | .0165 | .0079 | .0215 |
| | nDCG@10 | .0281 | .0136 | .0026 | **.0357** | .0282 | .0157 | .0302 |
| Aposdle | R@5 | .0 | .0 | .0 | **.0026** | .0 | .0 | .0 |
| | P@5 | .0 | .0 | .0 | **.0333** | .0 | .0 | .0 |
| | F@5 | .0 | .0 | .0 | **.0049** | .0 | .0 | .0 |
| | F@10 | **.0196** | .0 | .0151 | .0045 | .0 | .0045 | .0045 |
| | nDCG@5 | .0 | .0 | .0 | **.0042** | .0 | .0 | .0 |
| | nDCG@10 | **.0152** | .0 | .0103 | .0042 | .0 | .0036 | .0033 |

and the performance (F@5) of all considered algorithms but MP. This is especially visible when comparing KDD15 ($|AU_r| = 49.4$) and Aposdle ($|AU_r| = 1$). KDD15 is our only MOOC dataset. It differs predominantly through its density but also through the structural nature of the learning environment, where each course is hierarchically organized in modules, categories and learning resources.

**Table 3. Results of our tag recommender evaluation**. We see that the cognitive-inspired $BLL_{AC} + MP_R$ clearly outperforms its competitors (*RQ2*). *Note*: the highest accuracy values per dataset are highlighted in bold.

| Dataset | Metric | $MP_U$ | $MP_R$ | $MP_{U,R}$ | $CF_U$ | $BLL_{AC}$ | $BLL_{AC} + MP_R$ |
|---|---|---|---|---|---|---|---|
| BibSonomy | R@5 | .3486 | .0862 | .3839 | .3530 | .3809 | **.4071** |
| | P@5 | .1991 | .0572 | .2221 | .2066 | .2207 | **.2359** |
| | F@5 | .2535 | .0688 | .2814 | .2606 | .2795 | **.2987** |
| | F@10 | .1879 | .0523 | .2131 | .1875 | .2028 | **.2237** |
| | nDCG@5 | .3449 | .0841 | .3741 | .3492 | .3851 | **.4022** |
| | nDCG@10 | .3712 | .0918 | .4070 | .3693 | .4095 | **.4343** |
| CiteULike | R@5 | .3665 | .0631 | .3933 | .3639 | .4114 | **.4325** |
| | P@5 | .1687 | .0323 | .1829 | .1698 | .1897 | **.2003** |
| | F@5 | .2310 | .0427 | .2497 | .2315 | .2597 | **.2738** |
| | F@10 | .1672 | .0294 | .1825 | .1560 | .1797 | **.1928** |
| | nDCG@5 | .3414 | .0600 | .3632 | .3457 | .4016 | **.4140** |
| | nDCG@10 | .3674 | .0631 | .3926 | .3596 | .4221 | **.4385** |
| TravelWell | R@5 | .2207 | .0714 | .2442 | .1740 | .2491 | **.2828** |
| | P@5 | .1000 | .0366 | .1333 | .0800 | .1300 | **.1400** |
| | F@5 | .1376 | .0484 | .1724 | .1096 | .1708 | **.1872** |
| | F@10 | .1125 | .0388 | .1356 | .0744 | .1287 | **.1426** |
| | nDCG@5 | .2110 | .0717 | .2253 | .1622 | .2525 | **.2615** |
| | nDCG@10 | .2411 | .0800 | .2686 | .1730 | .2783 | **.2900** |
| MACE | R@5 | .1306 | .0510 | .1463 | .1522 | .1775 | **.1901** |
| | P@5 | .0576 | .0173 | .0618 | .0631 | .0812 | **.0812** |
| | F@5 | .0799 | .0259 | .0869 | .0893 | .1114 | **.1138** |
| | F@10 | .0662 | .0170 | .0692 | .0615 | .0829 | **.0848** |
| | nDCG@5 | .1146 | .0463 | .1296 | .1502 | .1670 | **.1734** |
| | nDCG@10 | .1333 | .0483 | .1477 | .1568 | .1835 | **.1902** |

Contradicting [13], which suggested to use MOOCs datasets to evaluate TEL recommendations, our findings indicate that recommender performance results calculated on MOOCs are not representative for other, typically sparse, TEL environments. This is especially true for small-scale environments such as Apos-dle, where the evaluation positively shows that algorithms based on implicit usage data do not satisfy the use case. For Aposdle, which has only six users, none of the considered algorithms showed acceptable results. While approaches based on individual user data ($CB_T$, SUSTAIN) may work in similar settings, we suppose this is hindered by the unfortunate association of topics, which do not describe the content of a resource but rather the application type (e.g., tem-plate) and the poor allocation of topics to resources which is on average 1.16. We believe that learning environments that serve only a very small number of users, such as often the case in work place or formal learning settings, should draw on recommendation approaches that build upon a thorough description of learner and learning resources as incorporated in ontology-based recommender systems.

### 4.2 Tag Recommendations (*RQ2*)

The tag recommender evaluation was limited to the four datasets of our study that feature tags. Contrary to the results of the resource recommender study, we can observe a clear winner, which performs best on all datasets and metrics as depicted in Table 3. $\mathrm{BLL}_{AC} + \mathrm{MP}_R$ combines frequency and recency of a user's tagging history, which is enhanced by context information and therewith also recommends tags that are new to a user. Because runtime and complexity are considered very important factors in most TEL environments [8], we also emphasize the results of $\mathrm{MP}_{U,R}$ that outperforms the comparably cost-intensive $\mathrm{CF}_U$ in three of four settings, and hence forms a good alternative for runtime-sensitive settings. An extensive evaluation of runtime and memory for tag recommendation algorithms can be found in [18].

## 5 Conclusion

This paper presents a data-driven study that measures the performance of six known recommendation algorithms and variations thereof on altogether six TEL datasets from different application domains. Learning settings cover social bookmarking, open social learning, MOOCs and workplace learning. First, we investigate the suitability of three state-of-the-art recommendation algorithms (MP, CF, CB) and two approaches suggested for the educational context (UCbSim, SUSTAIN). The algorithms are implemented on implicit usage data. Our results show that satisfactory performance values can only be reached for KDD15, the MOOCs dataset. This suggests that standard resource recommendation algorithms, originating from the data-rich commercial domain are not well suited to the needs of sparse-data learning environments (*RQ1*). In a second study, we evaluate computationally inexpensive tag recommendation algorithms that may be applied to support learners' tagging behavior. To this end, we computed the performance of MP, CF and a cognitive-inspired algorithm, $\mathrm{BLL}_{AC}$, on four datasets. Results show that a hybrid recommendation approach combining $\mathrm{BLL}_{AC}$ and $\mathrm{MP}_R$ clearly outperforms the remaining methods (*RQ2*).

**Limitations and Future Work.** This evaluation only covers performance measurements of resource and tag recommendation algorithms. Other relevant indicators, as described in [13], such as user satisfaction, task support, learning performance and learning motivation are not addressed in this research. Also, we would like to mention the restriction of data-driven studies to items that are part of a user's history (i.e., if a user did not engage with a specific learning resource in the usage data, the evaluation considers this resource as wrongly recommended). However, this might not be the case. Thus, for future work, we plan to validate our results in an online recommender study. We believe that this would allow us to measure the real user acceptance of the recommendations.

**Acknowledgments.** We would like to gratefully acknowledge Katja Niemann who provided us with the MACE and TravelWell datasets, as well as the organizers of KDD Cup 2015 and XuetangX for making the KDD dataset available. This work is funded by the Know-Center, the EU-IP Learning Layers (Grant Agreement: 318209) and the EU-IP AFEL (Grant Agreement: 687916). The Know-Center is funded within the Austrian COMET Program under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria.

## References

1. Drachsler, H., Verbert, K., Santos, O.C., Manouselis, N.: Panorama of recommender systems to support learning. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 421–451. Springer, Heidelberg (2015)
2. Khribi, M.K., Jemni, M., Nasraoui, O.: Recommendation systems for personalized technology-enhanced learning. In: Kinshuk, Huang, R. (eds.) Ubiquitous Learning Environments and Technologies, pp. 159–180. Springer, Heidelberg (2015)
3. Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., Koper, R.: Recommender systems in technology enhanced learning. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 387–415. Springer, Heidelberg (2011)
4. Verbert, K., Manouselis, N., Drachsler, H., Duval, E.: Dataset-driven research to support learning and knowledge analytics. Educ. Technol. Soc. **15**(3), 133–148 (2012)
5. Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., Duval, E.: Dataset-driven research for improving recommender systems for learning. In: Proceedings of LAK 2011, pp. 44–53. ACM (2011)
6. Fazeli, S., Loni, B., Drachsler, H., Sloep, P.: Which recommender system can best fit social learning platforms? In: Rensing, C., de Freitas, S., Ley, T., Muñoz-Merino, P.J. (eds.) EC-TEL 2014. LNCS, vol. 8719, pp. 84–97. Springer, Heidelberg (2014)
7. Niemann, K., Wolpers, M.: Usage context-boosted filtering for recommender systems in TEL. In: Hernández-Leo, D., Ley, T., Klamma, R., Harrer, A. (eds.) EC-TEL 2013. LNCS, vol. 8095, pp. 246–259. Springer, Heidelberg (2013)
8. Manouselis, N., Vuorikari, R., Van Assche, F.: Collaborative recommendation of e-learning resources: an experimental investigation. J. Comput. Assist. Learn. **26**(4), 227–242 (2010)
9. Bateman, S., Brooks, C., Mccalla, G., Brusilovsky, P.: Applying collaborative tagging to e-learning. In: Proceedings WWW 2007 (2007)
10. Kuhn, A., McNally, B., Schmoll, S., Cahill, C., Lo, W.-T., Quintana, C., Delen, I.: How students find, evaluate and utilize peer-collected annotated multimedia data in science inquiry with Zydeco. In: Proceedings of SIGCHI 2012, pp. 3061–3070. ACM (2012)
11. Klašnja-Milićević, A., Ivanović, M., Nanopoulos, A.: Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions. Artif. Intell. Rev. **44**(4), 571–604 (2015)

12. Manouselis, N., Drachsler, H., Verbert, K., Duval, E.: Recommender Systems for Learning. Springer, New York (2012)
13. Erdt, M., Fernandez, A., Rensing, C.: Evaluating recommender systems for technology enhanced learning: a quantitative survey. IEEE Trans. Learn. Technol. **8**(4), 326–344 (2015)
14. Lohmann, S., Thalmann, S., Harrer, A., Maier, R.: Learner-generated annotation of learning resources-lessons from experiments on tagging. J. Univ. Comput. Sci. **304**, 312 (2007)
15. Diaz-Aviles, E., Fisichella, M., Kawase, R., Nejdl, W., Stewart, A.: Unsupervised auto-tagging for learning object enrichment. In: Kloos, C.D., Gillet, D., Crespo García, R.M., Wild, F., Wolpers, M. (eds.) EC-TEL 2011. LNCS, vol. 6964, pp. 83–96. Springer, Heidelberg (2011)
16. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
17. Niemann, K.: Automatic tagging of learning objects based on their usage in web portals. In: Conole, G., Klobučar, T., Rensing, C., Konert, J., Lavoué, E. (eds.) Design for Teaching and Learning in a Networked World, vol. 9307, pp. 240–253. Springer, Heidelberg (2015)
18. Kowald, D., Lex, E.: Evaluating tag recommender algorithms in real-world folksonomies: a comparative study. In: Proceedings of RecSys 2015, pp. 265–268. ACM (2015)
19. Seitlinger, P., Kowald, D., Kopeinik, S., Hasani-Mavriqi, I., Ley, T., Lex, E.: Attention please! a hybrid resource recommender mimicking attention-interpretation dynamics. In: Proceedings of International World Wide Web Conferences Steering Committee, WWW 2015, pp. 339–345 (2015)
20. Trattner, C., Kowald, D., Seitlinger, P., Kopeinik, S., Ley, T.: Modeling activation processes in human memory to predict the use of tags in social bookmarking systems. J. Web Sci. **2**(1), 1–16 (2016)
21. Kowald, D., Lacic, E., Trattner, C.: Tagrec: towards a standardized tagrecommender benchmarking framework. In: Proceedings of HT 2014. ACM, New York (2014)
22. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in Folksonomies. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) PKDD 2007. LNCS (LNAI), vol. 4702, pp. 506–514. Springer, Heidelberg (2007)
23. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007)
24. Marinho, L.B., Schmidt-Thieme, L.: Collaborative tag recommendations. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (eds.) Data Analysis, Machine Learning and Applications, pp. 533–540. Springer, Heidelberg (2008)
25. Gemmell, J., Schimoler, T., Ramezani, M., Christiansen, L., Mobasher, B.: Improving folkrank with item-based collaborative filtering. In: Recommender Systems & the Social Web (2009)
26. Basilico, J., Hofmann, T.: Unifying collaborative and content-based filtering. In: Proceedings of ICML 2004, p. 9. ACM (2004)
27. Friedrich, M., Niemann, K., Scheffel, M., Schmitz, H.-C., Wolpers, M.: Object recommendation based on usage context. Educ. Technol. Soc. **10**(3), 106–121 (2007)
28. Niemann, K., Wolpers, M.: Creating usage context-based object similarities to boost recommender systems in technology enhanced learning. IEEE Trans. Learn. Technol. **8**(3), 274–285 (2015)

29. Kowald, D., Kopeinik, S., Seitlinger, P., Ley, T., Albert, D., Trattner, C.: Refining frequency-based tag reuse predictions by means of time and semantic context. In: Atzmueller, M., Chin, A., Scholz, C., Trattner, C. (eds.) MUSE/MSM 2013, LNAI 8940. LNCS, vol. 8940, pp. 55–74. Springer, Heidelberg (2015)
30. Anderson, J.R., Schooler, L.J.: Reflections of the environment in memory. Psychol. Sci. **2**(6), 396–408 (1991)
31. Love, B.C., Medin, D.L., Gureckis, T.M.: Sustain: a network model of category learning. Psychol. Rev. **111**(2), 309 (2004)
32. Benchmark folksonomy data from bibsonomy, Knowledge and Data Engineering Group. University of Kassel, 2013/2015. http://www.kde.cs.uni-kassel.de/bibsonomy/dumps
33. Stefaner, M., Dalla Vecchia, E., Condotta, M., Wolpers, M., Specht, M., Apelt, S., Duval, E.: MACE – enriching architectural learning objects for experience multiplication. In: Duval, E., Klamma, R., Wolpers, M. (eds.) EC-TEL 2007. LNCS, vol. 4753, pp. 322–336. Springer, Heidelberg (2007)
34. Vuorikari, R., Massart, D.: Datatel challenge: European schoolnet's travel well dataset. In: Proceedings of RecSysTEL 2010 (2010)
35. Beham, G., Stern, H., Lindstaedt, S.: Aposdle-ds a dataset from the Aposdle work integrated learning system. In: Proceedings of RecSysTEL 2010 (2010)
36. Marinho, L.B., Hotho, A., Jäschke, R., Nanopoulos, A., Rendle, S., Schmidt-Thieme, L., Stumme, G., Symeonidis, P.: Recommender Systems for Social Tagging Systems. Springer, New York (2012)
37. Sakai, T.: On the reliability of information retrieval metrics based on graded relevance. Inf. Process. Manage. **43**(2), 531–548 (2007)

# Discouraging Gaming the System Through Interventions of an Animated Pedagogical Agent

Thiago Marquez Nunes[1], Ig Ibert Bittencourt[2], Seiji Isotani[3], and Patricia A. Jaques[1(✉)]

[1] PIPCA - Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, Brazil
nunes.thiago@live.com, pjaques@unisinos.br

[2] IC - Universidade Federal de Alagoas (UFAL), Maceió, Brazil
ig.ibert@ic.ufal.br

[3] ICMC - Universidade de São Paulo (USP), São Carlos, Brazil
sisotani@icmc.usp.br

**Abstract.** Intelligent Tutoring Systems (ITSs) have been largely used in school settings and considered effective learning tools. However, students' performance might be impaired by their undesired behaviors. An example of these behaviors is "gaming the system", which happens when the student tries to mislead the system in order to advance faster in the tasks. Previous works have tried to treat this behavior by blocking student's actions; however, this restrictive approach has proved to be ineffective. We propose the use of animated pedagogical agents in ITSs as a non-restrictive approach to gaming the system. We believe that an animated pedagogical agent can discourage this behavior by taking two actions: (1) showing it is aware of when students are gaming, and (2) educating students about the negative impact of this behavior on their learning. We implemented this approach in a step-based algebra ITS, and a classroom experiment was conducted with 37 students who used the system for 50 min on average to solve linear equations. Although, due to the design restrictions of the experiment, we could not statistically demonstrate that the presence of the agent decreased the gaming of the system, descriptive statistics of the tutor log data show evidence of a possible positive effect of an aware animated pedagogical agent on students' behavior, indicating that this approach deserves further investigation.

**Keywords:** Animated pedagogical agents · Gaming the system · Intelligent tutoring systems

## 1 Introduction

Intelligent Tutoring Systems (ITSs) have been largely used in school settings, showing effectiveness as learning tools [2,9,11,12]. These systems use artificial intelligence techniques to create effective tutors that can provide individualized assistance for students, allowing them to learn at their own pace. ITSs have sought to achieve the same learning levels as one-to-one teaching, helping to increase students' performance by at least one letter grade [19,22,23,25].


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 139–151, 2016.
DOI: 10.1007/978-3-319-45153-4_11

Although research has shown ITSs are powerful learning tools, students' outcomes in ITSs might be impaired by occasional undesired behaviors, such as "off-task behavior", "gaming the system" (GTS) and others. A particular case is GTS, in which the student tries to mislead the system in order to advance faster in his or her tasks. The GTS behavior is undesired because it impairs student performance [7,10,24]. Studies have shown that students who game the system learn a third less than students who do not [5].

This behavior generally happens in two different situations: (i) when the student takes advantage of the hints provided by the ITS to obtain the correct answer without reading and trying to understand the exercise, and (ii) when the student tries all possible answers without making actual cognitive efforts to achieve the correct solution. In the first case, the student understands that he or she will receive the correct answer after a certain number of hint requests. In the second case, the student makes arbitrary attempts at the available answers until he or she accidentally finds the correct one. This second way of gaming is more often found in multiple-choice questions systems [7].

Previous works have tried to treat the gaming behavior by blocking the student's actions. This can be achieved by, for instance, disabling the "hint" button [1,6], and it has been shown to be ineffective. In fact, in these cases the student tends to get frustrated or angry and looks for other ways of gaming the system. Baker et al. [6] used animated pedagogical agents to tackle this issue.

Animated pedagogical agents (APAs) are software agents represented by an animated character that generally have a teaching role in a learning system. They interact with the user through gestures and facial expressions. In Baker's study, the APA is a dog named Scooter that becomes angry when it realizes that the student is gaming. One important remark here is that Scooter does not explain why it gets angry; it only shows facial expressions of anger. The student may not make the connection between Scooter's angry mood with his or her gaming behavior. Besides, Scooter does not try to educate students about the negative effects of gaming. Possibly because of that, the authors could not demonstrate that the agent discouraged gaming behavior.

Several studies have shown the benefits of using APAs in learning systems [2,3,13,15,17,18,20]. An animated agent can influence the learning experience and improve students' performance. They enhance the communication channel between tutors and students in addition to increasing the tutor's motivational ability and the student's empathy with the tutor. The presence of this kind of agent improves the student's motivation because it simulates the presence of a real teacher, which causes the student-agent interaction to flow more naturally and anthropomorphically. Furthermore, other studies (e.g., [4]) have suggested that non-restrictive approaches constitute an interesting hypothesis to be studied when it comes to GTS.

Similarly to what was proposed by [6], we use APAs as a non-restrictive approach to GTS in ITSs. However, unlike Scooter, our agent shows students that it is aware of their gaming behavior and tries to educate them about the negative effects of this behavior. We believe that an APA can discourage GTS if it shows that it is aware of students' actions, including when they are gaming. The APA can also educate students on the negative impact of this behavior on their learning.

To verify the impact of an "aware" APA on students' behavior, we integrated an APA into a step-based algebra ITS called PAT2Math [14]. This agent provides feedback related to the problem faced by the student and makes interventions when it realizes that the student is gaming. Both feedback and GTS interventions are delivered by the animated agent in the form of speech balloons. Besides, in the interventions, the agent displays a nonverbal behavior representing its emotion about the student's behavior (for instance, it is frustrated or angry).

## 2 Animated Pedagogical Agents as a Non-restrictive Approach to Gaming the System

This section presents our non-restrictive approach to deal with GTS in ITSs. The main goal is to discourage GTS through interventions of an APA. The agent monitors students' actions and, through its verbal and nonverbal behavior, shows it is aware that the student is trying to game the system. We believe that one of the reasons why students game the system is their impression that the tutor does not "understand" their actions ("It is only a machine!") and hence cannot act accordingly. Demonstrating awareness of students' behavior makes agents more believable, which in turn makes students more careful about their actions, similar to the behavior students present when interacting with a human teacher. Furthermore, because APAs improve the communication channel, motivational capability, and empathy between the tutor and the student [2,16,18], they are considered a more powerful approach to educating students on the negative effects of gaming the tutor than the display of a warning window on the screen. In fact, previous research [4] claims that this kind of proactive and non-restrictive approach using APAs is promising for dealing with GTS.

### 2.1 The Algebra Tutor

The APA has been integrated into a web-based algebra ITS that provides personalized assistance for students in the solving of step-by-step equations. Figure 1 shows the equation solving interface of the tutor. Item (a) shows the equation the student has to solve, and the equation lines below it are steps entered by the student. Item (b) corresponds to a symbol that represents the operation the student chose to solve the current step. The student chooses operations from panels (e) (for both first and second degree equations) or (f) (only for second degree equations). The panels in item (j) contain (in this order) arithmetic operations, undo and redo operations, score and performance in the ITS. Panel (g) contains buttons that allow the student to ask for a new equation, clear the last entered step, and edit the last step. The buttons on panel (h) allow students to ask for help. The student can request a hint (first button), ask the system to solve a step (second button), or even ask the system to solve the entire equation

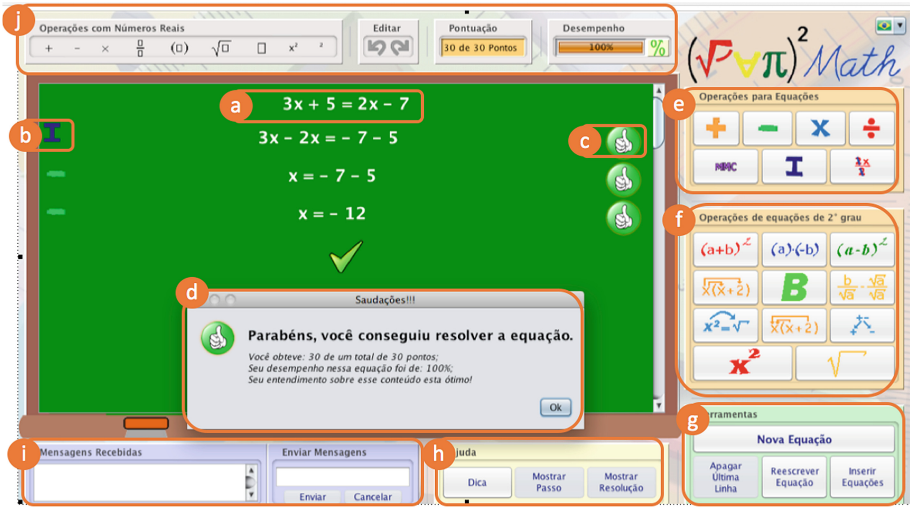



**Fig. 1.** Original version of PAT2Math with no APA

(third button). Finally, Panel (i) allows the student to exchange messages with a teacher, and panel (d) shows the pop-up window that is displayed when students succeed in solving an equation.

In each step, the student receives assistance from the tutor in one of the following ways: real-time minimum feedback (yes/no), on-demand hints, or immediate feedback on errors. The minimum feedback tells the student whether the step is correct or not, and it is displayed next to the current equation, as depicted in Fig. 1. Hints are given on-demand, whenever students click on the corresponding button, because, for instance, they do not know how to proceed to solve a given step. Immediate feedback is given when students enter an incorrect step. Similarly to the minimum feedback behavior shown in Fig. 1, the hint messages and immediate feedback are also given by the APA in a speech balloon.

The minimum feedback is provided by the tutor's rule-based expert system [14], which is able to solve linear and quadratic equations. Given an equation, the expert system can determine the possible next steps. PAT2Math's student model compares the step provided by the student with the possible solutions generated by the expert system to verify if the student's step is correct, similarly to what is done in model-tracing tutors [11]. The tutor also has a bug library, which allows the system to identify students' misconceptions.

To show the hints, the hint component calls the expert system passing the student's step as a function parameter. The expert system verifies if the step is correct, detects possible misconceptions, and returns a possible correct solution to the hint component with some additional information (e.g., which algebraic operation could be used to enter a next correct step in the equation, the student's misconception, etc.). If there is more than one possible solution, the expert system chooses the most correct one, i.e., the one that would lead to the solution of the equation in fewest steps. Then, the hint component interprets the solution, chooses a hint, and sends it to the interface component to display it.

The hint component stores the hints textually, and classifies them by level of detail. A given algebraic operation usually has five levels of detail that range from more generic hints (e.g., draws the student's attention to a term of the equation) to more specific hints (e.g., calls the expert system to provide the problem's solution). To customize the hints to the current equation, the help texts are templates that allow replacing certain marks for specific terms of the current step.

The APA was implemented with the framework Divalite [21]. This framework was chosen because it offers an easy, fast, and robust integration with web applications. The tutor was developed in HTML5 (client side) and Java (server side).

### 2.2 Integration Between the APA and PAT2Math

To implement the gaming-the-system detector inside the tutor, we needed to introduce restrictions to which GTS behavior causes the tutor to act. We chose to restrict the tutor's actions to when the student games the system by asking repeatedly for hints on the same step. This choice was based on the tutor's features; because it does not have multiple-choice problems, it is more difficult for the student to game the system by trying different answers.

After defining the scope of the research, the second step was to define which situations are considered GTS in the context of our algebra tutor. In this step, we were assisted by a math teacher who had already used our tutor in her classroom. After analyzing the log files of previous experiments with the tutor, we empirically determined that the gaming behavior generally happens when the student asks for two or more hints in the same step without interacting with the tutor (a possible form of interaction could be trying to solve the step). Although this inference mechanism might be considered simple, it is very similar to the one proposed by the GIFT framework [8], a well-known framework to detect when students are gaming. However, GIFT incorporated a time threshold between two hint requests to verify if students actually tried to read the hints.

The tutor monitors the student's actions and evaluates whether the system is being gamed or not based on the aforementioned mentioned restrictions. If the student asks for two successive hints without any action towards solving the equation (e.g., entering a step), the agent considers that the student has gamed the system and starts to act accordingly. However, it is the agent who chooses when to intervene about GTS. A probability factor determines how likely the student is expected to receive an intervention regarding the gaming behavior. By controlling the tutor's actions with this factor, our goal was to avoid that the agent shows a robotic behavior, which could disengage students.

The initial probability factor that determines the agent's intervention frequency is zero. Each time the tutor identifies the gaming behavior, this factor is increased by 15 %, up to a maximum value of 90 %. Hence, the system reaches the maximum probability of interventions after six attempts from the student at gaming the system. Each time the system detects that the student has gamed, it generates a real number between zero and one [0; 1]. If the generated number has a value lower than the frequency factor, a gaming intervention is shown by the agent; otherwise, the agent shows a hint. Both gaming interventions and hints are showed as messages of the agent in speech balloons. Due to the short duration of the experiment, we opted for not resetting the frequency factor during the session.

When the maximum factor is achieved, the tutor might show only messages regarding GTS, which would give the tutor a repetitive and unnatural behavior. To avoid this, an additional mechanism was implemented for when the maximum frequency is reached. For each message regarding the gaming behavior, the tutor provides a hint that is relevant to the student's task. Although we see this as a limitation in our experiment, we consider it necessary to keep the tutor's believability.

Moreover, to increase the agent's believability and to make the tutor's interventions more dynamic, the size of messages regarding GTS alternate between short and long. We created a total of 30 sentences (20 long sentences and 10 short sentences), which can be combined with seven different body behaviors of the agent. The messages of a given type are randomly chosen by the agent. Figures 2 and 3 illustrate some examples of these messages.

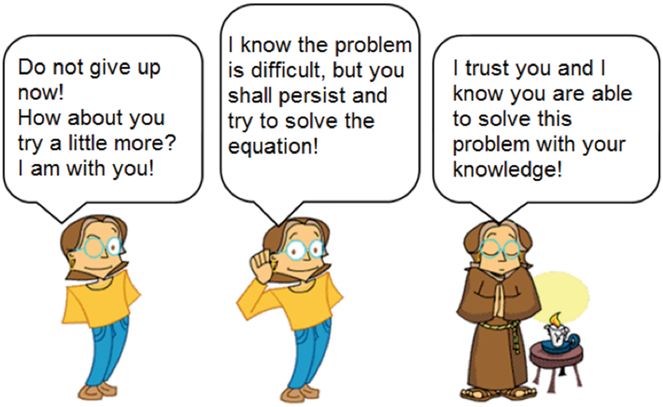



**Fig. 2.** Interventions of the agent showing that it is worried about the student's learning (translated from original messages in Portuguese)

From a technical point of view, the code modifications regarding GTS treatment in Pat2Math were performed directly in the tutor's interface model. That was done because this model is independent from the other system components. This allowed us to be more efficient in the integration process and to better validate our model before performing deeper changes in the tutor.


**Fig. 3.** The agent showing that it is aware the student is gaming (translated from original messages in Portuguese)

### 2.3 Agent's Interventions When it Detects Gaming the System

Once it is detected that the student is gaming, the agent starts to show the messages related to the student's conduct. We defined two types of strategies. The first strategy is educative and involves raising the student's awareness about the negative effects of gaming on learning. In this first type of message, the agent also shows that it is worried about the student's learning, as shown in Fig. 2. For the second strategy, the agent reveals that it suspects the student might be taking advantage by asking for help. In this second case, the goal is to discourage the student's gaming by demonstrating that it is "aware" of this type of behavior. For each student, the agent chooses one of these strategies and always uses it, assuming it as a personality trait (kind and attentive; or suspicious and cautious) to convey believability. Some examples of messages from the second strategy are shown in Fig. 3. Although the messages were originally shown in Portuguese, they were translated into English in this figure.

The strategies are composed of the agent's body behavior (Divalite animations) and the speech bubble messages, because Divalite's agents are not endowed with voice synthesis. These messages were created with the help of the math teacher and contain sentences similar to the ones that teachers use in the classroom.

## 3 Evaluation and Discussion

We conducted a classroom study to try and find indications that the agent's interventions were working (i.e., discouraging GTS). Thirty-seven Brazilian students from two $8^{\text{th}}$ year classes participated in the study.

Initially, the participants were informed about generic goals of the intervention. Basically, all they were told was that they would use an intelligent learning environment that teaches algebra and that we wanted to verify if the system works well. Students were also notified that they were free to abandon the experiment before or during its course. Moreover, because the average age of the students was 14 years, their parents were asked to sign an informed consent form (ICF) one week before the session.

Two versions of PAT2Math were implemented and used in this study. The first version was the original PAT2Math, with no agent. The second version was an extension of the original PAT2Math with the GTS 'aware' APA integrated into it. The agent had one of two possible personalities, it could be either suspicious or kind.

Unfortunately, random assignment was not appropriate for this intervention because students would understand they were using different versions of the system due to the presence/absence of the agent. Thus, this evaluation study followed a 'Non-equivalent control groups post-test only'. Thirteen students from the afternoon class were assigned to the control group (the tutor version with no agent), and 24 students of the morning class were assigned to the experimental group (the tutor version with the "aware" agent). From a total of 24 students who were assigned to the experimental group, 13 students used the 'kind' version of the agent and 11 students had access to the suspicious agent. The interfaces of the control and experimental groups are shown in Figs. 1 and 4, respectively.

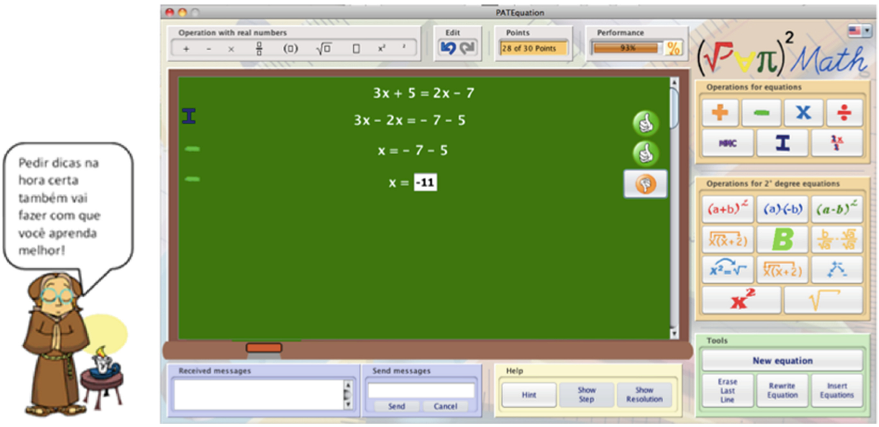



**Fig. 4.** The experimental group used the tutor's version with the gaming-the-system "aware" agent

The experiment was conducted at the school's computer lab. In the beginning of the session, the students had a brief explanation of how the tutor works. Afterwards, they were registered and started to interact with the system. The students did not receive any help from the experimenter regarding the problem-solving process, just about how to use the tutor.

The students used the tutor for 50 min on average. During this time period, they had to solve, step by step, a list of equations provided by the tutor. The equations were presented in a random order; this way, students were solving different equations at the same time. The tutor logged each student's actions.

The log files were analyzed to verify the occurrence of GTS during the interactions with the tutor. We identified how many times the students asked for help and how many times the APA intervened showing messages regarding GTS. The groups were evaluated according to five different parameters: the occurrence of GTS, the number of hints asked, the number of entirely solved equations, and the number of hints and messages regarding the GTS shown by the agent. Means (M), standard deviations (SD), and the total number of occurrences (#) for each of these parameters are listed in Table 1. For the evaluation, we considered any number of consecutive requests of help in the same step as only one occurrence of the gaming behavior. In this way, a single step cannot have more than one occurrence of gaming.

As the means in Table 1 indicate, students in the treatment group (considering data from students who interacted both with the suspicious and the kind agents) gamed the system as much as students in the control group, although experimental-group students asked for help more times. The number of equations solved was also slightly higher in the experimental group.

In the experimental group, comparisons between students who interacted with the suspicious-personality agent and students who interacted with the kind-personality agent show that the formers asked for hints fewer times. This impacted the GTS behavior of the students who used the kind agent, who had proportionally gamed more. It seems that the suspicious personality of the agent prevented students from asking for hints instead of only preventing them from gaming. Besides, the number of GTS messages shown by both versions of the agent are almost the same. This can be an indicator that the content of the message was the factor that impacted on student's behavior, and not the number of messages shown by the agent. We believe that the impact of the personality of the agent in students' behavior deserves further investigation.

Unfortunately, we could not apply a statistical test to verify if there is a statistically significant difference between the groups. As the students were not randomly assigned to control and experimental groups, many factors could jeopardize the internal validity of our quasi-experiment, which makes this type of experiment not recommended for inferential statistics.

However, considering the observations made during the experimental phase, we believe that the agent could have influenced students' behavior. The students who interacted with the suspicious agent asked for hints fewer times, and consequently gamed less. Besides, even though the students in the experimental group asked for hints more times, they gamed as much as students in the control group. This probably means they were trying to use the hints to learn how to solve the equations. Although students in the experimental group solved more equations, log analyses show few cases in which students solved the equation because the hint provided the final answer. Hence, there is no evidence that

**Table 1.** Descriptive statistics summarizing the number of GTS-related actions of the students

| Groups | # of times gaming the system | | |
|---|---|---|---|
| | M | SD | # |
| Experimental (Suspicious ag.) | 0.55 | 0.52 | 6 |
| Experimental (Kind ag.) | 1.85 | 1.86 | 24 |
| Experimental (Total) | 1.25 | 1.54 | 30 |
| Control | 1.27 | 1.62 | 19 |
| Groups | # of hints asked | | |
| | M | SD | # |
| Experimental (Suspicious ag.) | 4.18 | 3.19 | 46 |
| Experimental (Kind ag.) | 7.77 | 6.07 | 101 |
| Experimental (Total) | 6.13 | 5.19 | 147 |
| Control | 5.13 | 5.66 | 77 |
| Groups | # of agent's GTS messages shown | | |
| | M | SD | # |
| Experimental (Suspicious ag.) | 0.73 | 0.65 | 8 |
| Experimental (Kind ag.) | 0.77 | 1.24 | 10 |
| Experimental (Total) | 0.75 | 0.69 | 18 |
| Control | 0 | 0 | 0 |
| Groups | # of hints shown | | |
| | M | SD | # |
| Experimental (Suspicious ag.) | 3.45 | 2.91 | 38 |
| Experimental (Kind ag.) | 7.00 | 6.35 | 91 |
| Experimental (Total) | 5.38 | 5.29 | 129 |
| Control | 5.13 | 5.66 | 77 |
| Groups | # of solved equations | | |
| | M | SD | # |
| Experimental (Suspicious ag.) | 4.55 | 3.59 | 50 |
| Experimental (Kind ag.) | 7.46 | 4.03 | 97 |
| Experimental (Total) | 6.13 | 4.04 | 147 |
| Control | 5.20 | 4.35 | 78 |

the larger number of solved equations in the experimental group was due to the system's hints.

## 4 Conclusion

This paper has presented an original and non-restrictive approach to prevent students from gaming the system. Our approach consists of an animated agent that shows the students that it is uncomfortable about their gaming behavior and is concerned about the consequences of this behavior on their learning. Our main hypothesis is that students game the tutor because they consider it “only a machine” that is not aware of their actions and is not able to punish them for inadequate behavior. Students have interacted with two personalities of the animated agent: kind and attentive; or suspicious and cautious. The present paper also reinforces the use of non-restrictive approaches in dealing with undesired behaviors, turning the experience with the tutor into one that is more friendly and anthropomorphic.

Although we were unable to statistically prove a causal relation between the presence of the APA and a decrease in the gaming behavior, the descriptive statistics show a possible positive effect of an aware APA on students' behavior. Besides, it seems that the suspicious-personality agent inhibits students from asking help instead of making them game less. We believe that these results deserve further investigation and we are already planning a true experiment with a larger group of students randomly assigned to control and treatment groups.

In future studies, we plan to develop a more complex and consistent personality model for the agent. Besides establishing a better interaction between the tutor and the student, an APA endowed with a rich personality model will allow us to identify if there are personality traits that are more appropriate to deal with GTS. Also, we see the use of posture for the animated agent as a valid topic of research to improve this student-tutor relation.

Another possible work is the integration of the presented model with the GIFT framework, which is a well-known and established framework to detect disengaged behaviors [8]. An already tested model for detecting GTS and other off-task behaviors could increase the effectiveness of our approach.

**Acknowledgement.** This work is supported by the following research funding agencies in Brazil: CAPES, CNPq and FAPERGS.

## References

1. Aleven, V.: Helping students to become better help seekers: towards supporting metacognition in a cognitive tutor. In: Proceedings of German-USA Early Career Research Exchange Program (2001)
2. Arroyo, I., Woolf, B.P., Cooper, D., Burleson, W., Muldner, K.: The impact of animated pedagogical agents on girls' and boys' emotions, attitudes, behaviors and learning. In: 11th IEEE International Conference on Advanced Learning Technologies (ICALT), pp. 506–510 (2011)
3. Atkinson, R.K.: Optimizing learning from examples using animated pedagogical agents. J. Educ. Psychol. **94**, 416–427 (2002)
4. Baker, R.S.J.D.: Gaming the system: a retrospective look. Philippine Comput. J. **6**(2), 9–13 (2011)
5. Baker, R.S., Corbett, A.T., Koedinger, K.R.: Detecting student misuse of intelligent tutoring systems. In: Lester, J.C., Vicari, R.M., Paraguaçu, F. (eds.) ITS 2004. LNCS, vol. 3220, pp. 531–540. Springer, Heidelberg (2004)

6. Baker, R.S.J., Corbett, A.T., Koedinger, K.R., Evenson, S., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., Beck, J.E.: Adapting to when students game an intelligent tutoring system. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 392–401. Springer, Heidelberg (2006)
7. Baker, R., Corbett, A.T., Koedinger, K.R., Wagner, A.: Off-task behavior in the cognitive tutor classroom: when students game the system. In: ACM Conference on Computer-Human Interaction (CHI), pp. 383–390 (2004)
8. Baker, R.S.J.D., Rossi, L.M.: Assessing the disengaged behaviors of learners. In: Sottilare, R., Graesser, A., Hu, X., Holden, H. (eds.) Design Recommendations for Intelligent Tutoring Systems, pp. 155–166. U.S. Army Research Lab, Orlando (2013)
9. Beal, C.R., Walles, R., Arroyo, I., Woolf, B.P.: On-line tutoring for math achievement testing: a controlled evaluation. J. Interact. Online Learn. **6**(1), 43–55 (2007)
10. Cocea, M., Hershkovitz, A., Baker, R.S.J.D.: The impact of off-task and gaming behaviors on learning: Immediate or aggregate. Front. Artif. Intell. Appl. **200**, 507–514 (2009)
11. Corbett, A.T., Koedinger, K.R., Hadley, W.H.: Cognitive tutors: from the research classroom to all classrooms. In: Goodman, P. (ed.) Technology Enhanced Learning: Opportunities for Change, pp. 235–263. Lawrence Erlbaum Associates, Mahway (2001)
12. Goguadze, G., Melis, E.: Feedback in ActiveMath exercises. In: Proceedings of International Conference on Mathematics Education, pp. 1–7 (2008)
13. Hasegawa, D., Shirakawa, S., Shioiri, N., Hanawa, T., Sakuta, H., Ohara, K.: The effect of metaphoric gestures on schematic understanding of instruction performed by a pedagogical conversational agent. In: Zaphiris, P., Ioannou, A. (eds.) LCT 2015. LNCS, vol. 9192, pp. 361–371. Springer, Heidelberg (2015)
14. Jaques, P.A., Seffrin, H., Rubi, G., de Morais, F., Ghilardi, C., Bittencourt, I., Isotani, S.: Rule-based expert systems to support step-by-step guidance in algebraic problem solving: the case of the tutor PAT2Math. Expert Syst. Appl. **40**(14), 5456–5465 (2013)
15. Johnson, A., Ozogul, G., Reisslein, M.: Supporting multimedia learning with visual signalling and animated pedagogical agent: moderating effects of prior knowledge. J. Comput. Assist. Learn. **31**(2), 97–115 (2015)
16. Johnson, W.L., Rickel, J., Lester, J.C.: Animated pedagogical agents: face-to-face interaction in interactive learning environments. Int. J. Artif. Intell. Educ. **11**, 47–78 (2000)
17. Kim, Y., Baylor, A.L.: Research-based design of pedagogical agent roles: a review, progress, and recommendations. Int. J. Artif. Intell. Educ. **26**(1), 160–169 (2016)
18. Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A., Bhogal, R.S.: The persona effect: affective impact of animated pedagogical agents. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 359–366. Atlanta (1997)
19. Ma, W., Adesope, O.O., Nesbit, J.C., Liu, Q.: Intelligent tutoring systems and learning outcomes: a meta-analysis. J. Educ. Psychol. **106**(4), 1–18 (2014)
20. van der Meij, H., van der Meij, J., Harmsen, R.: Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry learning environment. Educ. Technol. Res. Dev. **63**(3), 381–403 (2015)
21. Sansonnet, J., Jaques, P.A., Correa, D., Braffort, A., Verrecchia, C.: Developing web fully-integrated conversational assistant agents. In: Cho, Y., Tarokh, V. (eds.) Proceedings of the 2012 Research in Applied Computation Symposium (RACS 2012), pp. 14–19. ACM, San Antonio, Texas (2012)

22. Steenbergen-Hu, S., Cooper, H.: A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. J. Educ. Psychol. **105**(4), 970–987 (2013)
23. Steenbergen-Hu, S., Cooper, H.: A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. J. Educ. Psychol. **106**(2), 331–347 (2014)
24. Tait, K., Hartley, J., Anderson, R.: Feedback procedures in computer-assisted arithmetic instruction. Br. J. Educ. Psychol. **46**(2), 161–171 (1973)
25. Vanlehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educ. Psychol. **46**(4), 197–221 (2011)

# Multi-device Territoriality to Support Collaborative Activities

## Implementation and Findings from the E-Learning Domain

Jean-Charles Marty[1(✉)], Audrey Serna[2], Thibault Carron[3], Philippe Pernelle[4], and David Wayntal[5]

[1] Université de Savoie Mont-Blanc, LIRIS, UMR5205, 69622 Villeurbanne, France
jean-charles.marty@liris.cnrs.fr
[2] INSA Lyon, LIRIS, UMR5205, 69622 Villeurbanne, France
audrey.serna@liris.cnrs.fr
[3] Université de Savoie Mont-Blanc, LIP6, UMR7606, 75252 Paris, France
thibault.carron@lip6.fr
[4] Université de Lyon 1, 69622 Villeurbanne, France
philippe.pernelle@univ-lyon1.fr
[5] HUCO, Campus Scientifique Technolac, 73370 Le Bourget-du-Lac, France
david.wayntal@huco.fr

**Abstract.** In our research, we consider complex Game Based Learning (GBL) scenarios where both individual and collaborative learning are addressed. In order to support these scenarios, personal devices (tablets, mobile phones) co-exist with shared devices (collaborative tabletops). New learning usages emerge in these multi-device environments where learners can swap from individual to collaborative tasks. In this context, new problems appear when one wants to design new GBL activities. One major issue refers to the combination of personal and collective workspaces. This notion also known as "territoriality" has been addressed in the literature, particularly for Collaborative Tabletop Workspaces. However, we need to extend and reconsider this notion when designing multi-device activities. For instance, providing users with both private and shared devices raises information visualisation issues. In this work, we present several aspects to consider for the design of GBL activities in this context: territory arrangement in multi device environments; inter-territory actions to manage information; and contextual information visibility for objects involved in the learning tasks. We then detail a case study used to apply our proposal. We have designed and enacted a scenario of a collaborative game to learn French grammar. Individual and collaborative tasks co-exist and are supported with a multi-device environment. We describe how the experiment was carried out and the main results deduced from students' answers to questionnaires.

**Keywords:** Collaborative learning · Personal workspace · Collaborative workspace · Tablet · Tabletop · Multi-device environment


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 152–164, 2016.
DOI: 10.1007/978-3-319-45153-4_12

# 1 Introduction

It is now commonly stated that Game Based Learning (GBL) has positive effects on learners' engagement (Hildmann et al. 2009 Kelle et al. 2011). Students feel more concerned and invested when the learning scenario is motivating, and this is particularly true with GBL scenarios (Pernin et al. 2014). With the evolution of technology and the emergence of new devices, the complexity of learning scenarios mixing personal and collaborative activities is considerably increasing (Dillenbourg et al. 2007). Our research interest focuses on the improvement of writing collaborative activities within GBL environments. According to the results described in Dillenbourg (1999), collaborative activities are enhanced when learners have both individual and collective workspaces. This clear separation of workspaces for collaborative activities taking place on tabletops has already been addressed through the "territoriality" notion (Scott 2004).

Based on previous studies, we propose to extend the concepts of territoriality to multi-device environments. We will address three research questions that are derived from this general objective. Firstly, we need to better identify which devices are appropriate to support individual and collective activities for learners. Then, we look at specifying the interaction part, detailing how to work with personal or collective data, how to transfer them from personal to collaborative workspaces, or conversely, from collaborative to personal workspaces, defining gestures and relevant interactions for performing the basic identified actions. Finally, we consider aspects related to information visualisation. As several people are involved in these collaborative learning activities, we may need to define views on several pieces of information where the information may be partly hidden to others. We therefore need to study what are the relevant information representations according to the user, his/her device and to the level of privacy of the data.

Section 2 presents the related work on which we ground our work. Previous research in the different mentioned domains is briefly presented. We explain how personal and collective workspaces are used in a GBL collaborative activity; then we describe the territoriality aspects. In Sect. 3, we present our contribution. We describe our proposal of territory arrangement in multi-device environments. This proposal was applied to the design of a particular learning situation evaluated in ecological situation with 54 students. In Sect. 4, we describe this case study and present participants' feeling about territoriality in a multi-device environment and interesting findings on design impacts of collective behaviours. Finally, we suggest new ideas to reinforce this research work.

# 2 Related Work

## 2.1 Collaborative Game-Based Learning

Collaborative learning is an activity where two people or more (a pair, a group, a class, etc.) learn something (taking a course, solving a problem, etc.) together (face to face, through computers, etc.) according to a definition given by Dillenbourg (1999). From there, he describes different kinds of workspaces that learners need to exploit in a collaborative activity. The first one is personal. Learners have their own workspace where all information is hidden to the other participants. The second one is collective. Learners interact and work on collaborative devices. The central objective here is to create collective knowledge, resulting from the group activity. We need both individual and collective workspaces since most of the pedagogical scenarios in CSCL contain both individual and collaborative activities. This clear separation between personal and collaborative workspaces led to a lot of work in the CSCL and CSCW communities (Häkkinen et al. 2012 Pinelle et al. 2003; Marty and Carron 2011). Sometimes, in highly complex scenarios, learning activities in virtual and physical locations take place alternatively, resulting in splitting the personal activities in virtual environments and the collaborative ones in real life, as it is the case in the Janus project (Loiseau et al. 2013).

### 2.2 Territoriality on Collaborative Devices

When one wants to support collaborative activities in digital environments, data visualisation and interaction becomes a central issue. In the collaboration process, access to information must be natural. We therefore need to know where to display personal and group information. Scott and colleagues (2004) explain these territoriality aspects on collaborative devices used by several actors placed all around the same equipment. They based their study on the analysis of interactions between participants around a table and the observation of which workspaces become naturally unavailable. Users are intended to collaborate in order to achieve a given activity. The study concludes that the table is divided in three different territories dedicated to three different goals: personal activity, group activity and storage. Firstly, personal territory is situated near the user. All items situated in this area are considered private and belong to him/her. Secondly, group territory is located at the centre of the table. Everyone is therefore located at the same distance from the shared items. Thirdly, storage territory is at the border of personal territories. Items are reserved by a user but can be reused by another. This notion of territoriality has been widely used in designing applications for tabletops and large horizontal shared displays. For instance, Antle et al. (2011) designed a serious game on a tabletop where three players have to regulate a village growth while preserving the environment. They defined territoriality quite naturally by locating each private workspace on the side of the tabletop and a group workspace in the centre. We perceive the same needs on top of improvement of writing ergonomics and privacy.

### 2.3 Collaborative Activities in Multi-device Environment

With the generalisation of mobile devices, personal activities have progressively moved to new devices, smaller and lighter. Simultaneously the emergence of multi-touch tabletops accommodates for easier collaborative activities. New environments supporting collaborative learning must consider these changes and provide the learners with flexible environments where multi-device activities can take place.

As an example, the Caretta project (Sugimoto et al. 2004) proposes to build a city in a collaborative way, taking environment issues into account. Students act as a team around the board game and decide mutually where to place houses, factories and trees. Each student also has his/her own tablet where the common environment is reproduced. Everyone can therefore simulate any action and visualise the consequences on the city-state. All predictions are individual and if their results are appropriate, actions can be proposed to the group.

Using multi-device environments raises several more complex issues such as supporting awareness for the group (the identification of who is doing what on a collaborative space) or sharing and exchanging information between devices. MacKenzie et al. (2012) describe a platform where several users can simultaneously show their computer desktops on the same large screen, forming a common presentation of several devices used at the same time. This work thus proposes to display personal workspaces to only one shared workspace. Seyed and colleagues (2012) present an overview of the main gestures for cross-device interaction. They classify the gestures according to the specific actions required in a multi-display environment. They analyse interactions between collaborative devices (tabletop, digital board, etc.) and personal devices (tablets, mobile phone, etc.).

Scott and colleagues (2014) give some insights on how to transfer information between tabletops and tablets without user identification. They propose to consider a virtual bridge between these devices. Personal areas displayed on the tabletop (representing the bridges) allow each user to move information from the tabletop to his/her tablet, by dragging and dropping this information onto his/her bridge area. Conversely, when s/he pushes information from his/her tablet, it appears on his/her bridge area on the tabletop and everyone can see it. The authors offered a solution to the transfer problem between devices, proposing a solution for identifying who is doing the action. We want to extend this solution by considering all the territoriality aspects raised by multi-device environments.

## 3 Extension of Territoriality to Multi-device Environments

We saw in Sect. 2 how personal and group workspaces are central requirements to perform collaborative activities. In this section, we aim at defining how to extend the territoriality aspects in multi-device environments composed of a shared horizontal display (tabletop for instance) and several personal devices (such as tablets).

In order to better understand the relevant items to address, we describe a “classical” scenario of collaborative activities: a member of the group often works individually and obtains partial findings. S/he can then propose these results to the rest of the group, work on the group ideas with the other members, and can retrieve some information issued from the collaborative process. In this example, each member of the group involved in the collaborative activity therefore needs support for making both collaborative and individual activities.

The research questions addressed here are structured according to the MVC[1] paradigm. Firstly, we define the model in terms of territory arrangement. We specify

[1] *Model View Controller.*

workspaces required for supporting individual or collaborative activities and how to map such workspaces on a device, on specific areas of a device, or on a set of devices. Secondly, we describe the control aspects, specifying the interaction protocol to handle information between personal and collaborative workspaces. Thirdly, we address the view aspect by managing *partial views* on pieces of information. Partial views can allow to hide private parts of the information.

### 3.1 Model: Territories Arrangement

In a collaborative activity, taking place on a tabletop, we usually allocate specific tabletop areas for each individual workspace (Scott et al. 2004). Members of a group can be somewhat bothered by this approach for two main reasons: a/everyone performs individual tasks on the tabletop, even if these tasks have no interest for the other members of the group, and b/the user's favourite tools accessible in a personal environment are not available on the tabletop. Furthermore, many users prefer to work on their rough ideas alone, without being observed by others.

Taking advantage of multi-device environments allows us to reduce these drawbacks by proposing an extension of individual workspaces (Fig. 1). We propose to distribute personal workspaces between two specific areas: the *strictly personal workspace* displayed on a personal device (e.g. a tablet or a mobile phone) for individual tasks and the *perceptible personal workspace* displayed on the tabletop for awareness reasons. The other members of the group need to be aware of who owns what and when.

Users can thus perform their tasks on their personal devices, in a well-known environment, out of sight from the other members of the group. However, on collaborative devices, the perceptible personal workspace contains a partial view of all the objects (e.g. their title, their outline) being used by this user, providing the other members of the group with awareness.

### 3.2 Control: Inter-territory Actions

The design activity for multi-device environments implies to select natural and intuitive gestures for cross-device interactions since users have to focus on their specific activities, completing complex tasks with several other people, etc. Different studies, summarized in Seyed et al. (2012), have pointed out that some gestures are particularly well adapted to transfer information between devices. We based our approach on these studies, and we proposed different interactions according to the workspace, the device and the tasks to perform. We distinguish three situations: within the collaborative workspace, from collaborative workspace to personal workspace, and the other way round. The selected gestures are now quite obvious especially for our students but we summarize them below.

*From Personal to Collaborative Workspaces.* To move information from personal devices to the tabletop, "swipe up" interactions seem well adapted. For easy understanding, as a participant cannot keep a focus on anything, anywhere, at any time, this

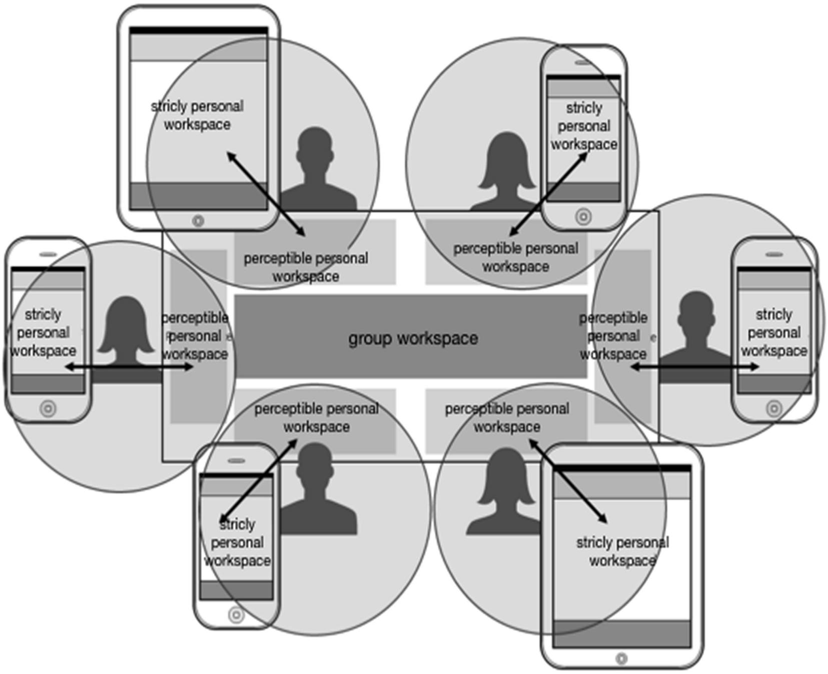



**Fig. 1.** Extension of the territoriality notion by connecting personal devices to the tabletop

information should appear close to his/her private workspace on the tabletop, *i.e.* in his/her perceptible personal workspace.

*Collaborative Workspace.* On tabletops, our interest is to enhance collaboration in the group. It means improving communication between participants. We offer attractive multi-touch properties of the device to make the different members of the group interact more. We propose to use well-known basic gestures such as “single tap”, “multi-tap”, “swipe”, “rotate” and “pinch”. Learners know how to use them and they will not be disoriented in front of the tabletop.

*From Collaborative to Personal Workspaces.* In order to transfer information from tabletop to tablet, we recommend using gestures like “swipe down”. The user puts his/her finger on the wanted area and brings it to his/her perceptible personal workspace. We prefer to use a bridge for transferring information since we need to know which user is performing the action (traceability issues).

### 3.3 View: Information Visibility

Designing applications for collaborative learning on multi-device environment induces changes for information visibility. This aspect is important since we need to adapt the visualisation to a particular device (responsive design) but also to consider what part of the information can be seen and by whom while designing our applications. We examine successively these visibility aspects through the different workspaces described previously.

*Strictly Personal Workspaces.* Personal devices (e.g. tablets) support personal activities. Information displayed on these devices is generally exclusively dedicated to the user and is not visible to the other members of the group. In some cases, during collaborative activities, a user can share information by showing directly the tablet content to his/her group. However, this usage tends to recede when using collaborative devices. The user prefers keeping his/her own information for him/herself, choosing what s/he wants to share. The simple interaction mechanisms proposed to exchange information between personal and collaborative devices (see Sect. 3.2) enables this behaviour.

*Group Workspace.* The group workspace corresponds to the central area of the shared display. In this area, each member of the group can see the shared information and interact with it. This workspace is the place where collective knowledge (Hadwin et al. 2010) is created. In the group workspace, users can choose to share their information with the group. Information must therefore be fully visible. This allows each member to interact with available information. They can move, zoom in, zoom out, rotate and even bring information to his/her private workspace. Due to motivation, discernible indicators must point out who the initial owner of a production is, and who made the modifications to this production. These awareness hints usually enhance collaboration (the group keeps a trace of who initiated an idea).

*Perceptible Personal Workspaces.* The parts of the personal workspaces located on tabletop, are displayed differently. We are at the edge between personal and collaborative devices, the bridges defined in (Scott et al. 2014). Each group member can see the others moving information to their private workspace. For awareness matters, this information remains visible on the private workspace but with fewer details. The "public" view of the information is specific to the activity and should be defined during the design process. For example, in case of textual information, we can choose to show only the first words of the text, while for an article, we can decide to exhibit the title, or to blur the information if it is private. Users can modify information later on their personal devices, but no such modifications will be shown to the other members. In some cases, we need to display a graphic animation showing that the object is being edited. This approach raises the problem of having the same piece of information duplicated on several devices with different views (more or less detailed) and is in our view, an effective extension to the theoretical concepts presented previously.

## 4 Case Study

We applied this proposal to a specific case study. This section presents the scenario of the activity and how the design method has been applied. The second part presents the results of an ecological experimentation of this activity carried out with first year students at the Institute of Technology in Chambéry (France). There were four classes of fourteen students, a teacher and three human observers.

### 4.1 A Scenario with Distinctive Steps

The scenario has been co-designed with teachers with the objective of reinforcing the "French grammar" level of their students. Students play a game where they have to perform several journalist tasks. Their objective is to be employed by a virtual newspaper office. To reach this goal, they follow a three-step scenario (collaboration, cooperation, and synthesis/debriefing).

During the first step, the class is split into two groups (homogeneous level). Geographically (or "territorially") speaking, they were in two contiguous rooms with one tabletop and several tablets (one for each user). The relative proximity between the groups creates natural emulation (they were able to see nearby the others discussing/working).

This first step is a collaborative task where the goal is to create news in brief (short texts also called "brèves" in the journalistic French dialect) describing news in relation to university life. We mainly apply our territoriality extension to this part. In this activity, the learners are intended to perform both individual actions (creation of articles: tablets will support personal activities) and collaborative actions (merging articles from different users, evaluating other members' articles). They need to share personal information with others or to collect shared objects (resulting from a collaborative activity) for working individually. Each learner can also annotate each item depending on grammar quality with three different smileys (good, medium, bad).

As a second step, students have to illustrate the articles they have created. A role is allocated to each member of the group. The illustration task is more a cooperative process with different complementary roles (headquarters, reporter/photographer, investigator). We tried for this task a more geographically distributed configuration: each role has a specific area of action, distant from the others.

During the last step, each group presents their work together. The committee distributes individual and collective points and elects the best group according to their ideas and their production. As shown in (Marton and Säljö 1976), this phase is key to reformulate the taught concepts for a deeper learning process and acts as a motivation goal: exposing their collaborative results to the other groups.

### 4.2 Experiment Results

We had scheduled the experimentation phase in June 2015. Its main objective was to observe if our proposal of territoriality arrangement in a multi-device environment was relevant in a learning scenario. We also were interested in observing the impact of design on collaborative behaviours (on group regulation for instance). Thus we analyzed participants' behaviours such as remarks, feedbacks, etc., and their impact on each member, with interesting exchanges concerning writing skills. Participants were also asked to fill out a questionnaire to give their overall impressions on playing the game and on how they experienced working together in a distributed environment. They were asked to rate different game design aspects using 5-point Likert scale and to answer several open-ended questions.

*User Experience About Territoriality and Collaborative Work.* Table 1 and Fig. 2 present the scores obtained to the different questions related to territoriality and collective work. Overall, participants' opinions were very high and homogenous (standard deviations never exceed 1.08). Participants enjoyed the game and the feedbacks collected were all positive. Everyone had a good experience working together in a distributed environment (Q7, Q12 and Q13). For instance, participant P11 declared: "Usually, I don't like working within a group, but this time it was great! I enjoyed the collaborative aspects".

**Table 1.** Questionnaire scores (means and standard deviation)

| | Question | Mean (sd) |
|---|---|---|
| Q7 | Did you have the feeling of having worked efficiently with the group? | 4.13 (0.56) |
| Q12 | Did you think that having a tablet and a tabletop is suitable for collaboration? | 4.28 (0.57) |
| Q13 | Did you like working with several devices? | 4.47 (0.58) |
| Q14 | Did the gesture to transfer information from personal to group workspace seem natural for you? | 4.42 (0.63) |
| Q15 | Did the gesture to transfer information from group to personal workspace seem natural for you? | 4.49 (0.58) |
| Q16 | Did you understand easily that you had a personal workspace on the tabletop? | 4.64 (0.56) |
| Q17 | Did you like having private information on your tablet that the others cannot see? | 4.26 (0.65) |
| Q18 | Did you find the voting process easy? | 4.12 (0.93) |
| Q19 | Did you find the gestures performed in the group workspace easy to perform? | 4.08 (0.68) |
| Q26 | Did you have difficulties in remote collaboration during the mobile activity? | 3.09 (1.08) |
| Q24 | Did you have difficulties in the transition between game sessions and real activities? | 1.79 (0.99) |

Regarding territoriality aspects, all the participants identified clearly the different workspaces and cross-devices interactions seemed natural (Q13, Q14, Q15). They also appreciated to own private information on their tablets (Q17). During the four sessions, we have observed that the students spent more time in editing their draft proposal alone, without stress, trying to do their best to propose a text that would interest the other members of the group. Finally, there is the same positive experience regarding the voting process and gestures performed on the group workspace on the tabletop (Q18, Q19). As soon as two or three texts were displayed on the tabletop, the students experimented the vote feature. No hesitation was noticed, but rather a change of attitude when they discovered that they had only a limited number of votes (5 positive and 5 negative). They decided to keep their voting tokens for later when really interesting (or unpleasant) ideas appeared.

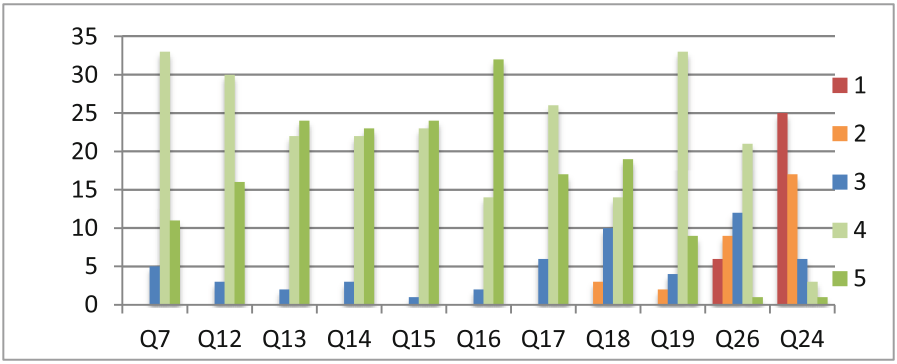



**Fig. 2.** Distribution of answers

*Design Impacts on Collaborative Behaviors.* During the experiments, we observed interesting results regarding the impact of the design and the scripting of the activity on participants' behaviours.

Firstly, we introduced a visual indicator of individual participation to the collective decision process. The indicator was represented as a red aura surrounding the personal workspace of each participant on the tabletop. Each time a participant was opposed to the rest of the group regarding the decision to make, the aura was growing. We observed efficient self-regulation of participants thanks to this indicator. When the students noticed that a red aura was growing around their "perceptible personal workspace", they asked for the meaning of this indicator. The impact of the explanation was the same in all the groups. They did not want to be responsible for a lack of collaboration. They started an explanation process about why they refused the proposal of the group, explaining what items of the text were not acceptable. The exchanges among participants were immediately richer than before and the red auras shrank.

Secondly, emulation, but no emergent competition feeling appeared between the two groups in parallel. Nevertheless, inside the groups, for the second phase, each role was associated to a specific territory: tabletop room for headquarters, inside the building for the photo-reporters and outside for the investigators (using drones to take photos): the first two roles were frustrated and even the photo-reporters were more interested in drones' control. Participants seemed to feel more difficulties to collaborate in this part of the activity (Q26) even if they remained taken up with the game scenario (Q24).

Finally, we observed quite surprisingly a strong impact of the tabletop characteristics on collective behaviours (Fig. 3). Participants using the tabletop with large borders stayed around the table to work even for personal work, talking together and forming a real group. Conversely, participants using the tabletop without border moved away from the table during individual tasks, trying to find a more comfortable way to elaborate their personal proposals, sitting down on the floor for instance. In this case, participants were not interacting. We then observed a less fluid collaboration process with very distinct phases (personal work then collaborative work). They were immersed in their individual task searching for the most comfortable conditions.

The swipe-up gestures sending the texts on the tabletop make the students become aware of the collaborative process. It is only once several texts have been sent to the tabletop that a student moved toward the tabletop, followed by the others. In fact, they understood at this stage that the collaboration activity (including votes) would take place around the table. It seems to us that it took more time for the group disseminated everywhere in the room to set up a real group atmosphere, but it did not alter the next steps of the activity: they immediately needed to exchange on the different pieces of information displayed on the tabletop.

**Fig. 3.** Illustration of collaborative activities on tablets and tabletop for our case study and the impact of unsuitability of a device (same phase).

## 5 Conclusion

In this article, we exposed the requirements for extending different aspects of territoriality concept for collaborative activities, especially in multi-devices environments. Tablets are used for personal tasks so that information on these devices is only visible by the owner. Tabletops are used for collaborative tasks. We therefore defined two kinds of workspaces on the shared display. The first one is private but partially perceptible by the others. Each user has thus his/her own workspace in front of him. The second one, at the centre of the tabletop, is public and all users can interact with it. In our proposal, we have defined both actions and information visualization for each workspace. Then, we have applied this concept for designing a particular learning scenario for improving French grammar level. Experimentations carried out reveal that the results in terms of collaborative activity have been particularly enhanced by this configuration. In addition, the concepts relative to extended territoriality have been extensively and easily used in the experiment. More generally, there is a real lack of such collaborative tools and we believe that this work opens a new way to design and develop such tools.

Finally, we would like to mention an important aspect of this experiment. It was included in a larger pedagogical session. The students started with an individual game activity in a GBL environment, where adapted small videos improved the students' French grammar level significantly: most of the assessments concerning such skills have been done in the game. The collaborative activity presented in this paper was used by the tutor as a re-enforcement of what has been learnt. The main topics to analyze here were thus related to improvement of collaboration.

**Acknowledgments.** We would like to thank the AIP Primeca for the support in this project, as well as the Rhone-Alpes French region.

## References

Antle, A.N., Bevans, A., Tanenbaum, J., Seaborn, K., Wang, S.: Futura: design for collaborative learning and game play on a multi-touch digital tabletop. In: Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 93–100. ACM, January 2011

Dillenbourg, P.: What do you mean by collaborative learning? Collaborative Learn. Cogn. Comput.Approaches, 1–19 (1999)

Dillenbourg, P., Tchounikine, P.: Flexibility in macro-scripts for computer-supported collaborative learning. J. Comput. Assist. Learn. **23**(1), 1–13 (2007)

Hadwin, A.F., Oshige, M., Gress, C.Z.L., Winne, P.H.: Innovative ways for using gStudy to orchestrate and research social aspects of self-regulated learning. Comput. Hum. Behav. **26** (5), 794–805 (2010). Advancing Educational Research on Computer-Supported Collaborative Learning (CSCL) Through the use of gStudy CSCL Tools

Häkkinen, P., Hämäläinen, R.: Shared and personal learning spaces: challenges for pedagogical design. Internet High. Educ. **15**(4), 231–236 (2012)

Hildmann, H., Uhlemann, A., Livingstone, D.: Simple mobile phone-based games to adjust the player's behaviour and social norms. Int. J. Mob. Learn. Organ. **3**(3), 289–305 (2009)

Kelle, S., Klemke, R., Specht, M.: Design patterns for learning games. Int. J. Technol. Enhanced Learn. **3**(6), 555–569 (2011)

Loiseau, M., Lavoué, E., Marty, J.C., George, S.: Raising awareness on archaeology: a multiplayer game-based approach with mixed reality. In: 7th European Conference on Games Based Learning (ECGBL 2013), pp. 336–343 (2013)

MacKenzie, R., Hawkey, K., Booth, K.S., Liu, Z., Perswain, P., Dhillon, S.S.: LACOME: a multi-user collaboration system for shared large displays. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion, pp. 267–268. ACM, February 2012

Marty, J.C., Carron, T.: Observation of collaborative activities in a game-based learning platform. IEEE Trans. Learn. Technol. **4**(1), 98–110 (2011)

Marton, F., Säljö, R.: On qualitative differences in learning: i—outcome and process. Br. J. Educ. Psychol. **46**, 4–11 (1976). doi:10.1111/j.2044-8279.1976.tb02980.x

Pernin, J.P., Mariais, C., Michau, F., Emin-Martinez, V., Mandran, N.: Using game mechanisms to foster GBL designers' cooperation and creativity. Int. J. Learn. Technol. **9**(2), 139–160 (2014)

Pinelle, D., Gutwin, C., Greenberg, S.: Task analysis for groupware usability evaluation: modeling shared-workspace tasks with the mechanics of collaboration. ACM Trans. Comput. Hum. Interac. **10**(4), 281–311 (2003)

Scott, S.D., Carpendale, M.S.T., Inkpen, K.M.: Territoriality in collaborative tabletop workspaces. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, pp. 294–303. ACM, November 2004

Scott, S.D., Besacier, G., McClelland, P.J.: Cross-device transfer in a collaborative multi-surface environment without user identification. In: 2014 International Conference on Collaboration Technologies and Systems (CTS), pp. 219–226. IEEE, May 2014

Seyed, T., Burns, C., Costa Sousa, M., Maurer, F., Tang, A.: Eliciting usable gestures for multi-display environments. In: Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, pp. 41–50. ACM, November 2012

Sugimoto, M., Hosoi, K., Hashizume, H.: Caretta: a system for supporting face-to-face collaboration by integrating personal and shared spaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 41–48. ACM, April 2004

# Refinement of a Q-matrix with an Ensemble Technique Based on Multi-label Classification Algorithms

Sein Minn[1(✉)], Michel C. Desmarais[1], and ShunKai Fu[2]

[1] Polytechnique Montreal, Montreal, Canada
{sein.minn,michel.desmarais}@polymtl.ca
[2] Huaqiao University, Quanzhou, China
fusk@hqu.edu.cn

**Abstract.** There are numerous algorithms and tools to help an expert map exercises and tasks to underlying skills. The last decade has witnessed a wealth of data driven approaches aiming to refine expert-defined mappings of tasks to skill. This refinement can be seen as a classification problem: for each possible mapping of task to skill, the classifier has to decide whether the expert's advice is correct, or incorrect. Whereas most algorithms are working at the level of individual mappings, we introduce an approach based on a multi-label classification algorithm that is trained on the mapping of a task to all skills simultaneously. The approach is shown to outperform the existing task to skill mapping refinement techniques.

**Keywords:** Student model · Skills modeling · Psychometrics · Q-matrix validation · Multi-label skills assessment

## 1 Introduction

Intelligent tutoring systems rely on efficient methods to assess the skills to perform tasks. These skills can involve factual knowledge, deep understanding of abstract concepts, general problem solving abilities, practice at recognizing patterns and situations, etc. Furthermore, a designer of a learning environment may focus on a particular subset of these skills. It might be the subset that is deemed appropriate for 10–12 years old kids endowed with a specific training. Or it might be a subset that more closely relates to a given topical or pedagogical perspective at the expense of alternative perspectives. For example a tutor may not care much about general problem solving abilities that require months to acquire, and focus on factual knowledge and rules that are easier to teach and assess, even though both the problem solving skills and factual knowledge are involved in the training and assessment material.

Whatever the motivation is for defining the skills behind the successful completion of tasks, a first point to emphasize is that for the same tasks, one skill definition may be considered appropriate for one context whereas another will be


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 165–178, 2016.
DOI: 10.1007/978-3-319-45153-4_13

required for another context. A second point to emphasize is that the definition of skills behind tasks, or the converse, the definition of tasks for a given set of skills, are non trivial and error prone processes.Therefore, tools to help a tutor, or a designer of a learning environment, validate a given mapping of skills to tasks would be highly valuable. Let us refer to this endeavour as the problem of *Q-matrix refinement*, where the Q-matrix represents the mapping of tasks to skills.

In this paper, we present a framework to help validate a Q-matrix called Multi-Label Skills Refinement (MLSR). We describe the method, setup, analysis, and results of a performance assessment of Q-matrix refinement.This approach can be considered an ensemble technique, since it combines refinements obtained from different algorithms to calculate its own refinements: Minimal Residual Sum Square (MinRSS), Maximum Difference (MaxDiff) and Conjunctive Alternating Least Square Factorization (ALSC). In addition, the approach uses features obtained from a large number of simulations with the refinements algorithms, and in particular an indicator of each algorithm's error rate over a given cell of the Q-matrix. The error rate computed from these simulations by using synthetic data, for which the ground truth is known.

The rest of this paper is organized as follow. Section 2 reviews the related work on the Q-matrices and techniques to validate them from data. Section 3 combining techniques with multi-label classification, Sect. 4 presents the error metric. Experimental results are found in Sects. 5 and 6 concludes and discusses future prospective.

## 2 Q-Matrices and Related work

An example of a Q-matrix mapping 11 items to 5 skills is given below. Item 4 requires skill 1 only, whereas item 11 requires skill 2 and 4. If all specified skills are required to succeed the item, the Q-matrix, it is termed conjunctive. If any of the required skill is sufficient to the item success, it is termed disjunctive. The compensatory version corresponds to the case where each required item increases the chances of success in some way. A well known model that relies on a conjunctive Q-matrix corresponds is the DINA model (Deterministic Input Noisy AND) and its disjunctive counterpart is the DINO model (Deterministic Input Noisy OR). In this paper, we focus in the conjunctive version of Q-matrices.

### 2.1 Q-matrix Refinement Techniques

Whereas we find a number of techniques to derive Q-matrices entirely from data (for eg. [1–4]), the current study focuses on a related problem: refining expert-given Q-matrices from data. The two techniques are closely related. The main difference can generally be considered as one of starting points: entirely data-driven Q-matrix definition starts from a random state, or from some predetermined state, whereas refinement techniques start from the expert's Q-matrix. However, very often, the general algorithms are the same (Table 1).

**Table 1.** Example of a Q-matrix

$$\begin{array}{c} \\ i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \\ i_6 \\ i_7 \\ i_8 \\ i_9 \\ i_{10} \\ i_{11} \end{array} \begin{array}{c} \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

We chose to use three Q-matrix refinement techniques that were studied in [5,6] for the purpose of comparison. They are state of the art techniques for static data: for which the student does not learn during data collection, as opposed to data from learning environments where it is expected the student will learn. The three techniques are described below.

**MinRSS:** Minimal Residual Sum Square (MinRSS) is from [2]. The algorithm first identifies the most likely skills profile of each individual based on a non parametric method introduced in [7]. Given a Q-matrix, this method finds the *ideal response vector* closest to the individual's real response vector based on the Hamming distance:

$$d_{hj}(r,\eta) = \sum_{j=1}^{J} |r_j - \eta_j| \tag{1}$$

where $r$ is the real response vector, $\eta$ is the ideal response vector, and $J$ is the number of items. Note that other measures such as entropy can be used in place of the Hamming distance.

Once the skills profiles are identified, the algorithm searches for Q-matrix that minimizes the residual sum of squares (RSS) between the predicted and the real results. It relies on a heuristic search that starts with the items that generate the most errors and stops when no changes reduces the the RSS.

This method is called MinRSS. It yields good performance under different underlying conjunctive models.

**MaxDiff:** de la Torre et al. [3,8] propose that a correctly specified q-vector for item $j$ should maximize the difference of probabilities of correct response between examinees who possess all the required skills and those who do not. Their approach relies on the DINA model:

$$P(X_j|\xi_j) = (1 - s_j)^{\xi_j} g_j^{(1-\xi_j)}$$

where $X_j$ is the probability of success to item $j$ and $\xi_j = 1$ if all skills required for that item are mastered, and $\xi_j = 0$ otherwise. The $s_j$ and $g_j$ parameters are respectively the slip and guess factors.

The approach consists in choosing the q-vector for an item $j$ that maximizes the difference in probabilities when all required skills are required and not:

$$q_j = \arg\max_{\alpha_l}[P(X_j = 1|\xi_j = 1) - P(X_j = 1|\xi_j = 0)] \tag{2}$$

de la Torre et al. [3] proposed a greedy algorithm that adds skills into a q-vector sequentially. This algorithm requires knowledge of $s_j$ and $g_j$ in advance. They are calculated by the EM (Expectation Maximization) algorithm.

**ALSC:** ALSC (Conjunctive Alternating Least Square Factorization) is introduced in Desmarais et al. [6,9]. The method relies on the standard Alternate Least Square technique to factorize student test results into a Q-matrix and a profile matrix. ALSC decomposes the results matrix $\mathbf{R}_{m\times n}$ of $m$ items by $n$ students as the inner product two smaller matrices:

$$\neg\mathbf{R} = \mathbf{Q}\,\neg\mathbf{S} \tag{3}$$

where $\neg\mathbf{R}$ is the negation of the results matrix ($m$ items by $n$ students), $\mathbf{Q}$ is the $m$ items by $k$ skills Q-matrix, and $\neg\mathbf{S}$ is negation of the the mastery matrix of $k$ skills by $n$ students (normalized for rows columns to sum to 1). By negation, we mean the 0-values are transformed to 1, and non-0-values to 0. Negation is necessary for a conjunctive Q-matrix.

The factorization consists of alternating between estimates of $\mathbf{S}$ and $\mathbf{Q}$ until convergence. Starting with the initial expert defined Q-matrix, $\mathbf{Q}_0$, a least-squares estimate of $\mathbf{S}$ is obtained:

$$\neg\hat{\mathbf{S}}_0 = (\mathbf{Q}_0^{\mathrm{T}}\,\mathbf{Q}_0)^{-1}\,\mathbf{Q}_0^{\mathrm{T}}\,\neg\mathbf{R} \tag{4}$$

Then, a new estimate of the Q-matrix, $\hat{\mathbf{Q}}_1$, is again obtained by the least-squares estimate:

$$\hat{\mathbf{Q}}_1 = \neg\mathbf{R}\,\neg\hat{\mathbf{S}}_0^{\mathrm{T}}\,(\neg\hat{\mathbf{S}}_0\,\neg\hat{\mathbf{S}}_0^{\mathrm{T}})^{-1} \tag{5}$$

iteratively until convergence. Alternating between Eqs. (4) and (5) yields progressive refinements of the matrices $\hat{\mathbf{Q}}_i$ and $\hat{\mathbf{S}}_i$ that more closely approximate $\mathbf{R}$ in Eq. (3). The final $\hat{\mathbf{Q}}_{\mathbf{i}}$ is rounded to yield a binary matrix.

## 3 Multi-label Skills Refinement

Each of the three techniques described above, MinRSS, MaxDiff, and ALSC, uses a substantially different algorithm from the others to refine a Q-matrix. In that respect, their respective outcome may be complementary, and we can hypothesize that they can be combined to provide a more reliable output than any single one. Furthermore, some algorithms are more effective in general, but may not be the best performer in all context. Defining the features that allows learning which algorithm provides the most reliable outcome in a given context is another objective of combining these techniques.

We first describe the data on which the multi-label skill refinement techniques are trained, and then describe the two algorithms.

### 3.1 Data to Train the Multi-label Skills Refinement Algorithms

Table 2 contains an excerpt of data used to train the multi-label skills refinement algorithms. Each line is a record for a single item to skills mapping. The right-most column contain the true labels. The left columns contain the suggested refinements from the different algorithms and *contextual factors* that may provide information about the most reliable technique refinement in a given context. They are:

**Stickiness** is the proportion of times a cell $i$ is misclassified by an algorithm $s$ when perturbating all other cells of the Q-matrix

$$St_{si} = \frac{\sum_{n=1}^{N}(r_n \neq p_n)}{N} \tag{6}$$

where $N$ is the set of perturbated cells (all cells but the target cell $i$), $r_i$ is the cell value in the original matrix and $p_i$ is the value in the pertubated matrix.

**Skills per row** indicates the number of skills required for a given item. An item may contain one or more skills.

**Skills per column** is the sum of the skills per columns. It is an indicator of how often this skill is required by the different items of the Q-matrix.

**Table 2.** Example of the data used for multi-label classification

| Items | MinRSS | | | | | | ... | Real values | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Prediction for skills $s_n$ | | | Stickiness for skills $s_n$ | | | ... | | | |
| | $s_1$ | $s_2$ | $s_3$ | $s_1$ | $s_2$ | $s_3$ | ... | $s_1$ | $s_2$ | $s_3$ |
| 1 | 1 | 1 | 0 | 0.04 | 0.04 | 0.00 | ... | 1 | 1 | 0 |
| 2 | 0 | 1 | 0 | 0.00 | 0.06 | 0.10 | ... | 0 | 1 | 1 |
| 3 | 1 | 1 | 1 | 0.20 | 0.05 | 0.00 | ... | 1 | 0 | 1 |
| 4 | 1 | 0 | 0 | 0.04 | 0.04 | 0.20 | ... | 1 | 0 | 0 |
| 5 | 1 | 0 | 1 | 0.00 | 0.04 | 0.04 | ... | 1 | 0 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

### 3.2 Multi-label Skills Algorithms

We transform the proposed outputs and contextual factors from three data driven techniques into a multi-label classification problem. We use synthetic data generated from 1000 permuted matrices for training and use real data for testing. The procedure for data generation of training and testing is shown in Fig. 1. The general idea is to introduce a perturbation in a Q-matrix and to run the refinement algorithms on the perturbed matrix to validate whether the perturbation

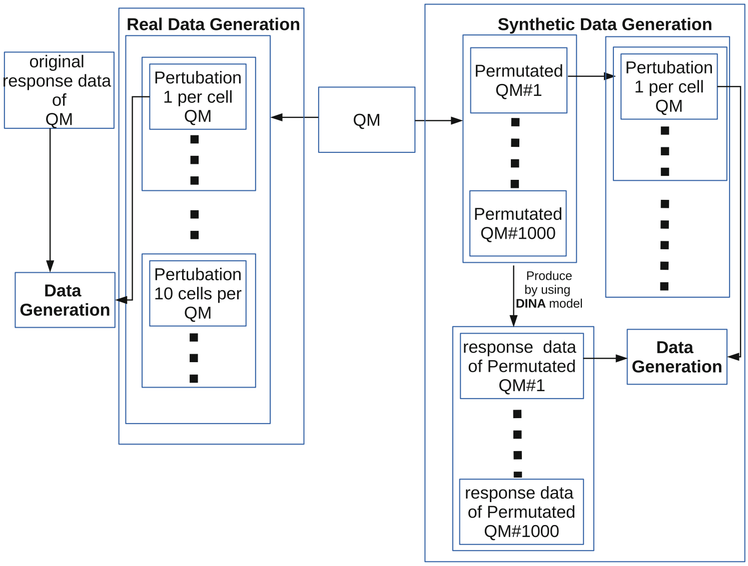



**Fig. 1.** Data Generation procedure of each Q-Matrix $QM_i$

is identified and whether false perturbations (false alarms) are introduced. From this process, we can measure the contextual factors of Table 2, namely the *stickiness* of a cell (tendency of generating a false alarm given a specific refinement algorithm), and which method is most reliable if an item has few or many skills, or if an skill is involved in many items or not. Given that the Q-matrices to generate the synthetic data are known, this provide the ground truth to do the training. Noise is introduced to make the data closer to real data and we use the original ratio of 0/1 in the perturbated matrix to create the 1000 permutations. See [6] for details.

Next, we follow the same approach as in [6], but instead of using a decision tree to predict a single cell in a Q-matrix, a multi-label classification algorithm is used for the predicting all skills of an item at once.

The generality of multi-label problems makes it significantly more complex to solve than traditional single-label (two-class or multi-class) problems. Only a few studies on multi-label learning are reported in the literature, which mainly concern the problems of text categorization, bioinformatics and scene classification.

Multi-label classification aims to predict a whole vector of labels at once, namely the item skills set in our case. We have a vector of skills for each item in our Q-matrices. So we can transform the proposed outputs of Q-matrices driven from the three refinement techniques and their contextual factors into multi-label classification problem and then we make final prediction by using those features. In this study, we use two multi-label classification methods: binary relevance method (Classifier chain method) [10] by using Naive Bayes classifier, and RAndom k-labELsets(Ensemble method) [11] by using the J48 decision tree algorithm.

**Binary Relevance Method with Naive Bayes.** The strategy of problem transformation is to use the one-against-all strategy by converting the multi-label problem into several binary classification problems. This approach is known as the binary relevance method (BR) [10]. A method closely related to the BR method is the Classifier Chain method (CC) proposed by Read et al. [10]. This method involves $Q$ binary classifiers linked along a chain. BR transforms any multi-label problem into one binary problem for each label.

Let us introduce some notation. Given an instance $X$ and its associated label set $l_i \subset |L|$, where its $l_i$ component of $|L|$ takes the value of 1 if $l_i \in |L|$ and 0 otherwise. In addition, let $N(x)$ denote the set of $x$ identified in the training set.

Hence this method trains $|L|$ binary classifiers $C_1, ..., C_{|L|}$. Each classifier $C_j$ is responsible for predicting the 0/1 association for each corresponding label $l_j \in\ L$.

BR with Naive Bayes (NB) method makes NB classifiers linked in a chain, such that the classifier for $l_i$ in the chain considers the classes predicted $l_1, l_2, ..., l_{i-1}$ from the previous classifiers as additional attributes. Thus, the feature vector for each binary classifier is extended with the class values (labels) of all previous classifiers in the chain. Each classifier in the chain is trained to learn the association of label $L_i$ given the features augmented with all previous class labels in the chain, $C_1$; $C_1$; $C_2$; ...; $C_{|L|}$. At classification time, the process starts at $C_1$, and propagates the predicted classes along the chain such that for $C_i$ it computes:

$$P(l_i) = \arg\max_{l_i} P(l_i|X, l_1, l_2, ..., l_{i-1}) \tag{7}$$

**RAndom k-labELsets with J48.** The ensemble methods for multi-label learning are developed on top of the common problem transformation or algorithm adaptation methods. The most well known problem transformation ensembles are the RAndom k-labELsets (RAkEL) system by Tsoumakas et al. [11]. RAkEL constructs each base classifier by considering a small random subset of labels and learning a single-label classifier for the prediction of each element in the power-set of this subset that transformed form multi-label problem.

In this experiment we use the single-label J48 classifier, an optimized implementation of the C4.5 or improved version of the C4.5. J48 constructs a Decision tree as an output.

## 4 Error Metric

The evaluation of methods for multi-label data requires different metrics than those used in the case of single label data. For the definitions of these metrics, we will consider an evaluation data set of multi-label examples $(x_i, Y_i), i = 1...m$, where $Y_i \subseteq L$ is the set of true labels and $Z_i$ is the set of predicted labels. This section presents metrics [12] that will be used in this experiment for the evaluation of our method.

**Hamming Loss** is a measure of how many times an instance label set is misclassified, i.e. a label not belonging to the instance is predicted or a label belonging to the instance is not predicted. The performance is perfect when $HammingLoss = 0$; the smaller the value of $HammingLoss$, the better the performance:

$$HammingLoss = \frac{1}{m}\sum_{i=1}^{m}\frac{|Z_i \Delta Y_i|}{M} \tag{8}$$

where $\Delta$ stands for the symmetric difference between two label sets. which is the theoretic equivalent of the exclusive disjunction (XOR operation) in Boolean logic for sets.

**Subset Accuracy:** To calculate the accuracy of vector of labels is truly classified. $SubsetAccuracy$ is defined as follows:

$$SubsetAccuracy = \frac{1}{m}\sum_{i=1}^{m} I(Z_i = Y_i) \tag{9}$$

**Example Based F-score** are calculated based on the average differences of the actual and the predicted sets of labels over all examples of the evaluation data set. The performance is perfect when $ExamplebasedF - score = 1$; the bigger the value ,the better the performance:

$$ExampleBasedF - score = \frac{1}{m}\sum_{i=1}^{m}\frac{2|Y_i \cap Z_i|}{|Z_i| + |Y_i|} \tag{10}$$

## 5 Experimental Study

For the sake of comparison, we use the same datasets as the ones used in Desmarais et al. (2015) [6,13]. It is a well known data set in fraction algebra from Tatsuoka's work (Tatsuoka, 1984) [13]. It consists 3 expert-driven Q-matrices and one SVD driven Q-matrix with a same data set. These allow us to analyze possibility of different models (Q-matrices) over the same data source. Table 3 provides the basic information and source of each dataset.

**Table 3.** Q-matrix for validation & explanation of category

| Q-Matrices | Number of | | | Description |
|---|---|---|---|---|
| | Skills | Items | Cases | |
| QM1 | 3 | 11 | 536 | Expert driven from [14] |
| QM2 | 5 | 11 | 536 | Expert driven from [3] |
| QM3 | 3 | 11 | 536 | Expert driven from [15] |
| QM4 | 3 | 11 | 536 | Data driven, SVD based |

**Table 4.** Hamming Loss result of Synthetic Data (single perturbation)

| QM | MinRSS | MaxDiff | ALSC | RAkEL.1 | BR.1 | RAkEL.2 | BR.2 | RAkEL.3 | BR.3 | RAkEL.4 | BR.4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| qm1 | 0.53 | 0.09 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| qm2 | 0.42 | 0.41 | 0.44 | 0.00 | 0.01 | 0.00 | 0.09 | 0.00 | 0.01 | 0.00 | 0.00 |
| qm3 | 0.63 | 0.64 | 0.55 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 |
| qm4 | 0.58 | 0.59 | 0.53 | 0.00 | 0.00 | 0.00 | 0.25* | 0.00 | 0.00 | 0.00 | 0.00 |

**Table 5.** SubSet Accuracy result of Synthetic Data (single perturbation)

| QM | MinRSS | MaxDiff | ALSC | RAkEL.1 | BR.1 | RAkEL.2 | BR.2 | RAkEL.3 | BR.3 | RAkEL.4 | BR.4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| qm1 | 0.19 | 0.85 | 0.18 | 1.00 | 0.98 | 1.00 | 1.00 | 1.00 | 0.98 | 1.00 | 0.99 |
| qm2 | 0.24 | 0.25 | 0.13 | 1.00 | 0.93 | 1.00 | 0.73 | 1.00 | 0.94 | 1.00 | 0.98 |
| qm3 | 0.00 | 0.02 | 0.00 | 1.00 | 1.00 | 1.00 | 0.91 | 1.00 | 1.00 | 1.00 | 1.00 |
| qm4 | 0.07 | 0.08 | 0.04 | 1.00 | 0.98 | 1.00 | 0.50*** | 1.00 | 0.98 | 1.00 | 0.98 |

**Table 6.** Macro averaged F-measure result of Synthetic Data (single perturbation)

| QM | MinRSS | MaxDiff | ALSC | RAkEL.1 | BR.1 | RAkEL.2 | BR.2 | RAkEL.3 | BR.3 | RAkEL.4 | BR.4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| qm1 | 0.54 | 0.90 | 0.54 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 |
| qm2 | 0.68 | 0.71 | 0.64 | 1.00 | 0.99 | 1.00 | 0.92 | 1.00 | 0.99 | 1.00 | 1.00 |
| qm3 | 0.06 | 0.10 | 0.16 | 1.00 | 1.00 | 1.00 | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 |
| qm4 | 0.37 | 0.37 | 0.42 | 1.00 | 1.00 | 1.00 | 0.66** | 1.00 | 1.00 | 1.00 | 1.00 |

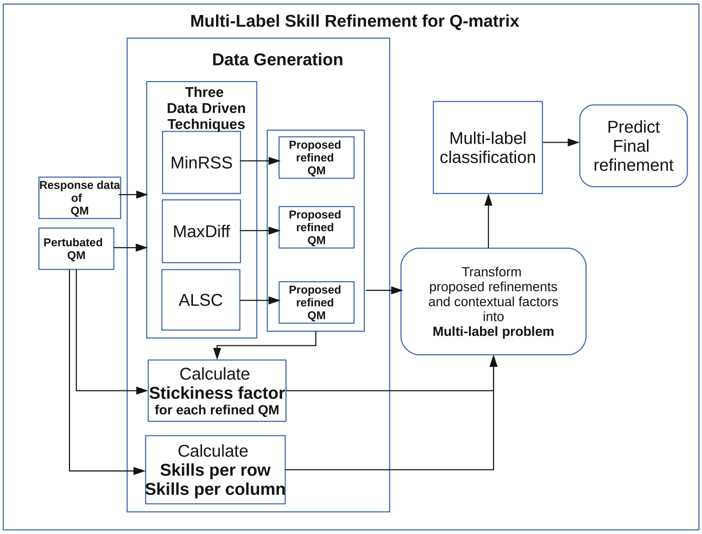



**Fig. 2.** Refinement procedure of each Q-Matrix $QM_i$

All experiments were done with 10 fold cross validation. We relied on the CDM [15] and NPCD packages which provided both the code for three basic data driven techniques and the data, and mulan [12] for multi-label classification.

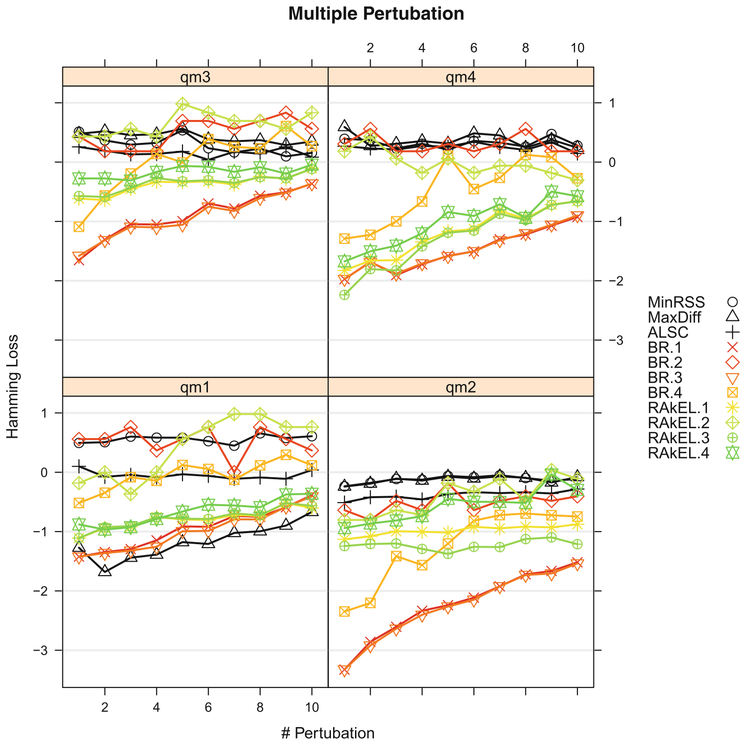



**Fig. 3.** Real Data: Logit value of Hamming Loss as a function of the number of perturbations (Color figure online)

We use Hamming loss, Subset Accuracy and Example based F-measure to assess the performance of the different algorithms.

The experimental results are reported in Tables 4, 5 and 6 for synthetic data, and in Figs. 3, 4 and 5 for real data. Four variations of the two multi-label approaches are reported for both real and synthetic data (BR.$n$ and RAkEL.$n$). They correspond to different training data. The four variations respectively contain:

BR.1/RAkEL.1: item number, outputs from three different basic algorithms
BR.2/RAkEL.2: item number, stickiness factors from three different algorithms
BR.2/RAkEL.3: combination of item number, outputs, row sum and column sums.
BR.2/RAkEL.4: combination of item number, outputs,stickiness factors, row sum and column sums.

For synthetic data, a single cell is perturbed. We can see from Tables 4, 5 and 6 that most of multi-label skill refinement methods can recover over 99 % for all Q-matrices and even the performance reaches 100 % in terms of subset accuracy and macro averaged F-measure. The standard deviations of all values

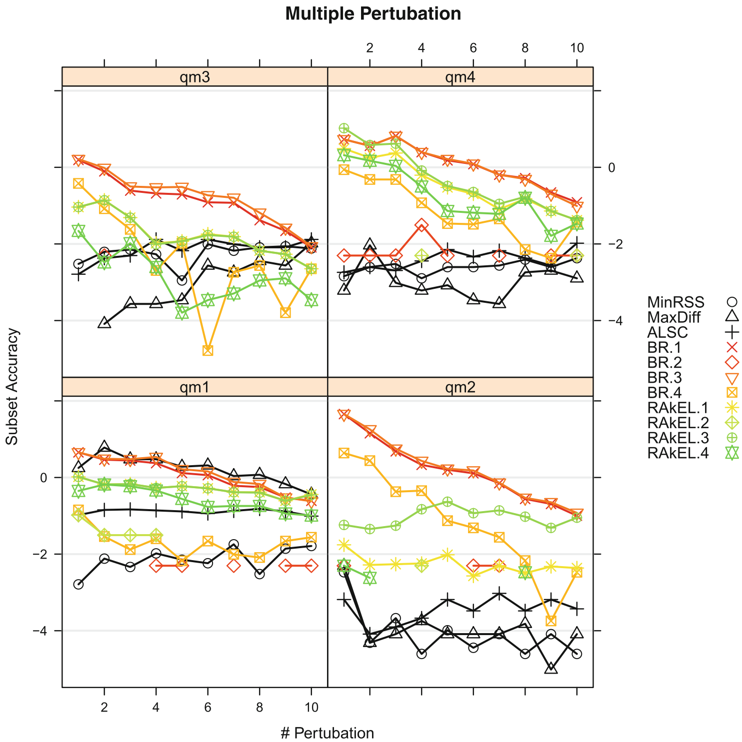



**Fig. 4.** Real Data: Logit value of Subset Accuracy as a function of the number of perturbations (Color figure online)

except the ones marked with stars are below 0.05 ($* \rightarrow sd < 0.01$, $** \rightarrow sd < 0.02$, $*** \rightarrow sd < 0.05$), which makes the vast majority of differences statistically significant. Clearly, all methods using multi-label refinement algorithms perform much better than any single method and the results are also substantially better than those of the single-cell decision tree method reported in [6].

For real data, multiple perturbations are introduced and the results are shown as figures to better visualize the trends as a function of the number of perturbations. A logit scale is used which can be considered a good estimate of the relative remaining error on a scale of $[0, 1]$ (for eg., it displays a relative error reduction in accuracy from 0.90 to 0.95 as similar to the reduction from 0.99 to 0.995). The black lines show the results of the three individual refinement algorithms, and the coloured lines show the multi-label algorithms results.

As expected, the performance declines with the number of perturbations. BR.1 and BR.2 show the best performances in general. However, the results for QM1 shows that the MaxDiff method has a performance relatively close to the these two methods, BR.1 and BR.1

These results reveal a trend in the performance of our method: it underperforms with fewer skills. For eg., the 5-skills QM2 shows a better performance

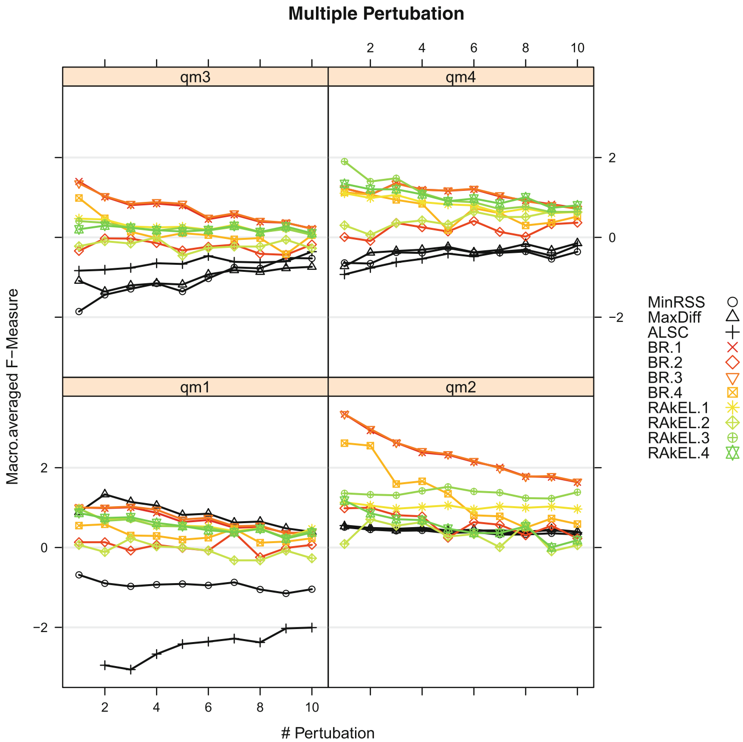



**Fig. 5.** Real Data: Logit value of example-based F-measure as a function of the number of perturbations (Color figure online)

than the 3-skills QM1, QM3 and QM4. Furthermore, QM1 has only two skills that really vary across items (skill 1 is required by all) and it is the method for which the performance of the multi-label approach is the worst.

## 6 Conclusion and Future Work

In this paper, we represent the multi-label skills to tasks refinement methods, that combine three data driven techniques and two multi-label classification techniques. Experiments with 3 expert driven Q-matrices and 1 Q-matrix driven from SVD, show the proposed refinement methods generally outperform the stand alone refinement algorithms. However, for real data, a Q-matrix with only two discriminant skills does not prove more effective than the MaxDiff refinement algorithm and the general pattern suggests the more skills involved, the better the BR.1 and BR.2 approaches perform.

As with previous work with ensemble techniques [6, 16], the experiments were conducted with static data, where the student does not learn during the data gathering process. Dealing with dynamic data, which is typical of traces collected from learning environments, imposes another challenge and a few researchers have done valuable work in that direction [17–19].

**Acknowledgements.** This work is funded by the NSERC Discovery funding awarded to the second author.

## References

## All links were last followed on June 20, 2016.

1. Barnes, T.: Novel derivation and application of skill matrices: the Q-matrix method. In: Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.D. (eds.) Handbook on Educational Data Mining, pp. 159–172. CRC Press, Boca Raton (2010)
2. Chiu, C.Y.: Statistical refinement of the Q-matrix in cognitive diagnosis. Appl. Psychol. Measur. **37**(8), 598–618 (2013)
3. de la Torre, J.: An empirically based method of Q-matrix validation for the DINA model: development and applications. J. Educ. Measur. **45**(4), 343–362 (2008)
4. Nižnan, J., Pelánek, R., Řihák, J.: Mapping problems to skills combining expert opinion and student data. In: Hliněný, P., Dvořák, Z., Jaroš, J., Kofroň, J., Kořenek, J., Matula, P., Pala, K. (eds.) MEMICS 2014. LNCS, vol. 8934, pp. 113–124. Springer, Heidelberg (2014)
5. Desmarais, M., Beheshti, B., Xu, P.: The refinement of a Q-matrix: assessing methods to validate tasks to skills mapping. In: Educational Data Mining (2014)
6. Desmarais, M.C., Xu, P., Beheshti, B.: Combining techniques to refine item to skills Q-matrices with a partition tree. In: Educational Data Mining (2015)
7. Chiu, C.Y., Douglas, J.: A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. J. Classif. **30**(2), 225–250 (2013)
8. de la Torre, J.: Dina model and parameter estimation: a didactic. J. Educ. Behav. Stat. **34**(1), 115–130 (2009)
9. Desmarais, M.C., Naceur, R.: A matrix factorization method for mapping items to skills and for enhancing expert-based Q-matrices. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS, vol. 7926, pp. 441–450. Springer, Heidelberg (2013)
10. Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. Mach. Learn. **85**(3), 333–359 (2011)
11. Tsoumakas, G., Katakis, I., Vlahavas, I.: Random k-labelsets for multi-label classification. IEEE Trans. Knowl. Data Eng. **23**(7), 1079–1089 (2011)
12. Tsoumakas, G., Katakis, I., Vlahavas, I.: Mining multi-label data. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 667–685. Springer, Heidelberg (2010)
13. Tatsuoka, K.K.: Rule space: an approach for dealing with misconceptions based on item response theory. J. Educ. Measur. **20**(4), 345–354 (1983)
14. Henson, R.A., Templin, J.L., Willse, J.T.: Defining a family of cognitive diagnosis models using log-linear models with latent variables. Psychometrika **74**(2), 191–210 (2009)
15. Robitzsch, A., Kiefer, T., George, A.C., Uenlue, A.: CDM: Cognitive Diagnosis Modeling, R package version 4.5-0 (2015)

16. Xu, P., Desmarais, M.C.: Boosted decision tree for Q-matrix refinement. In: 9th International Conference on Educational Data Mining, 6 June–2 July 2016, Raleigh, NC, USA (2016, to appear)
17. Matsuda, N., Furukawa, T., Bier, N., Faloutsos, C.: Machine beats experts: automatic discovery of skill models for data-driven online course refinement. Educ. Data Min. **2014**, 101–108 (2014)
18. González-Brenes, J.P.: Modeling skill acquisition over time with sequence and topic modeling. In: AISTATS (2015)
19. Aleven, V., Koedinger, K.R.: Knowledge Component (KC) approaches to learner modeling. In: Design Recommendations for Intelligent Tutoring Systems, p. 165 (2013)

# When Teaching Practices Meet Tablets' Affordances. Insights on the Materiality of Learning

Jalal Nouri(✉) and Teresa Cerratto Pargman

Computer and Systems Sciences, Stockholm University, Stockholm, Sweden
jalal.tessy@dsv.su.se

**Abstract.** Research on tablets in schools is currently dominated by the effects these devices have on our children's learning. Little has yet been said about how these devices contribute and participate in established school practices. This study delves into the questions of what do tablet-mediated teaching practices look like in Swedish schools and how are these practices valued by teachers? We collected data in four Swedish schools that were part of the one-to-one program financed by their municipalities. We apply qualitative and quantitative analysis methods on 22 deep interviews, 20 classrooms observations and 30 teachers' responses to an online survey. The study identifies a set of tablet-mediated teaching practices that lead to a deeper understanding of how affordances of media tablets configure contemporary forms of learning.

**Keywords:** Tablets · Affordances · Teaching practices · Mobile learning · Materiality of learning

## 1 Introduction

Mainstream research on mobile technology and media tablets in the area of education has so far focused on the implications of using mobile devices in formal learning activities [1, 2]. Roschelle et al. [3], and Chou, Block and Jesness, [4], have for example, reported on a correlation between use of mobile devices and enhancement of student engagement as well as progress of students' achievement. Other scholars in the field have underscored the proliferation of learning activities including interactive content creation [5]. Moreover, the open and easy access to information afforded by mobile devices while sitting in the classroom [6] as well as support for user-generated contexts [7] seem to modify power relations between teachers and students [8] and to have an impact on learners' shared epistemic agency [9]. Further, studies also point to the potential role of mobile technology to foster students' creativity and collaboration [10]. These studies have much contributed to understanding the value of using media tablets in teaching activities, however, most of them have approached media tablets and their affordances [11] as disembodied from their everyday use [9]. As such these studies have overlooked aspects associated to emerging teaching designs [12] and teaching practices [5]. Still underexplored is how tablet-mediated school practices emerge in the classroom as well as how they bring in new aspects that contribute to the


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 179–192, 2016.
DOI: 10.1007/978-3-319-45153-4_14

quality and meaningfulness of students' learning [13]. Taking a mediated action lens on technology affordances, [11], the paper presents a study we conducted in four Swedish schools which have been part of a 1:1 (one-to-one program) since 2012. The aim of the study was to understand what tablet-mediated teaching practices look like and how teachers value them. A mediated action perspective [11] is here chosen because it provides us with tools to look at the material characteristics of technology, in this particular case, media tablets. These material characteristics of technology thereafter called affordances are understood as a relational property of a three-way interaction between the person, mediational means, and cultural environment [6, 11]. Such conceptualization differs from others [14] as it situates affordances within a socio-cultural milieu and it emphasizes the dynamic and situatedness of the concept.

The present study delves into the materiality of the school activities identifying a set of tablet-mediated teaching practices that are entangled to the following affordances: persistence of the digital medium, multimodal character of the content of the applications and portability & ubiquity of media tablets. These material characteristics of the tablets afford a series of teaching and learning activities, we contend, play a central role in configuring the school practices observed in the study. As such, the study contributes to a deeper understanding of the weight that the specific design of tablets has on everyday teaching activities and practices. The paper also contributes to furthering current understanding of how media tablets, regarded as sociocultural artefacts, participate and configure contemporary forms of learning.

## 2 Description of the Methods and Context of the Study Chosen

### 2.1 Schools and School Subjects Targeted

We selected four Swedish elementary schools to conduct a study on emergent practices and transformations tied to the use of tablets in school classrooms. The schools selected obtained the tablets 3 years before we started our study in December 2013. Since then, we observed how and for which purposes the teachers had incorporated tablets in their teaching. These schools took part of the 1:1 tablet program providing students from 3rd grade to 9th grade with individual tablets (i.e. ipads). Two of the selected schools were public schools in the Stockholm area. They were chosen to participate in the program as they are considered "special schools" due to the heterogeneity of pupils' socio-economic, cultural backgrounds and languages proficiencies. The other two schools were located in Växjö, south of Sweden, and were private schools and comparatively more homogenous that the schools in Stockholm, at least in terms of pupils' socio-economic, cultural backgrounds and languages. The schools had in common a type of organization consisting among others of a dedicated leading group of teachers and IT pedagogues mainly responsible for spreading the use of tablets among teaching staff. As such, all the schools observed selected a group of teachers (10 % of the total amount of teachers in the school) to lead the introduction of the tablets in the classrooms through the following activities: (1) organizing workshops with the staff where learning platforms and main apps were demonstrated, (2) strategizing the introduction of the use of the media tablet into all school subjects, (3) discussing gained experiences with specific apps (4) choosing useful apps to purchase collectively. As such a dedicated group was actively leading the integration of the tablets in each of the schools studied schools.

The study focused specifically on the subjects of Natural Sciences and Mathematics as well as English and Swedish in grades 6–8 (age range 11–14 years old). We started to visit the schools in December 2013 and finished the data collection in December 2015.

### 2.2 Data Collection Methods

We conducted 22 deep interviews with school teachers (i.e., 60 min each), we observed 20 classrooms (i.e., 45 min each) and we collected 30 teachers' responses to an online survey. The teachers in the sample (n=30), consist of a rather experienced group of teachers, as they have taught for an average of 14 years (SD=10.11). The teachers interviewed were actively using the media tablets in their classrooms since the school became part of the 1:1 program in 2011.

The *interviews* of semi-structure nature were conducted at the schools. They consisted of questions covering four areas: 1- demographic questions (age, years of private tablet use, years of teaching, etc.) 2- questions about how teachers use tablets and how tablets support teaching, 3- questions about how pupils, according to the teachers, use tablets and are a support for learning, 4- questions about teachers' perceptions of benefits and disadvantages of using tablet in the classroom. A tape-recorder was used when conducting the interviews. The interviews were all conducted in Swedish.

Data were also collected through *field notes, photos and video-recorded events*. The school subject of the classrooms observed were: Natural sciences (chemistry and biology), Mathematics, English and Swedish. Both authors participated in the classrooms observations. We wrote notes, took photos (with the aim to provide a context to our notes and help us remember afterwards the classrooms observed), we also video-recorded when possible. The foci of the observations were: (1) type of activities conducted with and without the tablets, (2) tools (analog or digital- inclusively apps) employed; and specifically their main affordances and (3) observed tensions within teacher-tablet-pupil interactions.

With the purpose to examine how teachers perceived the value of and the relations between the thematic practices identified, we constructed a survey. The survey was sent to the participating teachers in the study. It consisted of 8 scales with a total of 53 items (see Table 1).

An exploratory factor analysis with principal component extraction was performed in an attempt to refine the instrument. After factor analysis, 12 items that did not load on any factors or were highly cross-loaded on multiple factors, were removed. Accordingly, the refined instrument used for analysis consisted of a total of 41 items. Overall, Cronbach's alphas were calculated for scales 2-8 with values ranging from 0.69 to 0.84. The survey was developed and administered through a web tool.

**Table 1.** Overview of the survey.

| Scale | Items | Focus |
|---|---|---|
| 1 | 22 | Demographic questions and general attitudes and perceptions of using tablets for teaching and learning |
| 2 | 5 | Teacher's perceptions of the benefit and frequency of using tablets for increasing student motivation and engagement |
| 3 | 4 | Teacher's perceptions of the benefit and frequency of using tablets for organizing learning |
| 4 | 7 | Teacher's perceptions of the benefit and frequency of using tablets for multimodal teaching and learning |
| 5 | 4 | Teacher's perceptions of the benefit and frequency of using tablets for documenting learning |
| 6 | 4 | Teacher's perceptions of the benefit and frequency of using tablets for assessing and providing feedback |
| 7 | 4 | Teacher's perceptions of the benefit and frequency of using tablets for communicating |
| 8 | 3 | Teacher's perceptions of the benefit and frequency of using tablets for distributing learning |

### 2.3 Data Analysis Methods

The interviews were fully transcribed and together with the field notes, photos and video-recordings from the classroom observations, were first independently coded and then collaboratively re-coded by both authors. We used procedures from content analysis [15] that supported the identification of conceptual threads from the text corpus obtained. The aim of our coding was the identification of tablet-mediated teaching practices. The content analysis resulted in the identification of a total of 7 themes corresponding to tablet-mediated teaching practices and a total of 25 specific categories of tablet-mediated pedagogical practices. The themes identified are: (1) organization of teaching and learning material; (2) documentation; (3) multimodal teaching and learning; (4) motivating pupils' engagement; (5) assessment and provision of formative feedback: (6) e-mail communication and (7) mobile learning.

## 3 Findings

This section reports on results from both qualitative and quantitative analysis performed on the interviews, field notes and the survey. The section introduces first, the 7 tablet-mediated teaching practices identified. That is followed by a presentation of the relation between teaching practice identified and teacher's perceptions of the value of using tablets in the classroom.

### 3.1 Tablet-Mediated Teaching Practices

The content analysis of 20 classroom field notes and 22 interviews with the teachers resulted into the identification of the seven thematic tablet-mediated pedagogical practices with associated categories of practices. See Table 2 for an overview of the identified thematic and category of practices. The table also displays the device's affordances we observed were associated to the practices identified.

**Table 2.** Overview of the tablet-mediated teaching practices in the schools studied

| Thematic Practice | Categories of practice | Tablet affordance |
|---|---|---|
| Organizing teaching and learning | Centralizing and sharing instructions, learning material and assignments | Persistence of the digital medium |
| Documentation | Supporting self- reflection, Supporting metacognition<br>Providing individualized learning, Increasing access parental insight | Multimodal channels<br>Persistence of the digital medium |
| Multimodal teaching and learning | Multimodal presentation of teaching and learning material, Self-construction of teaching and learning material, Re-using learning material from the Internet, Representing and visualizing facts, Supporting language learning (pronunciation through sound and comprehension through videos). Tackling reading and writing difficulties | Multimodal channels: sound, image, text |
| Motivation and engagement | Game-based learning and individualization | Multimodal channels<br>Portability of the mobile device |
| Assessment and provision of feedback | Collaborative assessment/feedback, Individual assessment/feedback, Class assessment/feedback, Automatic assessment/feedback | Multimodal channels<br>Persistence of the digital medium |
| E-mail communication | Teacher – Student | Persistence of the digital medium |
| Mobile learning | Flexibility and mobility | Portability and ubiquity of the mobile device |

Each of the practices identified were incorporated into the survey. We sent out the survey with the purpose to find out how teachers value such practices in terms of (1) use frequency of the tablet and (2) perceived usefulness in the classroom. See results in Table 3.

The teaching practice called organization of teaching and learning material dominated as the most valued tablet-mediated teaching practice followed by documentation and multimodal teaching and learning. Practices oriented to using the tablet for

**Table 3.** Overview of the tablet mediated practices in relation to how teachers valued each of them in their teaching.

| Practices and activities | Valued theme[a] |
|---|---|
| Organizing teaching and learning | M = 6.31, SD = 1.94 |
| Documentation | M = 5.56, SD = 1.92 |
| Multimodal teaching and learning | M = 5.17, SD = 1.93 |
| Motivation and engagement | M = 4.56, SD = 1.87 |
| Assessment and provision of feedback | M = 4.44, SD = 1.94 |
| E-mail communication | M = 2.44, SD = 1.76 |
| Mobile learning | M = 2.42, SD = 1.93 |

[a]Valued theme represents composite variables measuring how frequent and how useful thematic practices are on a scale from 1 to 8 according to the teachers.

motivating children and, as a support for assessing children's progress as well as for providing feedback, were also mentioned. Finally, using the tablet for e-mail communication and mobile learning were the least valued practices. In order to give the reader a sense of the tablet-mediated teaching practices identified, we describe each of them in detail in the following sections.

**Organization of Teaching and Learning.** From the analysis of the data, it Emerges that this practice is tightly related to the use of the learning management system (LMS) teachers use daily at the school. Specifically developed for tablets, LMS such as Schoolsoft, Learnify, I Tunes U, were daily used in the schools studied for creating and organizing learning material. In particular, the teachers we observed mentioned in the interviews the creation of instructions for individual assignments and group activities. Teachers explained to us they asked pupils to submit their assignments through the system so they can provide individual feedback that is then saved in the system for future consultation. Teachers also mentioned that material such as grading criteria, tests, homework, were uploaded and made available in the LMS facilitating centralization and distribution of learning material for the pupils. One of the teachers mentioned: *"Instead of handing out 300 papers every week, I have chosen to make all course material available in the system. I have uploaded instruction films, assessment material, homework, exams, everything. So, instead of referring to lost papers, I refer students to their tablets"* (Steve, natural science, Stockholm).

**Documentation.** Using tablets for documentation purposes was the second most valued thematic practice according to the teachers (M = 5.56, SD = 1.92). This practices involved for instance supporting self-reflection. In one of the class observations, we for example noted that students documented their lab work in chemistry with their tablets in form of text, tables and photos taken with the camera, that were then uploaded to the LMS system. The material available in the LMS was then commented by the teacher (after the activity), and collectively analysed by the entire classroom in a subsequent activity the next day. One of the teachers mentioned: *"the work students upload to the learning management system through their tablets is revisited for repetition and further analysis"* (Laura, Swedish, Stockholm).

Another aspect that the documentation practice supported was pupils' metacognition, in the sense, children were encouraged to create digital portfolios consisting of presentations combining text, images and audio. In the Swedish class, digital portfolios were then used by pupils to revisit words forms, adjectives, and idiomatic expressions and to stimulate pupils to reflect over their own learning progression. One of the teachers interviewed mentioned: *"I use digital portfolios that enable them [pupils] to monitor their own learning and compare their own performances".* (Martha, Swedish, Växjö).

Through the multimodality afforded by tablets' numerous apps, these devices facilitate teachers create digital portfolios and e-books. These portfolios and e-books support among others communication with parents who are interested in knowing what their children do at school.

**Mobile Learning.** Another way to organize teaching that was mentioned by the teachers was related to portability and ubiquity of the tablet. In this regard, we observed tablets facilitated allocating tasks to groups working in different rooms. Usually the stronger pupils, worked outside of the classroom while the teacher focused on the students with special needs inside the classroom. This was possible due to the fact that the information about the assignment was not only both displayed on the classroom whiteboard but also available on each children's tablets. Considering tablets' portability, one would expect that such affordance is often used to support mobile learning activities. In this study, we did surprisingly not found mobile learning activities were frequent or considered beneficial for learning (M = 2.42, SD = 1.93). However, the portability and ubiquity of the devices were mentioned as facilitating continuity of school activities, especially when pupils miss assignments due to absence or other. One of the teachers mentioned: *"It [mobile learning]helps when students are not here, if they are sick for instance, they can work on the same things as we do in the class from home. That helps us to reach the course objectives."* (Morten, English, Växjö).

**Multimodal Teaching and Learning.** rom the data analysed, it emerged that a large number of categories of practices was associated to the utilization of multimodal affordances of the tablets (i.e., sound, image, text). Using the tablet in the classroom was perceived by the teachers as a frequent and beneficial (M = 5.17, SD = 1.93) manner to teach as tablets by their affordances invite teachers to include multimodality in their teaching. Multimodal ways to teach and learn were exemplified by "presentations" pupils constructed using diverse applications for saving and managing photos, sound, video and text. Another central category of this practice was the creation of learning material such as: storytelling, e- books, portfolios, e-posters, movies, animations, interactive drawings made by the pupils. In the following excerpt, a teacher exemplifies one such emergent multimodal construction practice: *"They [pupils] looked at a video that is called "what does the fox say". Then they got the assignment to construct an own version of the video, and that can be done in different ways, they can for example sing, record themselves, record others, and then play it for the class in case they don't dare to stand in front of the class ... the tablet allows me to offer the students more ways to learn in the same classroom. It is not the case that all have to write texts"* (Lisa, English teacher, Stockholm). This example, illustrates the use of media tablets for practicing English while developing skills for expressing meaning beyond the text mode. Besides construction, it was revealed that media tablets frequently were used to find, consult and eventually reuse multimodal learning material available on the Internet. Using Internet in the classroom was especially appreciated as teachers mentioned, Internet extends the knowledge sources used in schools and the possibilities to present knowledge through different modalities. *"instead of as we did before telling the students to look it up in the book, we tell them to find the information on the internet in form of videos, images or texts".* (Petter, mathematics, Stockholm).

Teachers specifically mentioned the fascination pupils have for the image, which they believe helps pupils, especially those presenting weak language comprehension abilities in Swedish or English, with meaning-making processes. However, teachers also mentioned that the emotional relation pupils develop with the image motivated teachers to think seriously about how teach pupils to critically think about material and sources they find on the Internet or elsewhere. One activity that supports this goal is for instance the one implemented in the chemistry class where students were asked to take photos and make short films on the process of acidification. Once in the classroom, pupils were listening and discussing the information pupils provided through the films (containing children's own definitions of acidification process) and analysing the sources consulted and the accuracy of the content shared, with the purpose to find scientific indicators of acidification. (see Fig. 1).

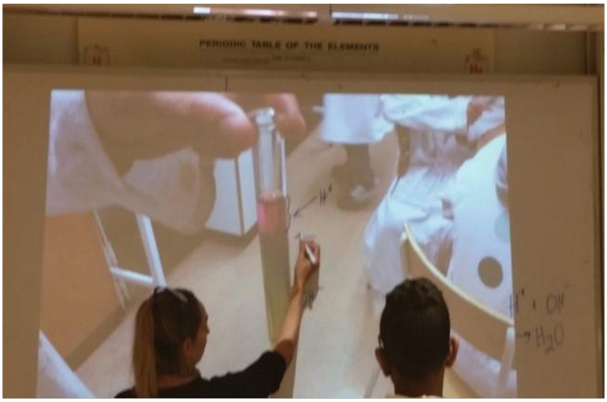


**Fig. 1.** Teachers and students analyzing results

Another category within the multimodal teaching and learning practice was language learning, facilitated by apps that were used to support pupils' pronunciation, vocabulary building, reading comprehension, spelling, and grammar in both Swedish and English. Most of these apps are indeed educational games. Several teachers valued particularly these educational games for children with other mother tongue than Swedish and for those diagnosed with dyslexia. According to the teachers a great majority of these games helped children to follow the teacher at almost the same pace than the rest of the class.

**Motivating Students' Engagement.** According to the teachers using tablets in the classroom seem to increase students' motivation and engagement. For instance, one teacher mentioned: *"We know that Ipads increases students' motivation, so when I feel that I need to increase students' engagement I let them work on the Ipads. That does not mean that I let them play games on the Ipads, they do serious work." (Sanna, Swedish and English teacher, Stockholm).* The teacher made reference to training language vocabulary and spelling activities given for instance on Fridays afternoon. At this particular time of the week, pupils are often tired and unfocused so the use of games for language training helps the class to engage with learning of vocabulary, spelling and pronunciation. In one opportunity, we observed how the entire classroom in the English class competed in groups performing tasks demanding vocabulary, pronunciation and spelling abilities. Children became extremely excited visualizing the scores and negotiating answers based on their language skills. In another opportunity we observed teachers in mathematics to use educational games in the classroom to train the class about arithmetic and geometry as well as identify which concepts needed to be reviewed; as the game makes available pupils' individual and group scores to teachers. As such games were used to get information on which pupils had more difficulties with a particular concept and test the overall class of most frequent errors. We also noticed, educational games were used to support collaborative learning through the resolution of for instance mathematical puzzles.

Multimodal educational games were also used in natural science. They often focused on multiple-choice questions and quizzes that pupils answered individually, or in dyads sharing a tablet. One of the teachers mentioned the value of these emergent practices in the classroom, stating: *"It is a fantastic activity. The students sometimes can't sit still and are jumping around because they are eager to know what the correct answers are."* Teachers in this study also mentioned they regarded educational games as a motivating tool challenging children to progress, visualize their own progression and work at their own pace, both in school and beyond.

**(Formative) Assessment and Provision of Feedback.** According to the teachers participating in our study, the integration of tablets in the classroom provided them with possibilities to systematically assess students and provide more accurate individual feedback on pupils' assignments. Teachers distinguished four categories of assessments and feedback practices, namely: (1) group assessment/feedback, (2) pupil assessment/feedback, (3) class assessment/feedback, and (4) automatic assessment/feedback. For instance, group work was assessed in different ways through the use of media tablets. One of the teachers of English let pupils video record their group conversations which later were assessed by the teacher who provided individual feedback to each group. That particular teacher emphasized the advantage of assessing video recordings of group discussions and conversations in the following way: *"By using video recordings I have documented what I base my assessment on. It allows me to forward and rewind. That is not possible when I observe a live discussion."* (Sanna, Swedish, English, Stockholm). The digital medium makes a difference for the teachers who can systematically save and retrieve, in this case group assignments and provide a more grounded feedback to the pupils. It is even the case teachers can show the recording to the pupils and engage a conversation on pupil's performance and teachers' assessment with the pupils. Another example was the one about assessing the whole class and providing feedback to the pupils that was facilitated by games application such as Kahoot. This application, we observed, enabled teachers to monitor both the progress of the class and each pupil individually and thus the possibility for teachers to provide feedback accordingly. The use of the digital material facilitated teachers to revisit pupils' performance and to adjust assessment and feedback provided, based on the evidence saved (i.e. audio files with pupil's conversations, pupil's reading of a text). In this case, the multimodality characteristic of the media tablets used for recording group or individual performance contributed to a more accurate assessment. The multimodality and persistence of the medium made of assessment and provision of feedback an evidence-based practice.

**E-mail Communication.** The analysis of the data showed that teachers do not really use tablets for communicating with the pupils through for instance e-mail. (M = 2.44, SD = 1.76). Teachers communicate with pupils face-to-face or mainly through instructions, assignments or feedback teachers provide on pupils' tasks available in the LMS.

## 3.2 Relation Between Identified Practices, Teachers' Perception of the Value of the Tablets in the Classroom and Transformation of Established School Practices

The results issued from the quantitative analysis revealed that teachers in general valued positively tablets in their teaching (M = 6.23, SD = 1.45) as they underscored tablets integrated in everyday pedagogical activities facilitate pupils to engage with school tasks and assignments (M = 6.14, SD = 0.36), motivate pupils (M = 6.31, SD = 1.52), and improve pupils' school performance (M = 5.72, SD = 1.93). The quantitative analysis also indicates that some teachers agree with the statement that tablets transform established teaching practices at school (M = 4.36, SD = 1.86). In order to examine teachers' appreciations in detail, a step-wise multiple regression analysis was used. The purpose of such analysis was to explain the variance in teachers' perceptions vis-à-vis the educational benefit of using tablets. In total, 88.3 % of the variance could be explained by a linear combination of the following variables: motivation & engagement, assessment & feedback, years of private tablet use (teachers), and multimodal learning and teaching (see Table 4).

**Table 4.** Step-wise regression analysis

| Variable | β | t | p |
|---|---|---|---|
| Motivation & engagement | .45 | 6.79 | 0.01 |
| Assessment & Feedback | .26 | 2.40 | 0.02 |
| Years of private tablet use | .28 | 3.19 | 0.06 |
| Multimodal learning | .41 | 5.14 | 0.01 |
| F(df, 6, 24) = 28.61<br>$R^2$ = 0.915 | | | |
| Adjusted $R^2$ = 0.883 | | | |
| Standard error = 0.455 | | | |

These results show that the more teachers perceived tablets increase -pupils' motivation, -support assessment & feedback, as well as -multimodal teaching and learning, the more they had used tablets for private use and the more they were to perceive educational benefits of tablets in their teaching. As one can notice, the strongest predictors were using tablets for stimulating student motivation and engagement, and using tablets for multimodal learning and teaching. Furthermore, independent sample t-tests were performed to investigate possible differences between teachers who perceived that tablets improve students learning and those who did not put in relation to how they valued the thematic practice. Significant differences were found only to regards to two themes of practices, namely: multimodal teaching and learning and motivation and engagement. Teachers who perceived the use of tablets improve student learning ($M = 5.20$, $SD = 1.00$) valued tablet-mediated multimodal teaching and learning practices significantly more than teachers who did not ($M = 3.72$, $SD = 1.06$), $t(28) = 3.13$, $p < 0.05$. Teachers who perceived that the use of tablets improve students' performance ($M = 4.13$, $SD = 1.17$) also valued the benefits of using tablets for increasing motivation and engagement significantly more than teachers who did not ($M = 2.34$, $SD = 0.98$), $t(28) = 3.99$, $p < 0.01$.

# 4 Discussion

Looking at the results obtained on the tablet-mediated practices identified and how teachers value the tablet in such practices, one can wonder why teachers have such a positive view on the use of tablets in the classrooms? Also, why are they pointed at these specific tablet-mediated teaching practices? We discuss the following questions elaborating on the three following points namely: teachers' interest and enthusiasm in introducing a new artefact into the ecology of the classroom, associating digitalization with educational progress as well as materiality of teaching practices.

## 4.1 Teachers' Interest and Enthusiasm

The teachers who agree to participate in the study were teachers who were highly motivated to introduce digital tools into their school and teaching practices. The teachers we interviewed and observed were passionate by introducing change into their workplace and accepted with much enthusiasm to participate in the study because they had something to show us. Furthermore, these teachers have invested much time and efforts in organizing hands-on workshops, pedagogical seminars and meetings so teaching staff could share experiences, knowledge and skills in relation to the use of media tablets in classrooms activities.

## 4.2 Associating Digitalization with Educational Progress

The teachers seem to associate digitalization of school practices to modernization and educational progress. Many of them mentioned more than once the imperative of adopting a tool that has become central in the children's everyday life (i.e. smart telephone or tablet). They conveyed a sense of "duty" of teaching children with and through tools that are part of children's worlds and Swedish society. Teachers also recognized media tablets introduce tensions into the classroom (i.e. entertainments games and social media platforms that are call "toys" by the teachers), but these were played down in their discourse.

### 4.3 Materiality of Teaching Practices

Results obtained about the types of tablet-mediated teaching practices identified as well as teachers' positive view of using tablets in the classroom, we contend, are associated to three main affordances tablets bring into the school: persistence of the digital medium, multimodality of the content and portability-ubiquity.

For instance, digital persistence [16] was referred to several times by teachers when explaining about the "mess" that distributing A 4 papers creates among pupils. Almost all the teachers agreed that handling school material electronically was a more effective way to save, search and centralize material. The persistence of the medium is not a detail when reflecting on how teachers value practices such as organization of teaching and documentation with tablets. For example, they can address pupils who often forget or loose teachers' homework, instructions or other type of information important to be aware of, to a shared and persistent workspace where all this material is centralized, accessible, organized and searchable. In that respect, the tablet via the use of LMS participates in building a socio-technical infrastructure [17] that organizes classroom work and facilitates face-to-face interaction and communication with the children.

Multimodality through the integration of sound, image to the written text is another affordance that is tightly connected to the emergence of multimodal teaching and learning practice [18]. This practice in particular was related to motivating pupils' to engage with school assignments and facilitating pupils' to construct learning material – instead of consume it. The entry of the image and sound into the classroom, through for instance Internet, it was also mentioned a reason for teaching pupils how to engage with emotional content in more critical ways. Furthermore, multimodality introduced an evidence-based assessment teachers found more reliable accurate and easier to share with pupils and eventually with interested parents. Also, the quantitative analysis performed indicated that a high valuation of multimodal teaching and learning practices by teachers was a significant predictor for perceived educational benefit of using tablets. Thus, this particular material affordance seems to be central for teachers and to a large extent explain their interest and enthusiasm in using tablets daily in their classrooms.

Portability and ubiquity of the tablet were associated with mobile learning, a practice that spoke of how teachers could organize teaching in different rooms and concentrate on the weakest groups. The fact that pupils can bring the tablets home facilitate for those who can not assist to the school or have troubles following tasks and assignments during classroom time.

Finally, we see an intricate relation between media tablets' affordances [6] and emergent tablet-mediated teaching practices [5]. Such a relationship, needs further examination as it underscores the value of design of digital devices in configuring today's school practices [9, 12, 16].

These devices once adopted in the classroom, influence, via their specific material characteristics current practices that oriented toward the construction of school knowledge [12, 19]. A thoroughly understanding of tablets' affordances through the analysis of teaching practices at schools will thus help researchers and designers in the TEL field better cognize how digital tools are transforming contemporary forms of learning. We thus believe research studies on the materiality of tablet-mediated practices are most than welcome at this stage, as Nordic schools for the most part, have entered the complex process of digitalization.

**Acknowledgements.** This work has been funded by the project Places with a research grant provided by the Swedish Research Council, Educational Sciences Program.

## References

1. Cerratto-Pargman, T., Milrad, M.: Beyond innovation in mobile learning:towards sustainability in schools. In: Traxler, J., Kukulska-Hulme, A. (eds.) Mobile Learning: The Next Generations, pp. 154–178. Routledge, London (2016)
2. Nouri, J., Cerratto-Pargman, T.: Characterizing learning mediated by mobile technologies: a cultural-historical activity theoretical analysis. IEEE Trans. Learn. Technol. **8**(4), 357–366 (2015)
3. Roschelle, J., Penuel, W.R., Yarnall, L., Shechtman, N., Tatar, D.: Handheld tools that "informate" assessment of student learning in science. J. Comput. Assist. Learn. **21**(3), 190–203 (2005)
4. Chou, C.C., Block, L., Jesness, R.: A case study of mobile learning pilot project in K-12 schools. J. Educ. Technol. Dev. Exch. **5**(2), 11–26 (2012)
5. Jahnke, I., Cerratto-Pargman, T., Furberg, A., Järvelä, S., Wasson, B.: Changing teaching and learning practices in schools with tablet mediated collaborative learning: nordic. In: European and International Views. CSCL2015, pp. 889–893. Gothenburg, Sweden (2015)
6. Kaptelinin, V.: Affordances and design. The interaction design foundation (2014)
7. Johri, A., Olds, B.M.: Situated engineering learning: Bridging engineering education research and the learning sciences. J. Eng. Educ. **100**(1), 151–185 (2011)
8. Jahnke, I., Norqvist, L., Olsson, A.: Digital didactical designs of learning expeditions. In: Rensing, C., de Freitas, S., Ley, T., Muñoz-Merino, P.J. (eds.) EC-TEL 2014. LNCS, vol. 8719, pp. 165–178. Springer, Heidelberg (2014)
9. Cerratto-Pargman, T., Knutsson, O., Karlström, P.: Materiality of online students' peer-review activities in higher education. In: Proceedings of CSCL 2015, pp. 308–315. Gothenburg, (2015)
10. Nouri, J., Cerratto Pargman, T., Eliasson, J., Ramberg, R.: Exploring the challenges of supporting collaborative mobile learning. Int. J. Mob. Blended Learn. **3**(4), 54–69 (2011)
11. Kaptelinin, Nardi: Affordances in HCI: toward a mediated action perspective. In: Proceedings of CHI 2012. Texas, USA (2012)
12. Laurillard, D.: Teaching as a Design Science: Building Pedagogical Patterns for Learning and Technology. Routledge, NY (2012)
13. Fenwick, T., Edwards, R., Sawchuk, P.: Emerging approaches to educational research: Tracing the socio-material. Routledge, London (2011)
14. Norman, D.: Affordance, conventions, and design. Interactions **6**(3), 38–43 (1999)

15. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(2), 77–101 (2006). ISSN 1478-088
16. Erickson, T.: Persistant conversation: an introduction. J. Comput. Mediated Commun. **4**, 308–325 (1999)
17. Star, S.L.: The ethnography of infrastructure. Am. Behav. Sci. **43**(3), 377–391 (1999)
18. Selander, N.S.: Conceptualization of multimodal and distributed designs for learning. In: The Future of Ubiquitous Learning, pp. 97–113. Springer, Heidelberg (2016)
19. Sörensen, E.: The Materiality of Learning: Technology and Knowledge in Educational Practice. Cambridge University Press, New York (2009)

# A Peer Evaluation Tool of Learning Designs

Kyparisia A. Papanikolaou[1(✉)], Evagellia Gouli[1], Katerina Makrh[1], Ioannis Sofos[2], and Maria Tzelepi[3]

[1] School of Pedagogical and Technological Education, Athens, Greece
kpapanikolaou@aspete.gr, lilag@di.uoa.gr, kmakrh@ppp.uoa.gr

[2] National Technical University, Athens, Greece
gs.sofos@gmail.com

[3] National and Kapodistrian University of Athens, Athens, Greece
tzelepimaria@yahoo.com

**Abstract.** In this paper, we focus on the value of assessment of students' artifacts as a means of evaluating and at the same time cultivating the various types of knowledge proposed by Technological Pedagogical Content Knowledge (TPACK) in a teacher education course on teaching with technology. In particular, we present the PeerLAND (Peer Assessment of Learning Designs) environment, which enables users to author technology-enhanced lessons (learning designs) and participate in peer assessment activities. Its design rationale is based on a learning designs' evaluation framework inspired by TPACK; Its function is twofold: (a) a pedagogical assessment and peer assessment mechanism and (b) a research tool as a lens, highlighting subtle perspectives of learning designs related to the complex, synthetic fields of knowledge needed by teachers in order to teach with technology. Initial evidence for the effectiveness of the process adopted in PeerLAND in cultivating learning design skills is provided.

**Keywords:** Evaluation methods for TEL · Learning design · Peer assessment

## 1 Introduction

Literature in the field of teacher education suggests that teachers' preparation in integrating digital technology in their teaching practice should support knowledge construction as regards their subject domain, pedagogical practices and technology, as well as the interrelation among these. This is the core idea of the Technological Pedagogical Content Knowledge model, widely known as TPACK [1], which theoretically ascribes the knowledge needed by teachers to teach with digital technologies. The field, however, is open to further development on ways to cultivate this knowledge [2–4] and also, to invent suitable means to evaluate it [5–7].

Research suggests two main approaches to evaluate the various types of knowledge included in TPACK which are based either on self-assessment of teachers participating in training programs, drawn through questionnaires and interviews, or on more objective measures such as observation and evaluation of their performance through their works and productions [5, 6]. Following the first direction, several self-reference


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 193–206, 2016.
DOI: 10.1007/978-3-319-45153-4_15

measurement scales have been suggested, an example being Schmidt et al. questionnaire [7], or open-ended questionnaire instruments. An interesting proposal towards this direction is a framework synthesizing TPACK with Activity Theory in order to examine teachers' activity in real educational/working conditions through interviews that capture self- evaluation based on a multi-factor lens [8].

Following the second direction, research is more scarce and focuses (a) on observation organized, coded and analyzed on the basis of the TPACK framework, (b) on problem solving by teachers on the basis of specific educational scenarios, (c) on analysis of teachers' artifacts, these being either lesson plans, portfolios or reflective journals. However, in the aforementioned research, assessment is carried out by teachers (trainees) themselves or by experts and the focus is often on specific TPACK fields, without explicitly addressing the skills and competencies expected to be cultivated by participants. Moreover, learning design is compartmentalized and dealt with as a separate domain from that of evaluation, thus undermining the continuum and integrative character of the two processes. This perspective is also supported by several learning design tools that focus on specific aspects of the process such as Learning Designer and Cloudworks. The Learning Designer (LD) [9] allows users to upload existing learning scenarios, or lesson plans, or create new ones, it analyses them and helps teachers recognize how much content or how many activities in their learning designs are dedicated to particular pedagogic practices, such as acquisition, reflection, practice, collaboration and production. Cloudworks is a social networking site in which clouds and bundles of clouds (cloudscapes) are used as valuable mediating artefacts to help guide discussion and sharing of learning and teaching ideas [10]. Although product evaluation is an issue in the aforementioned tools, in LD evaluation is approached as a self refection process focusing on particular pedagogic practices whilst in Cloudworks as a social activity with the main focus on the use of technology to support learning and teaching activities.

Aiming to contribute to the areas of teacher education and learning design, in this paper we present the online environment PeerLAND (Peer Assessment of LeArNing Designs) which supports the development and peer evaluation of learning designs on the basis of TPACK. A basic actor in the evaluation process is, in this case, the peer teacher trainees. They participate in a peer review process that is considered as an integral part of training in learning design with digital technologies. The paper concludes with elements of PeerLAND evaluation by postgraduate students in the context of a course on the use of digital tools for distance learning.

## 2 PeerLAND Overview

The scope of the web-based environment PeerLAND is students' support in creating and evaluating learning designs enhanced with digital technologies. There are two types of users of PeerLAND: (a) those that can be both authors of designs belonging to specific groups/classes and reviewers that have the right to submit reviews on designs of their group, and (b) instructors that may additionally formulate groups of students and move from one group to another.

As far as the design process is concerned, it demands the integration of *content knowledge, pedagogy and technology*. In particular, an educational scenario as a paradigm of a learning design, is a flexible, ill-structured plan of learning activities supporting students to achieve specific learning goals and work with the concepts involved. It also refers to ways to use tools, to organize individual/collaborative tasks and social orchestration, as well as the structure, sequence and content of tasks, place and time settings of the learning context [11]. Especially, designs align with constructivist approaches focusing on students' active involvement with digital technologies and their effective support towards specific learning goals through appropriate technological and pedagogical tools. Below we provide an overview of PeerLAND functionality for authors and reviewers.

**Author Environment.** Users are supported in the representation of educational scenarios. In particular, the structure of the learning designs hosted in PeerLAND is organised in multiple levels such as scenario, concept, activity (see Figs. 1 and 2), as pedagogical and technological aspects need to be clearly presented. Authors are supported to first articulate concise learning outcomes and then choose techniques, tools & resources, types of activities to include, as well as knowledge processes that students should cultivate in order to reach the expected outcomes. To this end, a list of constructivist techniques and technological tools are proposed, enabling authors to select the most appropriate for their scenario and argue about them. They can also check the compatibility of selected knowledge processes and activity types with their design. This way, authors are supported to build a pedagogically sound design in order to develop, evaluate or extend their scenario in an e-learning platform.

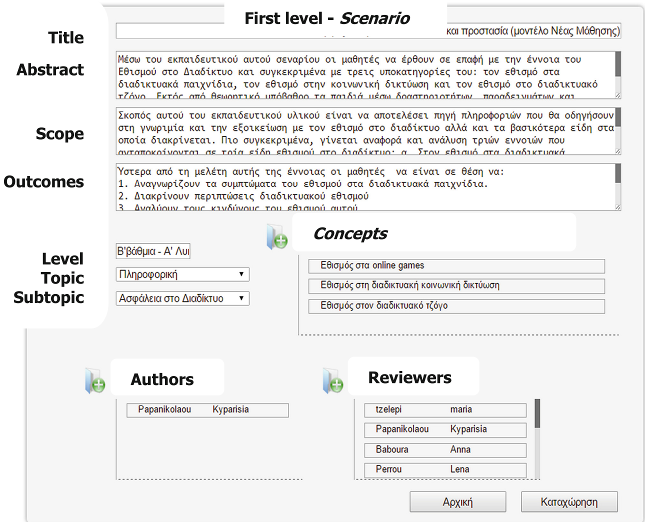



**Fig. 1.** First-level form of a scenario: *basic elements* such as abstract, scope, learning outcomes educational level, topic, sub-topic, *concepts involved, authors and reviewers.*

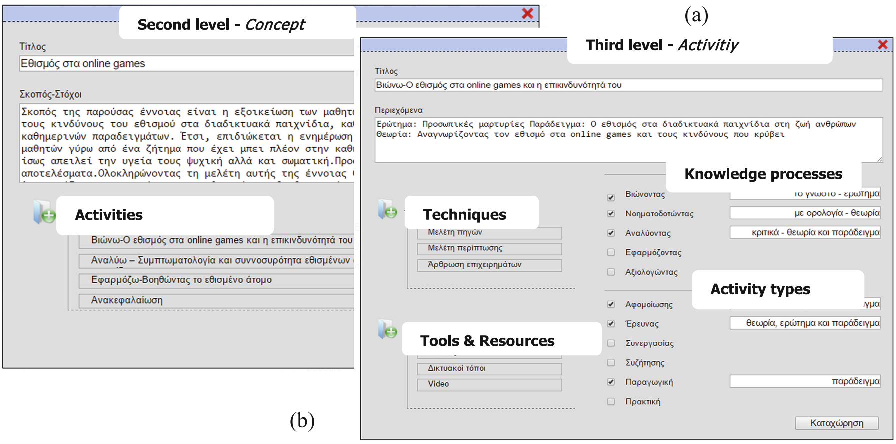



**Fig. 2.** (a) Second-level Form at concept level: *scope and outcomes, activity structure*, (b) Third-level Form at activity level: *didactic techniques, tools and resources, knowledge processes and activity types.*

This level of formality, apart from the articulation of basic design blueprints, also allows the “translation” of these blueprints into meaningful course structures.

**Peer-review Environment.** Users are supported to act as reviewers and participate in peer-evaluation tasks on scenarios authored by specific users-authors. The instructor can also act as a reviewer.

Reviewers evaluate scenarios using the TPACK [1] framework for thinking about what knowledge the authors have developed on the integration of technology into teaching. TPACK acknowledges three interdependent components of teachers’ knowledge, namely technological knowledge (TK), pedagogical knowledge (PK), and content knowledge (CK), as well as their intersections reflecting teachers’ understanding of teaching content with appropriate pedagogical methods and technologies such as pedagogical content knowledge (PCK), technological content knowledge (TCK), technological pedagogical knowledge (TPK). All these types of knowledge form the technological pedagogical content knowledge (TPACK).

The authors’ knowledge is estimated by each reviewer through the artifact they have developed, i.e. an educational scenario, and submitted in PeerLAND. The reviewers can check both the representation of the scenario in PeerLAND as well as the real scenario developed in an e-learning platform such as Moodle. The evaluation framework adopted is presented in the next section posing criteria per type of knowledge (as these appear in Fig. 3, forms TPACK (1/3) and TPACK(2/3)). Reviewers can also submit qualitative comments on the strengths and weaknesses of a scenario (see Fig. 3, TPACK(3/3)). The quantitative evaluation of the various aspects of the scenario is combined with the qualitative comments supporting authors to make revisions.

Numerical data make comparative presentation of reviewers’ comments possible (see Fig. 4). To this end PeerLAND builds a reference with numerical representations

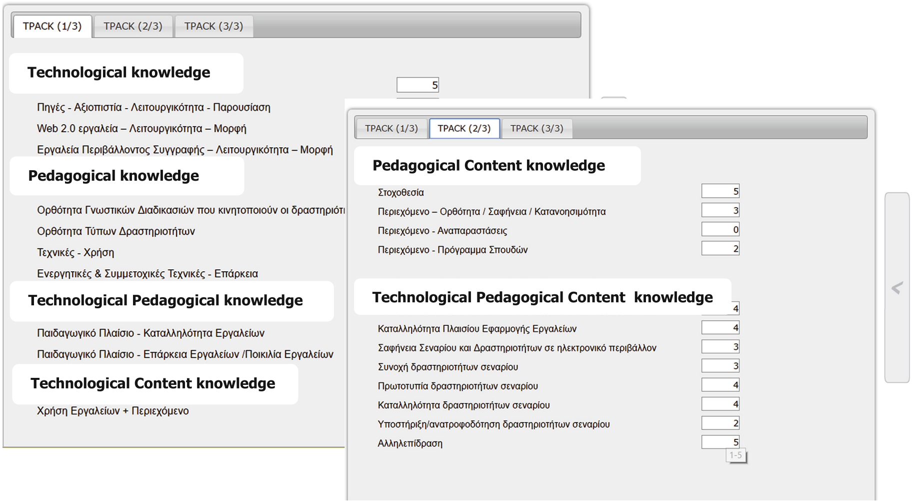


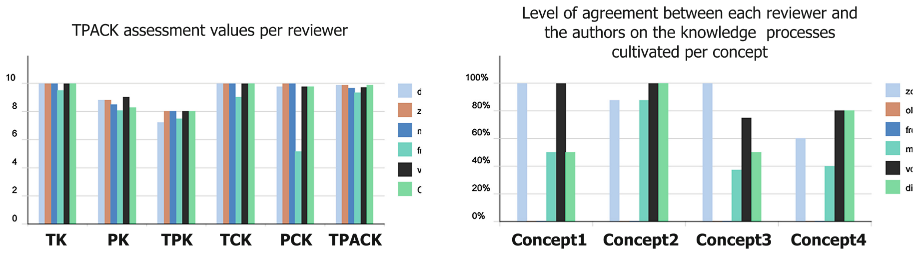

**Fig. 3.** Evaluation forms involving criteria about all the knowledge fields involved in TPACK organised in two Tabs, TPACK(1/3) and TPACK(2/3).

of the evaluations submitted per criterion and type of knowledge (see Fig. 4) as well as graphical representations of the level of agreement between each reviewer and the authors on the knowledge processes cultivated per concept (see Fig. 5). For example, as appears in Fig. 5, two of the reviewers agree with the authors on the first concept, whilst other two agree with half of the knowledge processes proposed.

**Technical Aspects.** PeerLAND is based on internet technologies aiming to enable users access it remotely without the need to install any particular software. To this end, client – server architecture was used. In particular, the server side layer has been developed in php programming language making use of MySql DBMS to store and manage available information, while an Apache server serves web content. The client side layer has been developed for web browser access following W3C standards. The application was developed in HTM, CSS, JavaScript and client – server communication is done with AJAX requests.

| | u_1 | u_2 | u_3 | u_4 | u_5 |
|---|---|---|---|---|---|
| **Technological knowledge** | 10.00 | 10.00 | 10.00 | 9.50 | 10.00 |
| **Pedagogical knowledge** | 8.80 | 8.80 | 8.50 | 8.05 | 9.00 |
| **Technological Pedagogical knowledge** | 7.20 | 8.00 | 8.00 | 7.50 | 8.00 |
| **Technological Content knowledge** | 10.00 | 10.00 | 10.00 | 9.00 | 10.00 |
| **Pedagogical Content knowledge** | 9.75 | 10.00 | 10.00 | 5.15 | 9.75 |
| **Technological Pedagogical Content knowledge** | 9.85 | 9.90 | 9.65 | 9.35 | 9.70 |

**Fig. 4.** Comparative presentations of evaluation data of a group of reviewers, i.e. u_1, u_2, u_3, etc., per knowledge dimension of TPACK




(a) (b)

**Fig. 5.** Graphical representations of evaluation data of a scenario from six reviewers according to (a) the knowledge fields of TPACK, (b) the degree of agreement of each reviewer with the authors concerning knowledge processes cultivated at concept-level. Names of reviewers have been eliminated.

## 3 Framework of Evaluation

In the PeerLAND environment, TPACK is used as a framework for cultivating, understanding and measuring teachers' knowledge on the development of educational scenarios enriched with digital technology. The evaluation framework proposes specific criteria for each knowledge dimension of the TPACK framework related to the learning design process. These criteria extend the factors proposed by [12] and adapt the instrument proposed in [7] in order to address the technological and pedagogical principles of the teacher training context proposed in [13]. The aim is to use a simple but also complete enough mechanism for teachers in order to describe their designs and evaluate them. In the first case, users as authors are enabled to describe the structure of the scenario and to undertake specific pedagogical and technological decisions during the design of activities, among a variety of alternatives. This process aims at cultivating learning design skills throughout the development process of an educational scenario. In the second case, the users as reviewers are enabled to reflect on educational scenarios of their peers, assess the use and understanding of TPACK by numerically rating (in most cases) the educational scenario according to specific criteria regarding technology and teaching. This process aims at promoting sharing and reflection on pedagogical and technological decisions taken through the design process. Peer evaluation tasks can thus enable self-evaluation through comparing what has been done in a scenario to what has been assessed in another scenario.

In Table 1 the criteria proposed per type of knowledge according to TPACK are presented. Along with each criterion the question that the reviewer has to answer in order to rate it, appears. In the case of $X_1$ and $X_2$ criteria of the Pedagogical Knowledge, their value is estimated automatically based on the comparison of the authors' and the reviewer's perspectives reflecting the level of agreement between the two. In particular, the ratings of the Pedagogical knowledge criteria 'Correctness of Knowledge Processes cultivated by the activities' (see Table 1 - Pedagogical knowledge – $X_1$) and 'Correctness of Types of Activities' (see Table 1 - Pedagogical Knowledge – $X_2$) are calculated as the level of agreement between the values posed by the authors and the reviewer of the scenario about the knowledge processes cultivated by each activity according to the New Learning framework [14] and the type of activity based on the categorization proposed by Laurillard [11]. The remaining criteria are numerically assessed in a range from 1 to 5 as this range is also adopted in the TPACK instrument [7]. The value of each criterion is associated with a weight ($w_i$) that can be altered by the instructor reflecting the current context and its priorities on the learning design process.

**Table 1.** Evaluation Criteria based on TPACK: *How do I understand that the authors of a particular scenario have developed specific skills/abilities?*

| **Technological Knowledge (TK)** |
|---|
| $TK = w_1 * X_1 + w_2 * X_2 + w_3 * X_3 + w_4 * X_4$ ***where*** |
| $X_1$: **Tools – Functionality – Form:** *Are the learning objects of the scenario created with specific tools (e.g. standalone software, web-based software that cannot be integrated in sites) functional? Is their form/presentation adequate?* |
| $X_2$: **Resources - Credibility - Functionality - Presentation:** *Are the links proposed to web resources active? Are they credible and valid? Is their reference complete and appropriate (e.g. based on an international standard like APA style providing data of last visit, followed by a short description etc.)?* |
| $X_3$: **Web 2.0 tools – Functionality – Form:** *Are the learning objects of the scenario created with specific Web 2.0 tools (e.g. glogster, timeline, word cloud, video etc.) functional? Is their form/presentation adequate?* |
| $X_4$: **Authoring Environment Tools – Functionality – Form:** *Have the tools of the authoring environment been appropriately incorporated in the activities of the scenario?* |
| **Pedagogical Knowledge (PK)** |
| $PK = w_1 * X_1 + w_2 * X_2 + w_3 * X_3$ ***where*** |
| $X_1$: **Correctness of Knowledge Processes cultivated by the activities**: *Do the activities of the scenario cultivate the knowledge processes referred by the authors?* |
| $X_2$: **Correctness of Activities ' Type** : *Do the activities of the scenario offer to students the learning experience described by the type of each activity as it this is characterized by the authors?* |
| $X_3$**: Techniques -** $X_3 = w_{31} * X_{31} + w_{32} * X_{32}$ *where* |
| $X_{31}$: **Didactic Techniques - Use**: *Is activities' context appropriately and adequately described?* |
| $X_{32}$: **Active & participatory Techniques - Adequacy**: *Do activities of the scenario promote active involvement and students' interaction?* |
| **Content Knowledge (CK)** |
| We assume that the students know well their 'content' e.g. subject matter that they will teach. |
| **Pedagogical Content Knowledge (PCK)** |
| $PCK = w_1 * X_1 + w_2 * X_2 + w_3 * X_3 + w_4 * X_4$ ***where*** |
| $X_1$: **Learning outcomes**: *Do the aims of the scenario and/or (overall) the activities cover the suggested knowledge processes and activity types? Are the learning outcomes attainable through the* |

**Table 1.** *(Continued)*

*means/context described in the scenario?*

$X_2$: **Content – Correctness / Accuracy / Understandability:** *Is the content and context of the activities scientifically correct, accurate and understandable? Are there accurate guidelines for students on how to use the content provided; Are there clear explanations of difficult concepts;*

$X_3$: **Content - Representations:** *Is the content of the activities characterized by multiformity? Does the content include various representations such as images, charts, tables, video, simulations, lists, comments? Do the learner-centered teaching techniques/methods adopted such as inquiry, problem solving,concept mapping, discussion, collaboration, promote the multiformity of the content? Do these representations and instructional strategies evoke students' interest? Are they suitable for the knowledge process they support? Are they appropriate in the activity context they are embedded?*

$X_4$: **Content - Curriculum** (if necessary): *Does the content of activities follow the proposed schedule and the curriculum;*

**Technological Pedagogical Knowledge (TPK)**

**$TPK = w_1 * X_1 + w_2 * X_2$ *where***

$X_1$: **Pedagogical Context - Tools Appropriateness**: *On the basis of the given pedagogical framework (knowledge processes, activity types, teaching techniques), are the proposed technological tools (tools, resources, Web 2.0 tools, "Other" tools) considered as appropriate, supporting the aims of knowledge processes, the activity types and teaching techniques within which they are embedded / integrated? Do they appropriately address the audience they target?*

$X_2$: **Pedagogical Context - Tools Adequacy/Tools Variety**: *On the basis of the given pedagogical framework, are the technological tools suggested considered as adequate to support the aims of the knowledge processes, activity types and instructional techniques within which they are embedded? Is the number/set of proposed tools characterised by variety?*

**Technological Content Knowledge (TCK)**

**$TCK = X_1$ *where***

$X_1$: **Tools Use + Content**: *Do the technological tools involved in the scenario add to the various activities multiformity, variety and alternative representations of information?*

**Technological Pedagogical Content Knowledge (TPACK)**

**$TPACK = w_1 * X_1 + w_2 * X_2 + w_3 * X_3 + w_4 * X_4 + w_5 * X_5 + w_6 * X_6 + w_7 * X_7 + w_8 * X_8$ *where***

$X_1$: **Appropriateness of technological tools' integration based on their potential**: *Are the potential and functionalities of the proposed technological tools adequately integrated in the scenario/activities? In case of the use of an e-platform for the implementation of the scenario, are the variety of its tools adequately used in the activities context?*

$X_2$: **Appropriateness of Learning Context for the technological tools**: *Have the technological tools been appropriately integrated in order to serve the learning outcomes of the scenario/activities and the pedagogical framework?*

$X_3$: **Accuracy of Scenario and Activity representation in the e-platform:** *Are there adequate guidelines on the time schedule of the various activities, on students' prior and prerequisite knowledge, on the importance and outcomes of the scenario?*

$X_4$: **Activity coherence:** *Is the activity sequence coherent as it is represented in the e-platform?*

$X_5$: **Originality of activities:** *Does the integration of technology within the scenario/activities take*

**Table 1.** *(Continued)*

| |
|---|
| *place in an innovative way, so that students' creativity is encouraged?* |
| $X_6$: **Activity appropriateness:** *In the activities implemented in the e-platform, is the wording of the assignments simple, with concise guidelines and friendly to the specific audience they address? Are there graphics, shapes, tables, footnotes where needed, adequate graphical annotations? Is the aesthetic presentation of activities considered as appropriate?* |
| $X_7$: **Support/feedback offered:** *Is the support provided to students (tools, media, guidelines, organisation of in teraction, communication, point out of difficult concepts, system feedback, helpful observations, e.t.c) throughout the scenario considered as appropriate ? Are the students enabled to achieve the aims of the activities?* |
| $X_8$: **Interaction:** *Is the interaction among students enhanced through the particular activities of the scenario? Is the collaboration/interaction of students (e.g. adequate guidelines, roles & collaboration script/argumentation, use of appropriate tools etc.) well organised in order to succeed? Is the cultivation of collaborative skills appropriately supported?* |

## 4 Empirical Evaluation of PeerLAND

The PeerLAND environment was used by 13 students in a postgraduate course on technology enhanced distance education of the National and Kapodistrian University of Athens during the academic year 2015–16. The students came from different disciplinary areas, such as primary education, foreign language education, literature and linguistics, geology and other technical fields.

In the particular course, emphasis was put on the role of the student as a designer of learning activities and educational scenarios enriched with digital technology. The aim was to design an environment that sought to support learning design activities with respect to the technology integration to the educational process by teachers [15]. Throughout the course, students are involved in individual and group activities. Given the fact that this is a non-homogenous class, group work is organized around collaboration of interdisciplinary teams towards developing a learning design in the form of an educational scenario.

The various fields of TPACK knowledge were cultivated in an integrative way, through activities offering opportunities for synthesizing subject matter knowledge, its pedagogy and technology [13]. The digital technologies used during the course are distinguished in three main types [16]: (a) virtual classroom environment that enables real class processes, (b) technologies being themselves a learning focus (Web 2.0 tools for developing learning objects, web-based resources) and (c) technologies as means to implement learning designs such as learning design and course authoring tools.

Students initially worked as students in Moodle, where the virtual class of the course was set up, and in INSPIRE*us* [17], in order to acquaint themselves with the basic functions of elearning systems and also with personalisation. Next they worked in groups, as authors of learning materials towards the goal of designing and implementing educational scenarios with potential for personalized support. For this purpose, they initially used the environment of Learning Designer [9] in order to design the educational scenario, as well as Moodle, where they were involved in a peer review activity for the scenarios they had just designed. Then they developed their scenarios in Moodle. There followed a second round of peer evaluation during which the authors transferred their scenarios in the PeerLAND environment and their reviewers (being the same as those in the first phase) deposited their evaluations. Each group received an evaluation report from the system synthesizing all the reviews' submitted about their scenario. Finally, students filled in evaluation questionnaires about their learning experience with PeerLAND, in order to evaluate its potential both as authors and as reviewers.

## 4.1 Methodology

The evaluation of PeerLAND was based on a questionnaire, constructed and tailored for the needs of this research. The questionnaire is organized in two parts with the aim of evaluating the environment on the basis of both the authoring process of learning designs and the reviewing process. With regards to the authors' view, students were assigned to assess the contribution of the environment in representing and designing educational scenarios, on a 5-level Likert scale from (1) – totally agree to (5) – totally disagree (see Table 2, Part A. Evaluating as authors - Questions 1–6). They also assessed the contribution of the environment in the scenarios' improvement as it enabled the provision of multiple forms of feedback by multiple reviewers on the basis of TPACK fields (see Table 2, Part A. Evaluating as authors - Questions 7-13). In addition, with regards to the reviewers' view, students were assigned to assess –on a similar Likert scale- the contribution of the environment on the facilitation of the evaluation process through the TPACK framework, the sharing and comparison with other reviewers and with the authors themselves (see Table 2, Part B. Evaluating as reviewers, Questions 14-22). The data analysis of the 13 questionnaires included adding up students' choices per question.

Figure 6 presents a standard deviation plot which represents the average (central point in the line) of the 13 students' answers per question (axis X), as well as the standard deviation (distance between the edges of the line) among the answers per question.

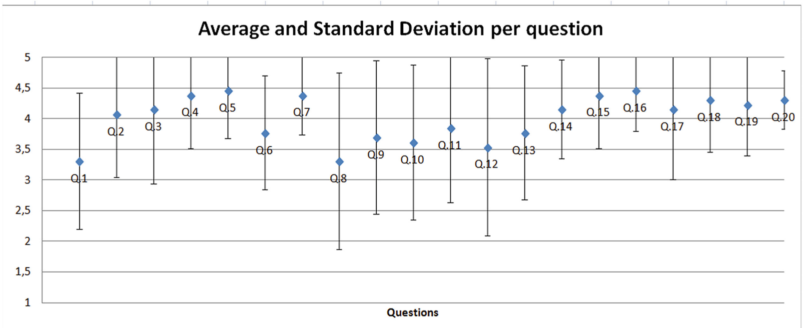



**Fig. 6.** Standard deviation plot of students' answers to the PeerLAND evaluation questionnaire

Students' answers move on average above value "3" in all questions. Especially encouraging are their answers in questions 4 and 5 which are about the support provided on the pedagogical aspect of the scenario and to questions 16 and 20, which refer to the potential of cross-comparison of evaluations among reviewers, where the average in answers surpasses value "4" and the variance is trivial. See also Table 2 where almost all answers to questions 4, 5, 16 and 20 appear positive. Thus, the integration of PeerLAND in a learning context should take place in a way that enables both reflection and the shaping of a learning design. As regards students' answers in questions 1, 8 and 12 in Fig. 6, where there is a relatively low average in students' answers and a high variation, the interpretation about questions 1 and 12 relates to the course context, whilst for question 8, the interpretation relates to the kind of data given to the authors (see also Table 2). More specifically, concerning the course context, students used PeerLAND after having completed the design of their scenario in Learning Designer and its implementation in Moodle (also having in the middle completed a peer evaluation process in Moodle). This resulted in using PeerLAND in order to represent a complete, ready structure–a fact which didn't function in support to its formation and shaping-. Students also didn't have the time to proceed to substantial changes. In an open discussion about their learning experience with PeerLAND, they stated that the use of the environment would be more constructive midway during the first stages of the scenario development in Moodle, in order for the comments to contribute to its improvement. Students' acknowledge the added value of PeerLAND in supporting peer evaluation and improvement both from the authors' and reviewers' perspective as this results also from their positive answers to questions 7 and 14 in Table 2. At this point emerges the need for purposeful integration of PeerLAND in an appropriate context, so that its dual function to best support both the authoring of learning scenarios, their evaluation and their subsequent improvements. As regards question 8 in Fig. 6, the detailed numerical representations of comparative data coming from authors and reviewers on every concept, seem not to have been positively evaluated by authors, data that would be really useful for the instructor of the course. This finding leads to a guideline on the re-examination of the adjustment of available information, depending on the role of each user within the environment (i.e. student-author, student-reviewer, instructor).

**Table 2.** Students' answers per choice and question

| **PART A. Evaluating as Authors** | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **1. PeerLAND helps me organize or confirm the structure of a scenario, as it gradually leads me to its construction.** | **1** | **1** | **6** | **3** | **2** |
| 2. The tools/functionalities offered by PeerLAND help me design activities, as they inform me of the teaching techniques that I can use. | 0 | 1 | 3 | 3 | 6 |
| 3. The tools/functionalities offered by PeerLAND help me design activities, as they inform me on the tools/resources that I can use. | 1 | 0 | 2 | 3 | 7 |
| **4. The tools/functionalities offered by PeerLAND help me design activities, as they inform me on pedagogically meaningful types of activities.** | **0** | **1** | **0** | **5** | **7** |
| **5. The tools/functionalities offered by PeerLAND help me design activities, as they** | **0** | **0** | **2** | **3** | **8** |

**Table 2.** *(Continued)*

| | | | | | |
|---|---|---|---|---|---|
| **inform me on the knowledge processes that can be supported by specific activities.** | | | | | |
| 6. PeerLAND offers a user-friendly and easy to use environment for authors of educational scenarios. | 0 | 2 | 1 | 8 | 2 |
| 7. I consider as important the potential offered by PeerLAND to validate or to improve my design (techniques, activity types, knowledge processes), through comparing my view to this of each reviewer per activity. | 0 | 0 | 1 | 6 | 6 |
| **8. The way reviews on my scenario are presented (in the form of a table with agreement percentages between my estimation and that of each reviewer) helped me to reflect on my design.** | **3** | **0** | **2** | **6** | **2** |
| 9. The comparative presentation of reviews on my scenario, per knowledge field and per criterion based on TPACK helped me to reflect on my design. | 2 | 0 | 0 | 9 | 2 |
| 10. The graphical presentation of reviews on my scenario based on TPACK, comparatively showing the evaluations of my reviewers helped me to reflect on my design. | 2 | 0 | 1 | 8 | 2 |
| 11. The comments I received on my scenario based on TPACK helped me to reflect on its positive and negative elements. | 1 | 1 | 1 | 6 | 4 |
| **12. The feedback I received through the evaluations of my scenario based on TPACK helped me to redesign it.** | **1** | **3** | **2** | **2** | **5** |
| 13. I consider as important the potential offered by PeerLAND to receive quantitative evaluations on the degree I developed each knowledge field according to TPACK. | 1 | 0 | 3 | 6 | 3 |
| **PART B. Evaluating as Reviewers** | | | | | |
| 14. The potential of peer reviewing and sharing scenarios offered by PeerLAND allows authors-reviewers to share their ideas. | 0 | 1 | 0 | 8 | 4 |
| 15. As a reviewer, I consider as important the potential offered by PeerLAND to export my evaluation as a .pdf file. | 0 | 1 | 0 | 5 | 7 |
| **16. As a reviewer, I consider as important the potential offered by PeerLAND to compare my evaluation to those of other peer reviewers.** | **0** | **0** | **1** | **5** | **7** |
| 17. As a reviewer, I consider as important the potential offered by PeerLAND to compare my evaluation to that of the authors (agreement percentage) about the design of each activity separately, but also with the respective evaluations of peer reviewers. | 1 | 0 | 1 | 5 | 6 |
| 18. As a reviewer, I consider as important the potential offered by PeerLAND to compare my evaluation to that of other peer reviewers through the specific assessment values per knowledge field and per TPACK criterion. | 0 | 1 | 0 | 6 | 6 |
| 19 As a reviewer, I consider as important the potential offered by PeerLAND to compare my overall evaluation to that of other peer reviewers though graphical representations based on TPACK. | 0 | 1 | 0 | 7 | 5 |
| **20. PeerLAND offers an accessible and user friendly working environment for reviewers of educational scenarios.** | **0** | **0** | **0** | **9** | **4** |

## 5 Conclusions

PeerLAND provides students the opportunity to work as authors with tools that support them in representing pedagogical and technological aspects of educational scenarios and then as reviewers assessing their peers' works according to specific criteria that reflect various perspectives of technology integration with pedagogy on the subject matter. This way training on ICT integration in teaching run on a continuum of design to evaluation. Initial evidence of the usefulness of PeerLAND gathered in a post-graduate course where students worked with PeerLAND undertaken the roles of authors and reviewers. Students used several learning design and content authoring environments throughout the course, also participating in peer assessment activities. This context matched the learning goals of a postgraduate course for pre-service teachers on technology enhanced learning. However, in a real-world environment with in-service teachers, the integration of environments, that was technically impossible at this phase of the research, would be an approach worthy to be pursued. In this line of research, we intend to further investigate how such a peer-feedback component may extend and support the process of course design and development in an e-learning platform.

As far as the PeerLAND evaluation is concerned, students acknowledged the support offered by PeerLAND in designing and improving their designs. They highlighted the value of the process of knowledge building on learning design. However, the context in which PeerLAND and the reviewing process is integrated is critical in order to support reflective cycles of design and improvement. Students argue that using PeerLAND at the first stages of the design process would be more effective for the final product. Our future plans include the collection of peer assessment results for the scenarios submitted. The analysis of this dataset and its cross-examination with results from self reports will provide more evidence on the value of peer evaluations through PeerLAND. Finally, another challenging research goal that we also consider is to extend the qualitative feedback in a way that could be comparable among different reviewers and the authors.

## References


1. Mishra, P., Koehler, M.J.: Technological pedagogical content knowledge: a framework for integrating technology in teacher knowledge. Teach. Coll. Rec. **108**(6), 1017–1054 (2006)
2. Jimoyiannis, A.: Designing and implementing an integrated technological pedagogical science knowledge framework for science teachers professional development. Comput. Educ. **55**(3), 1259–1269 (2010)
3. Tzavara, A., Komis, V.: Design and implementation of educational scenarios with the integration of TDCK: a case study at a department of early childhood education. In: Angeli, C., Valanides, N. (eds.) Technological Pedagogical Content Knowledge: Exploring, Developing, and Assessing TPCK. Springer, New York (2015)
4. Valtonen, T., Kukkonen, J., Wulff, A.: High school teachers' course designs and their professional knowledge of online teaching. Inf. Educ. **5**(2), 301–316 (2006)

5. Chai, C.-S., Koh, J.H.-L., Tsai, C.-C.: A review of technological pedagogical content knowledge. Educ. Technol. Soc. **16**(2), 31–51 (2013)
6. Koehler, M.J., Mishra, P., Kereluik, K., Shin, T.S., Graham, C.R.: The technological pedagogical content knowledge framework. In: Spector, J.M. et al. (eds.) Handbook of Research on Educational Communications and Technology, Springer Science + Business Media, New York (2014)
7. Schmidt, D.A., Baran, E., Thompson, A.D., Mishra, P., Koehler, M.J., Shin, T.S.: Technological pedagogical content knowledge (TPACK): the development and validation of an assessment instrument for preservice teachers. J. Res. Technol. Educ. **42**, 123–149 (2009)
8. Terpstra, M.: TPACKtivity: an activity-theory lens for examining TPACK development. In: Angeli, C., Valanides, N. (eds.) Technological Pedagogical Content Knowledge: Exploring, Developing, and Assessing TPCK. Springer, New York (2015)
9. Laurillard, D., Charlton, P., Craft, B., Dimakopoulos, D., Ljubojevic, D., Magoulas, G., Masterman, E., Pujadas, R., Whitley, E.A., Whittlestone, K.: A constructionist learning environment for teachers to model learning designs. J. Comput. Assist. Learn. **29**(1), 15–30 (2013)
10. Conole, G., Culver, J.: The design of cloudworks: applying social networking practice to foster the exchange of learning and teaching ideas and designs. Comput. Educ. **54**(3), 679–692 (2010)
11. Laurillard, D.: Teaching as a Design Science: Building Pedagogical Patterns for Learning and Technology. Routledge, New York (2012)
12. Oster-Levinz, A., Klieger, A.: Indicator for technological pedagogical content knowledge (TPACK) evaluation of online tasks. Turk. Online J. Distance Educ. **11**(4), 47–71 (2010)
13. Papanikolaou, K., Gouli, E., Makri, K.: Designing pre-service teacher training based on a combination of TPACK and communities of inquiry. In: Proceedings of 5th World Conference on Educational Sciences (WCES-2013) (2014). Procedia Soc. Behav. Sci. **116** 3437–3442
14. Kalantzis, M., Cope, B.: New Learning: Elements of a Science of Education, 2nd edn. Cambridge University Press, Cambridge (2012)
15. Garreta-Domingo, M., Hernández-Leo, D., Mor, Y., Sloep, P.: Teachers' perceptions about the HANDSON MOOC: A learning design studio case. In: Conole, G., et al. (eds.) EC-TEL 2015. LNCS, vol. 9307, pp. 420–427. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24258-3_34
16. Papanikolaou, K.A., Makri, K., Magoulas, G.D., Chinou, D., Georgalas, A., Roussos, P.: Synthesizing technological and pedagogical knowledge in learning design: a case study in teacher training on technology enhanced learning. Int. J. Digit. Literacy Digit. Competence **7** (1), 19–32 (2016)
17. Papanikolaou, K.: Constructing interpretative views of learners' interaction behavior in an open learner model. IEEE Trans. Learn. Technol. **8**(2), 201–214 (2015)

# Learning in the Context of ManuSkills: Attracting Youth to Manufacturing Through TEL

Stefano Perini[1(✉)], Maria Margoudi[2], Manuel Oliveira[3], and Marco Taisch[1]

[1] Department of Management, Economics and Industrial Engineering (DIG), Politecnico di Milano, Milan, Italy
{stefano.perini,marco.taisch}@polimi.it
[2] HighSkillz Ltd, Chatham, UK
maria.margoudi@highskillz.com
[3] Department of Industrial Management, SINTEF, Trondheim, Norway
manuel.oliveira@sintef.no

**Abstract.** The manufacturing industry plays a pivotal role in the European economy and global competitiveness. Although production technologies and processes are continuously being improved towards the vision of factories of the future, there is a dire shortage of human capital due to skills shortage and mismatch. This paper presents the results of studies carried out in Europe on how to leverage ICT (virtual reality, serious games, teaching factory and simulations) to increase the awareness and interest of young talent in manufacturing education.

A total of 24 field experiments was conducted across 5 European countries with a sample of 461 students of different age-groups, namely primary (children), secondary (teenagers) and tertiary education (young adults) with particular focus on secondary-school students. The analysis of the results are encouraging, demonstrating an impact achieved, in particular with regards to the youngsters with lowest levels of awareness and interest.

**Keywords:** Technology enhanced learning · Awareness · Interest · Manufacturing

## 1 Introduction

The manufacturing industry is a foundational pillar of the European economy, with 1 in every 10 companies employing 30 million people and generating 1,620 billion euros of added value[1]. The recognition of the strategic importance of manufacturing is embodied in the European Factories of the Future (FoF[2]) Public-Private partnership (PPP). The sector is a key driver for innovation, productivity and job creation, but

[1] http://ec.europa.eu/eurostat/statistics-explained/index.php/Manufacturing_statistics_-_NACE_Rev._2.
[2] http://ec.europa.eu/research/industrial_technologies/factories-of-the-future_en.html.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 207–220, 2016.
DOI: 10.1007/978-3-319-45153-4_16

despite the investment in research and innovation, industry suffers from a significant shortage of talent and skills mismatch. As documented by a survey of over 400 CEOs all over the world managed by Deloitte and the U.S. Council on Competitiveness [1], talent-driven innovation is nowadays the most important driver of global manufacturing competitiveness.

The dramatic importance of the human component for the prosperity of future manufacturing has been also highlighted by McKinsey & Company [2], which identifies the building of innovative workers' skills as one of the four key areas to focus on for the empowerment of manufacturing. For the second edition of the US Manufacturing Institute report [3], Deloitte points out that the problem of Skills Gap in Manufacturing has reached a boiling point with two thirds of US companies surveyed reporting a shortage of available qualified workers, leading to the destruction of 5 % of current US manufacturing jobs (600,000 jobs) due to a lack of qualified candidates. This shortage is not limited to the US: a similar survey organized by IDC [4] in May 2013 in Europe and North America, demonstrates that the same shortage appears in each developed country. For IDC, "People are the Opportunity and the Barrier: among the most critical barriers hampering the Factory of the Future strategy, manufacturers identified the challenge of finding skilled people with more than 70 % of share." Both symptoms and root causes of this skills gap appear to be similar in North America and in Europe. This presents an employment paradox [5], even in countries with high unemployment, there is an increasing number of employers reporting difficulty in filling manufacturing jobs [6]; the different functions within a single organization, with the engineering/technical ones among the most affected [7]; the educational attainments, with a more critical shortage of high-skill and medium-skill workers rather than low-skill ones [2].

The root causes of the skills shortage identified have been widely explored both in literature and in practice, leading to the identification of different elements, e.g. an aging workforce, an outdated strategic workforce planning, a limited efficiency of life-long learning, a poor perception of manufacturing among the young generation, the volatility and rapid transformation of work [8]. Therefore, there is a need to address the education and training to increase the supply of young talent to the European manufacturing industry, which implies also an increase of societal appeal of manufacturing to attract young talent. The use of innovative technologies in education and training play an important role to support the fast pace of change affecting the manufacturing industry, so for example new approaches for managing knowledge and developing skills is required so that the manufacturing decision making can be dispersed in the production level.

The purpose of the paper is to present the insights of the ManuSkills project, a European FP7 funded project, in leveraging innovative ICT technologies to attract young talent to manufacturing and increase their competences. A set of 24 field experiments were conducted across 5 European countries with a sample of 461 students of different age-groups, targeting primary (children), secondary (teenagers) and tertiary education (young adults). The experiments involved the use of serious games, virtual reality, teaching factory and simulation. The paper is divided into a further five distinct sections, starting with the distinction between awareness and interest (Sect. 2). The description of the ICT applications is provided in Sect. 3, whilst an overview of the evaluation approach of the field experiments and a discussion of the key results are presented in Sects. 4 and 5 respectively. Finally, conclusions are presented in Sect. 6.

## 2 Awareness and Interest Creation

The manufacturing skills gap is linked to the negative perception of manufacturing that youngsters hold and stems from wrong conceptions around its basic concepts. Indeed, as already pointed out, in different studies related to engineering field, youngsters usually don't have a clear perception of what advanced technical careers actually imply [9]. We argue that misconceptions about modern manufacturing can avoided by targeting two different notions that lead students to a more conscious and founded choice of their studies and profession: awareness and interest.

The concept of awareness has been deconstructed and thoroughly analyzed by different epistemological fields, such as psychology [10], marketing [11] and education [12], which is indicative of the significance and the cross-disciplinary nature of the term. The most widespread and commonly accepted definition of the term describes awareness as the ability to perceive, feel or be conscious of events, objects, thoughts, emotions or sensory patterns. In the educational field, when awareness gets associated to learning, it becomes equated with a person's ability to make forced-choice decisions above a chance level of performance [13]. For the purposes of the current study, the definition of awareness as the understanding that an individual has formed around a specific concept, as part of unconscious learning [14] was adopted. To summarize, the concept of awareness is strongly linked to consciousness, identified as the state or ability to perceive, feel or to be conscious of events, concepts or objects without necessarily proceeding to the level of understanding.

However, mere awareness of concepts is not sufficient enough to initiate a change of attitudes towards manufacturing. An important aspect that should be also taken into account is the creation of interest. Defined as the content-specific motivational characteristic composed of intrinsic feeling-related and value-related valences [15], interest is considered as something more than the passive awareness of a given domain, i.e. the active engagement and involvement of the youngster towards the presented concepts.

The difference and the link between the concepts of awareness and interest stems from the communication and marketing sector. The AIDA communication model [16] associates awareness with a first level of attention capturing. However, interest is the natural consecutive stage, where once the attention is captured actual interest building on a topic or domain can be initiated. In fact, it is worthwhile to notice that even though these two concepts are complementary, the one doesn't necessarily imply the presence of the other. Consequently, even though an individual might have high awareness levels on manufacturing, he/she might on the other hand have a scant interest and vice versa.

## 3 ICT Applications

ManuSkills ICT applications aim at targeting different age groups, hence pursuing for each of them the specific educational objectives of awareness and interest. The age groups identified are three, i.e. children (10–12 years old), teenagers (13–18 years old) and young adults (university students).

For children and teenagers, the assessment and improvement of the actual levels of awareness and interest about manufacturing turn out to be fundamental, trying to stimulate and then let them consider the possibility of a study and then a career in manufacturing field.

For young adults, awareness and interest consolidation and improvement are also definitely relevant, since youngsters are here still in an educational environment, and therefore even though their future career in manufacturing is probable is still not certain, since the attraction to other domains (e.g. consultancy, real estate, banking and finance) where manufacturing engineers could work should also be taken into account. Furthermore, not all university courses necessarily address only manufacturing, but often propose an interdisciplinary offer where manufacturing is one of the main components (e.g. engineering management, technology management degrees).

The six ICT applications support the field experiments are captured in Fig. 1, each of which will be summarily described. More information about the ICT applications can be found at http://demo.manuskills.org/.

**BrickPlanner** ***(age group: teenagers and young adults).*** The BrickPlanner is a serious game where the student is given a million euros to build a toy manufacturing

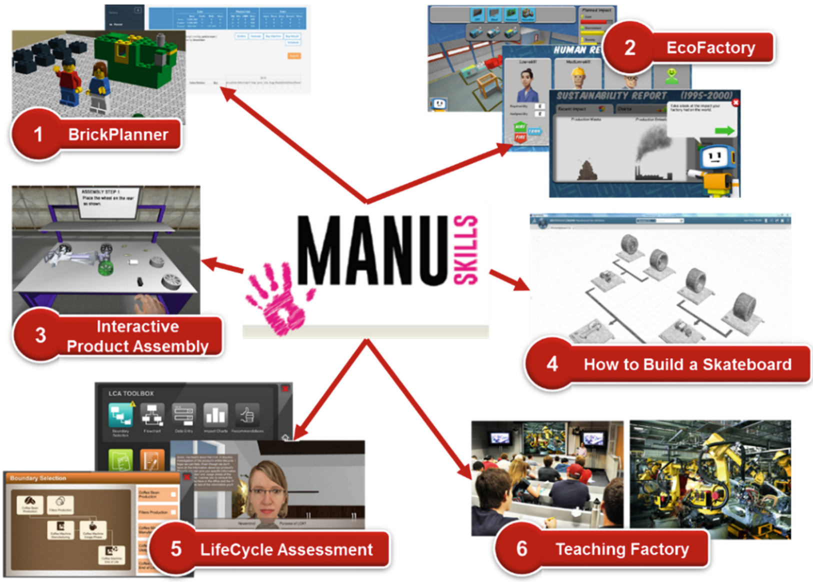



**Fig. 1.** Overview of the six ICT applications used in ManuSkills

company that is successful. In the process, they are given ten challenges, starting with a simple moulding machine and a production order. Gradually, the student builds a manufacturing company and addresses the complexity of dealing with multiple orders, including rush orders.

**EcoFactory** ***(age group: children and teenagers).*** EcoFactory is a serious game where the student assumes the role of CEO of a factory and they have three turns to make their company economically viable and sustainable. With each turn, the student may make choices concerning the design of a product, purchase of manufacturing machines and hiring of qualified staff. Once all decisions are made, the simulation advances 5 years and then a report is given indicating how sustainable the business is in terms of profit, environment and society.

**Interactive Product Assembly** ***(age group: teenagers and young adults).*** With the Interactive Product Assembly, the student is shown the basic principles be-hind the manual assembly process of a product for achieving maximum efficiency. The IPA provides a 3D environment where the student is challenged in putting together a radio-controlled car by using virtual reality. The application keeps score of the some key attributes that need to be respected during assembly which are the time it takes to complete the task and the correct sequence of parts.

**How to build a skateboard?** ***(age group: teenagers).*** The students have been hired by a start-up selling "do it yourself" skateboards on the internet. Their challenge is to create an assembly manual that will be delivered with the skateboard parts, using professional 3D software.

**LCA Game** ***(age group: young adults).*** In the LifeCycle serious game, the student assumes the role of sustainability manager tasked by the CEO to do the LifeCycle Assessment report on a coffee machine. The student is required to collate the data from multiple sources, which include information systems, production cells on the shopfloor and talking with different stake-holders within the factory. The student needs to make sense of the information obtained to determine what is correct, updated and unbiased. As a result, the student creates the LCA report with a set of recommendations to improve the sustainability of the product in terms of economic, environment and societal impact.

**Teaching Factory** ***(age group: young adults).*** The Teaching Factory approach, targeting young adults, aims at a broader use of novel learning methods for the introduction of young engineers to a wide spectrum of manufacturing problems. To achieve this it uses real life production for teaching purposes with training services delivered on a virtual basis. The "factory-to-classroom" operating concept of the Teaching Factory, aims at transferring knowledge from the factory to the classroom, this operation is carried out through the adoption of an industrial project in the context of academic practice, bringing together, in overlapping time and context, the industrial and academic practices. The industrial project can have varied but fixed duration that is relevant to the problem in the industrial side. This problem is deriving from a specific set of tasks, included in the product /production lifecycle and the students work on a solution for this problem.

## 4 Evaluation Approach

To evaluate the effectiveness of the ICT applications developed, a pretest-posttest quasi-experimental approach was used [17]. The experimental design procedure was composed of four steps:

- A pretest session aiming at assessing the initial levels of awareness, interest (and knowledge) of participants;
- The main TEL-based educational activity where the students engaged with the specific application developed
- A posttest session soon after the activity in order to re-assess the levels of awareness and interest of students
- A post-posttest session administered from three to six months afterwards.

A total of 461 students were randomly chosen among 21 educational institutions across five European countries (France, Italy, Switzerland, Denmark, Greece) which accepted to participate in the initiative. The distribution across the targeted age groups was: 43 participants from children age group, 218 from teenager's age group and 200 from young adult's age group. The 24 field experiments were done between February and November 2015.

For data collection, a combination of both quantitative and qualitative methods were used according to the specific age groups addressed and the variables considered. A "General Questionnaire" was also used in order to collect information on the profile and background of participants (e.g. age, gender, etc.).

The awareness about manufacturing of participants was assessed for teenagers and young adults through the "EiE Engineering and Science Attitudes Assessment" [18] that was designed in order to examine students' attitudes towards science and engineering and knowledge of general engineering concepts and technology. The questionnaire originally consisted of 20 items, which we reduced to 17, since some questions were considered to be out of scope. For the same reason, the phrasing of some of the questions was altered in order to readapt them to manufacturing domain.

For children, the "Draw-a-Factory" evaluation was used. It was based on "Draw-an-Engineer Test" [19] and readapted in order to focus it on manufacturing. In particular, the Draw-a-Factory test was based on the analysis of a drawing supported by three complementary open-ended questions. In particular, the following four questions were asked, i.e. "*Close your eyes and imagine a factory… Open your eyes. On the attached sheet of paper, draw what you imagined"* (drawing), *"Describe the factory in the picture. Write at least two sentences"* (open-ended question), *"List at least three words/phrases that come to mind when you think of this factory"* (open-ended question), and *"What kind of things do you think that happen in this factory on a typical day?"* (open-ended question). Post and post-post semi-structured interviews to investigate more in detail children's awareness and support the interpretation of the drawings were also used.

The interest about manufacturing was assessed for all the age groups through the "STEM Semantics Survey" [20]. The questionnaire was created to evaluate Interest in science, technology, engineering, mathematics, and STEM careers. It was slightly readapted by replacing engineering subscale with manufacturing one. Therefore, the 5 items for Engineering were changed into Manufacturing ones, and the 5 items for STEM Careers were changed into Manufacturing Careers ones. In addition, in order to cover the specific needs of the age group involved, further changes on the phrasing of the questionnaire were done. In particular, the word "Mundane" was replaced with the word "Dull", in order to simplify the meaning of the word and avoid possible misunderstandings by the participants.

The results of all questionnaires were normalized, reporting them on a 0 to 100 scale in a linear way, and analyzed by means of paired-sample t-tests. The significance level was set at 0.05.

## 5 Findings and Discussion of Results

In the next paragraphs, main findings about the impact of ManuSkills ICT applications on the awareness and interest of children, teenagers and young adults are summarized and the related implications are discussed.

### 5.1 Children

EcoFactory was the ICT Application targeting also children between 10 and 12 years old. Assessment of awareness was done by means of a qualitative approach, i.e. content analysis on the pre and post drawn and written answers of participants, and consequent comparison. An example of the pre and post drawings of a boy aged 11 representing his idea of a factory is illustrated in Fig. 2. In the pre-drawing the factory is polluting and it is indicated by the child as "a gloomy factory that pollutes the sky with all the gas that produces with the functioning of its machineries" and inside it "they build all the objects that make without realizing that they pollute". In the post-drawing, the factory is not polluting anymore and even the sun can be noticed at the right top of the box. The factory is now described as "shining and not polluting" and inside it "people work in harmony without toxic things but all created by means of eco-sustainable machineries".

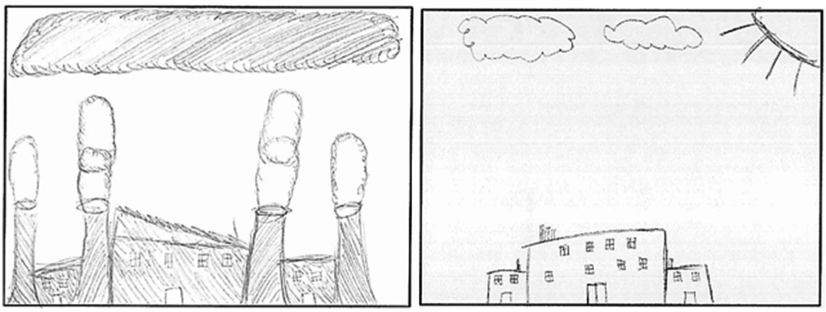


**Fig. 2.** Example of pre-post drawings

For each couple of pre-post drawings, the overall result achieved was identified and formalized in terms of Positive, Negative and No Impact. The results of this analysis showed a Positive impact on 33 participants (77 %), a Negative impact on 4 participants (9 %), and No impact on 6 participants (14 %).

The impact was significantly positive also on interest ($IN_{PRE} = 49.33$; $IN_{POST} = 62.77$) ($t = 2.7995$; $p < 0.01$). Therefore, the effects of the engagement with EcoFactory leading to a more realistic and up-to-date perception of manufacturing, were accompanied also by an increased attraction of the participants for its contents.

## 5.2 Teenagers

For teenagers, the average initial level of awareness ($AW_{PRE} = 65.43$) for manufacturing was higher than that of interest ($IN_{PRE} = 59.91$) ($t = 4.7866$; $p < 0.001$). This supports the initial idea that students can have a given level of awareness about manufacturing, but it doesn't imply necessarily their effective interest for the domain.

The engagement with ManuSkills ICT Applications allowed a general improvement of awareness for manufacturing ($AW_{POST} = 68.22$) ($t = 3.7602$; $p < 0.001$).

The impact on manufacturing interest was higher ($IN_{POST} = 64.90$) ($t = 4.2268$; $p < 0.001$). This can be explained by the fact that changes in awareness might require more time than that planned for ManuSkills activities in order to take place, while an interest in the participants can be instilled by means of interactive activities showing specific aspects of manufacturing domain.

An interesting aspect is that the greater impact for both awareness and interest was achieved on the participants with the initial lowest levels. This is showed in Figs. 3, 4, 5, and 6 where the pre and post results of awareness and interest are divided according to the three categories of Low (0–60; *in yellow in the figures*), Medium (61–80; *in blue in the figures*) and High (81–100; *in green in the figures*), and then compared. In particular, participants initially in the Low awareness category moved to the two other categories, i.e. Medium and High awareness, while participants initially in the Medium interest category moved to the High interest one.

Therefore, all the ManuSkills activities targeting teenagers were suitable to support students initially not confident with the concepts proposed. Again, this can be explained by the high level of interactivity provided by all the ICT applications and by the active participation requested in order to achieve the objectives defined.

## 5.3 Young Adults

For young adults, the average initial levels of awareness ($AW_{PRE} = 72.31$) and interest ($IN_{PRE} = 71.35$) were similar ($t = 0.6882$; $p > 0.05$) and both higher than those of teenagers for the same variables ($t = 5.4069$; $p < 0.001$) ($t = 5.4299$; $p < 0.001$). In this case this seems a reasonable result since students are closer to the working world and they already did a big choice in entering a STEM course. Nevertheless, as already pointed out, room for improvement should be also considered, since they can still eventually decide not to definitely enter in the manufacturing world after university.

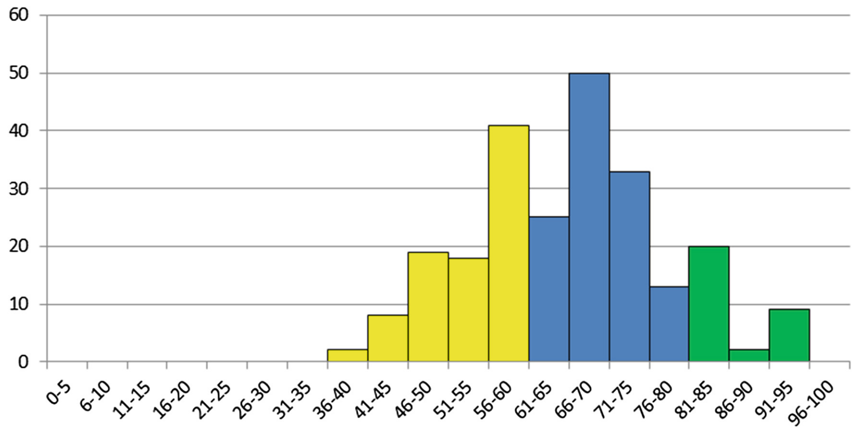



**Fig. 3.** Pre-awareness teenagers (*Number of participants per level of awareness*) (Color figure online)

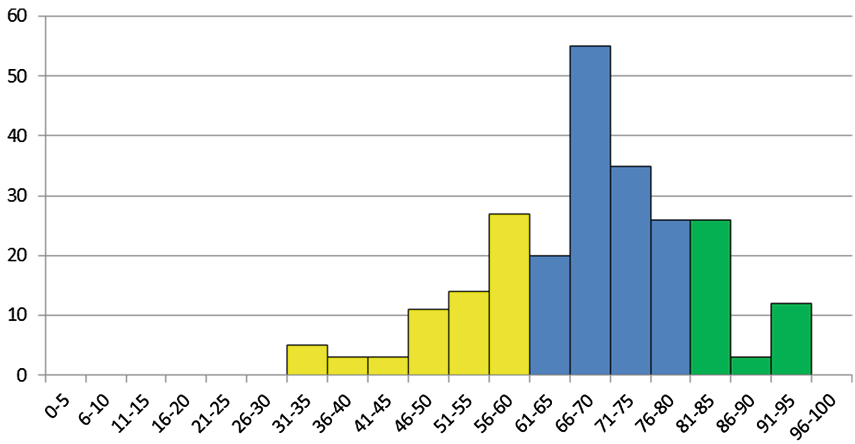



**Fig. 4.** Post-awareness teenagers (*Number of participants per level of awareness*) (Color figure online)

The engagement with ManuSkills ICT Applications also allowed an improvement of awareness for manufacturing ($AW_{POST} = 75.01$) ($t = 3.8395$; $p < 0.001$). This is interesting because is a similar result to that of obtained for teenagers. The fact that students in this case were all already involved in STEM university courses might contribute to the explanation.

The impact on manufacturing interest ($IN_{POST} = 75.39$) ($t = 3.7174$; $p < 0.001$) was similar to that on awareness and also to that on teenagers' interest. The second fact should be particularly noticed, since it shows that even university students can benefit in terms of interest from the ICT Applications developed. In fact all those applications (i.e. LCA Game, BrickPlanner, Interactive Product Assembly and Teaching Factory)

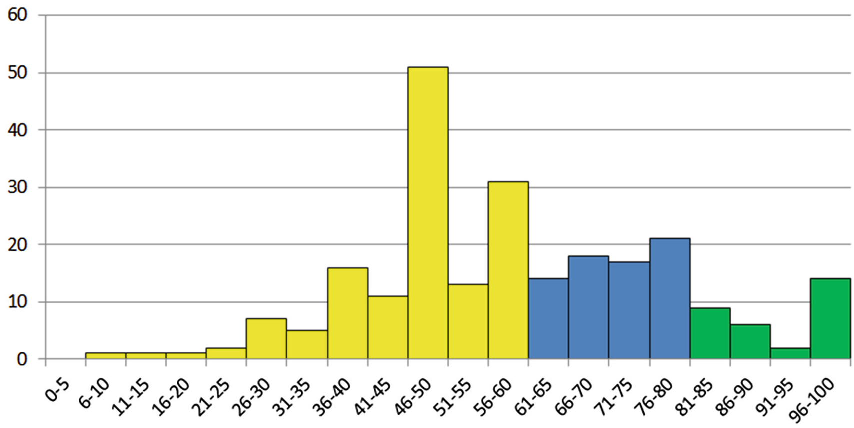



**Fig. 5.** Pre-interest teenagers (*Number of participants per level of interest*) (Color figure online)

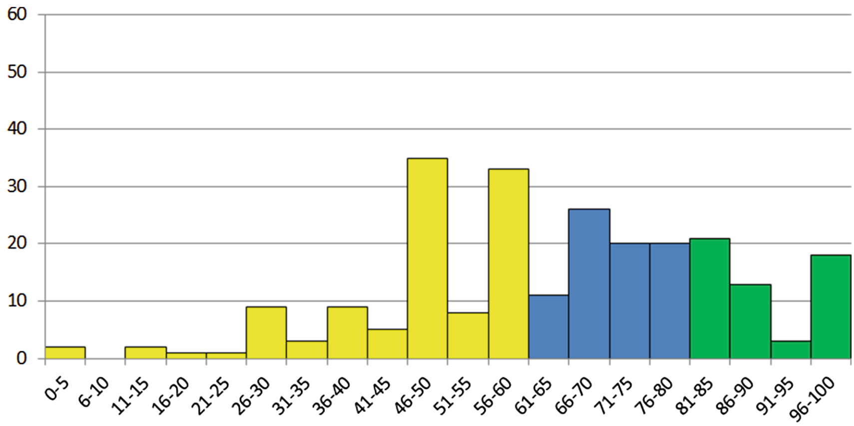



**Fig. 6.** Post-interest teenagers (*Number of participants per level of interest*) (Color figure online)

enable the representation of manufacturing concepts otherwise hardly communicable by means of traditional teaching approaches.

Also for young adults, the greater impact for awareness and interest was achieved on the participants with the initial lowest levels (Figs. 7, 8, 9, 10). Also in this case, participants initially in the Low awareness category moved to the two other categories, i.e. Medium and High awareness, while participants initially in the Medium interest category moved to the High interest one.

Therefore, all the ManuSkills activities targeting young adults were suitable to support an increase in their awareness and interest about manufacturing, even though they were students already engaged in a STEM higher education path. This fact supports the idea that even in this situation the use of non-traditional ICT-based teaching

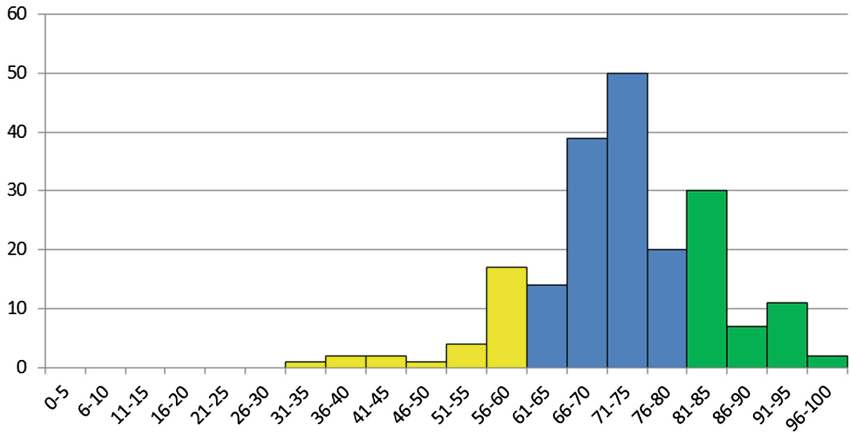



**Fig. 7.** Pre-awareness young adults (*Number of participants per level of awareness*) (Color figure online)

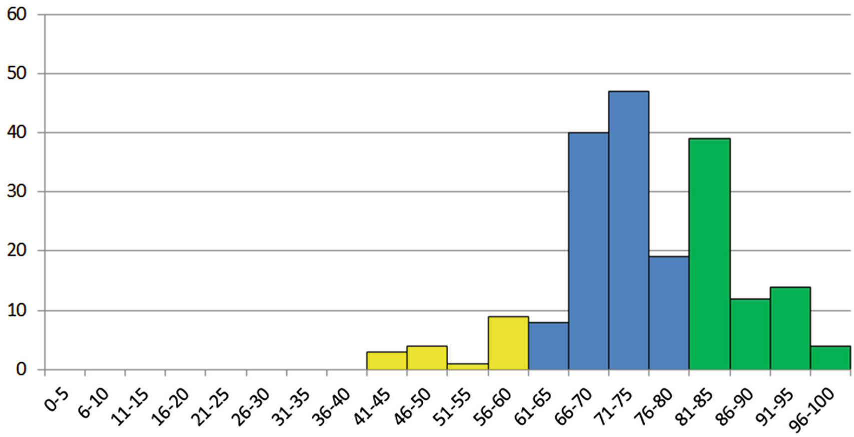



**Fig. 8.** Post-awareness young adults (*Number of participants per level of awareness*) (Color figure online)

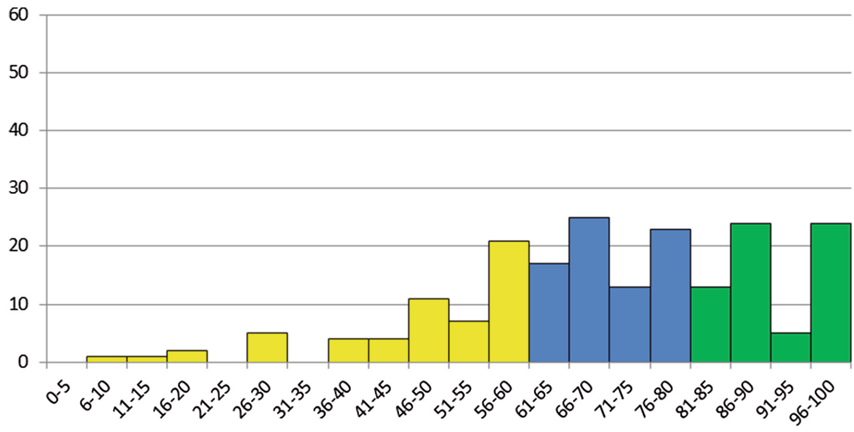



**Fig. 9.** Pre-interest young adults (*Number of participants per level of interest*) (Color figure online)

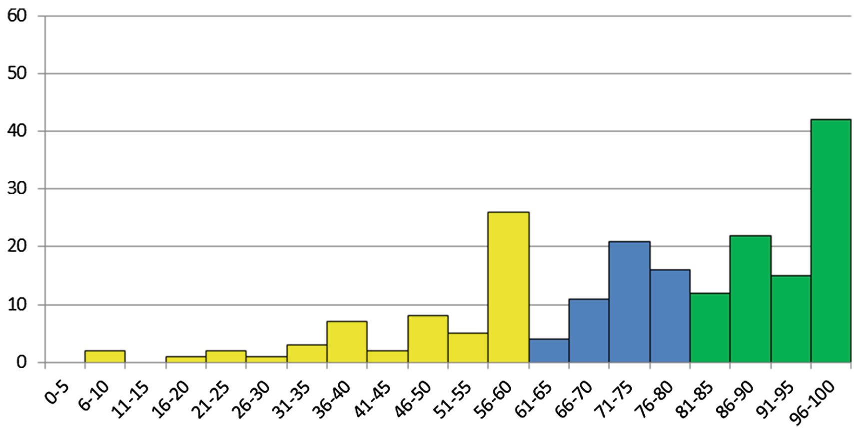



**Fig. 10.** Post-interest young adults (*Number of participants per level of interest*) (Color figure online)

approaches can be useful in order to present to students specific topics and positively affect their perception towards the manufacturing domain.

## 6 Conclusion

The problem of skills gap is becoming always more urgent in manufacturing. Among the several actions that can be implemented to solve the issue, the increase of awareness and interest of young talent in manufacturing is considered pivotal. The field experiments conducted within the ManuSkills project showed that the use of proper interactive ICT applications targeting children, teenagers and young adults, has a positive impact on that process.

Despite these encouraging preliminary results, further work should be still done on several points. In particular, the specific impact of the single delivery mechanism (e.g. serious game, simulation, virtual reality and teaching factory) on both awareness and interest should be understood more in detail, in order to see the specific differences among the activities. In addition, long-term effects of the engagement of young talent with the ICT-supported activities should be explored, in order to understand the retention of the changes in perception also some months after the first contact with them. As a further long term investigation, the hoped effective connection between a change in awareness and interest for manufacturing and a change in awareness and interest for a career in this field should be analyzed. Finally, all the above-mentioned results should be characterized for each single age group, thus finding how to properly introduce in the existing STEM curricula the ICT-supported activities proposed in order to support the definition of long-term awareness and interest programmes specifically targeting manufacturing.

## References

1. Deloitte U.S. and the Council on Competitiveness: Global Manufacturing Competitiveness Index (2013)
2. McKinsey & Company: Manufacturing the future: the next era of global growth and innovation (2012)
3. Deloitte and The Manufacturing Institute: Boiling point? The skills gap in U.S. manufacturing (2011)
4. Manenti, P.: The journey toward the people-intensive factory of the future. IDC Report (2013)
5. Perini, S., Oliveira, M., Costa, J., Kiritsis, D., Kyvsgaard Hansen, P.H., Rentzos, L., Skevi, A., Szigeti, H., Taisch, M.: Attracting young talents to manufacturing: a holistic approach. In: Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World IFIP Advances in Information and Communication Technology, Ajaccio (2014)
6. World Economic Forum: The future of manufacturing - Opportunities to drive economic growth (2012)
7. Economist Intelligence Unit: Plugging the skills gap - Shortages among plenty (2012)
8. Skevi, A., Szigeti, H., Perini, S., Oliveira, M., Taisch, M., Kiritsis, D.: Current skills gap in manufacturing: towards a new skills framework for factories of the future. In: Grabot, B., Vallespir, B., Gomes, S., Bouras, A., Kiritsis, D. (eds.) Advances in Production Management Systems. IFIP AICT, vol. 438, pp. 175–183. Springer, Heidelberg (2014)
9. Frehill, L.M.: Education and occupational sex segregation: the decision to major in engineering. Sociol. Q. **38**(2), 225–249 (1997)
10. Merikle, P.M., Smilek, D., Eastwood, J.D.: Perception without awareness: perspective from cognitive psichology. Cognition **79**(1), 115–134 (2001)
11. Huang, R., Sarigollu, E.: How brand awareness relates to market outcome, brand equity, and the marketing mix. J. Bus. Res. **65**(1), 92–99 (2012)
12. Marton, F.: Learning and Awareness. Psychology Press, Hove (1997)

13. Merikle, P.M.: Toward a definition of awareness. Bull. Psychon. Soc. **22**(5), 449–450 (1984)
14. Schmidt, R.: Consciousness and foreign language learning: a tutorial on the role of attention and awareness in learning. In: Attention and awareness in foreign language learning, pp. 1–63 (1995)
15. Schiefele, U.: Interest, learning, and motivation. Educ. Psychol. **26**(3–4), 299–323 (1991)
16. Rawal, P.: AIDA marketing communication model: stimulating a purchase decision in the minds of the consumers through a linear progression of steps. In: IRC's International of Multidisciplinary Journal of Research in Social and Management Sciences (2013)
17. Cook, T.D.: Quasi-experimental design. In: Wiley Encyclopedia of Management (1979)
18. Gibbons, S., Hirsch, L., Kimmel, H., Rockland, R. Bloom, J.: Middle school students' attitudes to and knowledge about engineering. In: Proceedings of the 2004 International Conference on Engineering Education, FL (2004)
19. Knight, M., Cunningham, C.: Draw an engineer test (DAET): development of a tool to investigate students' ideas about engineers and engineering. In: Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition (2004)
20. Tyler-Wood, T., Knezek, G., Christensen, R.: Instruments for assessing interest in STEM content and careers. J. Technol. Teach. Educ. **18**(2), 341–363 (2010)

# Does Taking a MOOC as a Complement for Remedial Courses Have an Effect on My Learning Outcomes? A Pilot Study on Calculus

Mar Pérez-Sanagustín(✉), Josefina Hernández-Correa, Claudio Gelmi, Isabel Hilliger, and María Fernanda Rodriguez

School of Engineering, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna, 4860 Macul, Santiago (RM), Chile
{mar.perez, jmhernal, cgelmi, ihillige, mfrodri3}@uc.cl

**Abstract.** This paper presents the results of a pilot study about students' adoption and learning outcomes of 4 MOOCs proposed as a complementary resource for traditional remedial courses on calculus. While the MOOCs were not mandatory, traditional remedial courses were required for those freshmen failing a diagnostic exam. The effects on 589 freshmen students were investigated. The data analysis shows that up to 16 % of the students were active in the MOOCs under study, mostly during the days before taking the diagnostic exam that preceded the traditional face-to-face remedial courses. Trace data about learner actions within the platform were collected as well as the students' scores. According to a statistical comparison of the students' exam scores and their interaction behavior with the MOOCs, we observe that active students had more chances of passing the diagnostic exam and skipping the required remedial courses. However, we found no significant differences on the remedial course exam scores between the students that were active in the MOOCs and those that were not. These findings suggest that MOOCs are a good solution to strengthening skills and reviewing concepts, but that more guidance is needed when used as a complement to traditional f2f courses.

**Keywords:** Moocs · Remedial courses · Higher education · Pilot study · Adoption · Learning outcomes

## 1 Introduction

Massively Open Online Courses (MOOCs) present new opportunities for facilitating teaching and learning [14]. MOOCs allow flexible learning anytime and anywhere, diversifying the variety of tasks that can be included in any course structure [15]. Lately, several case studies have documented different ways in which elite universities have integrated these courses into their curricula, broadening their teaching and learning strategies by implementing blended or hybrid learning approaches [5, 8, 17].

Two trends were observed in these case studies. The first trend (1) is using MOOCs as a complement of traditional teaching. For example, a study shows how Stanford University integrated MOOCs in a traditional course by asking students to watch video lectures, participate in discussion forums, complete quizzes and program assignments


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 221–233, 2016.
DOI: 10.1007/978-3-319-45153-4_17

in an online platform [13]. 26 students had to complement their learning with information about topics not addressed in the MOOC. The results show that students' attendance increased by 20 % and their engagement with the course content increased by 40 % [3]. Another example along these lines was developed by the University of Washington, which introduced blended learning in a traditional biology class. They were able to reduce its fail rate from 17 % to 4 %. Furthermore, the approval rates of the course increased from 14 % to 24 % since the initiative [2].

On the other hand, (2) MOOCs are used as remedial courses. Examples of these are the zero level courses developed by some universities. Universidad Carlos III de Madrid [11] analyzed the effect of a zero level course. In this experience, students took a diagnostic and a final exam, and the results indicated that students increased by 21 % the score in the final exam after the course. Regardless of other case studies in North America and Europe [1, 7], the effect of the MOOCs deserves further exploration in other countries to enrich current literature.

In order to contribute to the understanding of the MOOC-based models that use MOOCs to complement or substitute traditional remedial courses, this paper reports on the findings of a pilot study at the School of Engineering in Pontificia Universidad Católica de Chile (UC-Engineering). Specifically, we investigated the effects of 4 MOOCs on calculus for freshmen. From now on we call these MOOCs "service MOOCs" according the framework proposed in [17]. That is, MOOCs that students take voluntarily (partially or completely), and as a complement to the curriculum or a traditional course but no institutional recognition is given for completing this MOOC. In Sect. 2, we describe the context in which this study was carried out, as well as the research questions addressed. Also in this section, we describe the participants of the study, the data gathering techniques and the procedures we used for the analysis. In Sects. 3 and 4, respectively, we report the main results obtained and the lessons learned from the study as well as its limitations. Finally, in Sect. 5, we present the main conclusions, and future avenues. Altogether, this work provides a better understanding of the effects of this type of MOOC-based initiatives in terms of students' adoption and learning outcomes.

## 2 The Pilot Study

### 2.1 Context and Research Questions

About 600 Freshmen College students are admitted to the UC-Engineering every year. In order to get accepted in this program, students must be in the top positions of their high school ranking, besides demonstrating outstanding achievement in high school and in an admission exam that evaluates their knowledge in math, science, and language. Even so, students come with different understanding of basic calculus concepts, and their knowledge on these topics is often insufficient to successfully address the calculus courses that are imparted in the first year.

In the recent years, UC-Engineering freshmen have been required to take a calculus diagnostic exam right after they are informed that they have been admitted. The exam is divided into 4 modules: Algebra and Functions (M1), Trigonometry (M2), Polynomials and Complex Numbers (M3), and Sequences and Series (M4). Students that fail in a specific content are required to take a 2-day traditional course on each of the failed modules. In these courses, professors reinforce main theoretical topics, besides facilitating students' learning with guided exercises. After each course, students have to take a final exam to evaluate their progress in the respective module content.

Although this strategy has been a way of promoting students' calculus readiness, the experience from the last two years has shown some limitations: (1) low participation rates in the required remedial courses due to the fact that students that do not live in Santiago had difficulties to attend face-to-face courses; and (2) lack of individualized instruction considering that not all the students need to review the same topics. In order to address these limitations, last year the school decided to produce a service MOOC for each module and offer them as a complementary support for students' learning in the specific theoretical concepts. Since participating in the MOOCs was voluntarily, the main objective of this study was to analyze the impact of this initiative both in terms of students' adoption and learning outcomes. Specifically, two research questions were addressed:

- **RQ1. What is the students' adoption of this MOOC initiative?** This question aims at studying the students' use of the MOOCs in terms of their interactions with the course content in order to better understand who, how and when they use the provided courses.
- **RQ2. What are the effects of participating in the MOOCs in terms of students' learning outcomes?** This question aims at better understanding two aspects: (1) whether or not using the online platform before the diagnostic exam gives the students a better probability of passing it; and (2) whether or not students that use the MOOCs have better scores in the traditional remedial courses' final exams.

### 2.2 Description of the Pilot Study

The pilot study took place at UC-Engineering between December 27th 2015 and 29th January, 2016. The MOOCs were produced by 3 teaching assistants and were deployed in the Open edX platform as part of the UC-Engineering online initiative[1]. The MOOCs did not follow the same structure than the traditional remedial courses. Nonetheless all the contents of the MOOC were designed to align with the learning objectives and topics addressed in the traditional remedial courses. The MOOCs were all open to anyone interested, both from and outside the UC-Engineering.

The MOOCs were available before the students knew that they had been admitted in UC-Engineering. MOOCs were announced by e-mail and flyers a week before releasing the admission results to all those that had manifested their interest in studying at UC-Engineering. Additional outreach to students involved posting in the official Engineers' web page, so all prospective students were informed that they could register on the platform and take MOOC. Once accepted, all freshmen were registered in the

[1] Open edX Platform 'Ingeniería UC Online': http://online.ing.uc.cl/.

MOOC provider platform during the admission day, so all of them could access the 4 MOOCs. All the MOOCs are self-paced, so no restrictions or deadlines were proposed. Students were also informed that the participation in the MOOC courses was voluntary. Students were required to take a diagnostic exam to assess their prior knowledge and skills in calculus. Depending on their results on the diagnostic exam, students had to attend the mandatory specific remedial courses that were imparted traditionally before the first semester begins. Table 1 shows a time line of the different milestones in this case study, showing also the duration of each traditional remedial course and the dates of the final exams that the students took after participating in a required course to evaluate their progress in the respective content.

**Table 1.** Pilot study timeline

| Dates | Activity/Milestones |
|---|---|
| 27th Dec. 2015–10th Jan. 2016 | Dissemination effort via e-mail, web-page and flyers to potential engineering students |
| 11th Jan. | Publication of the Admission Results (00:00 h)<br>Presentation session of the accepted students and registration to the platform. |
| 13th Jan. | Calculus Diagnostic Exam |
| 14th Jan. | Publication of exam results |
| 18th Jan.–20th Jan. | M1 (Algebra and Functions)<br>Final exam of the traditional course M1<br>Link to the complementary service MOOC M1:<br>http://online.ing.uc.cl/courses/PUC/EINP001/2015_EINP001/info |
| 20th Jan.–25th Jan. | M2 (Trigonometry)<br>Final exam of the traditional course M2<br>Link to the complementary service MOOC M2:<br>http://online.ing.uc.cl/courses/PUC/EINP003/2015_EINP003/info |
| 25th Jan.–27th Jan. | M3 (Polynomials and Complex Numbers)<br>Final exam of the traditional course M3<br>Link to the complementary service MOOC M3:<br>http://online.ing.uc.cl/courses/PUC/EINP004/2015_EINP004/info |
| 27th Jan.–29th Jan. | M4 (Sequences and Series)<br>Final exam of the traditional course M4<br>Link to the complementary service MOOC M4:<br>http://online.ing.uc.cl/courses/PUC/EINP002/2015_EINP002/info |

### 2.3 Participants and Sample

Although the MOOCs were open to anyone, in this study we only took as a sample for the analysis those students that were admitted in UC-Engineering and took the diagnostic exam on calculus. 589 students (N = 589) took the diagnostic exam on calculus. Those who passed the diagnostic exam (Students Passing Diagnostic, SPD) and those who did not (Students Failing Diagnostic, SFD) were the sample of analysis of our study. Since not all attended the remedial courses if they failed the exam, we separated the sample into two groups: students that attended the traditional remedial courses

(Students Attending Remedial, SAR), and distinguished among those who passed the corresponding final exam (Students Passing Remedial, SPR) and those who did not (Students Failing Remedial, SFR) (Table 2).

**Table 2.** Number of Students in each phase according to mathematical content.

| Course | Diagnostic exam | | Traditional remedial courses | | |
|---|---|---|---|---|---|
| | SPD | SFD | SAR | SPR | SFR |
| M1 | 504 (86 %) | 85 (14 %) | 64 | 53 (83 %) | 11 (17 %) |
| M2 | 170 (29 %) | 419 (71 %) | 281 | 219 (78 %) | 62 (22 %) |
| M3 | 261 (44 %) | 328 (56 %) | 223 | 208 (93 %) | 15 (7 %) |
| M4 | 325 (55 %) | 264 (45 %) | 171 | 104 (61 %) | 67 (39 %) |

### 2.4 Data Collection and Analysis

The data gathered from the sample of study came from many different sources. First, we worked with the students' **scores in the diagnostic exam** (ScoresDE-M1, ScoresDE-M2, ScoresDE-M3 and ScoresDE-M4) and the scores obtained at each final exam of the required course (ScoresRE-M1, ScoresRE-M2, ScoresRE-M3 and ScoresRE-M4). These exams contemplate a 0–100 % scale, where a 100 % score would mean that they got every question right, and students passed the exams if they got a score of 50 % or higher.

The **students' activity and interaction** patterns with the MOOCs are represented by the number of movements each student made in each MOOC before the diagnostic test and during the required courses. The movements were extracted from the MOOCs' computational logs, where every action or movement each student does in the platform is registered (Logfiles). The numbers of active and non-active students are the measures of "adoption" in this study.

The **students' prior knowledge** was defined as the students' admission scores composed by: Math (MAT), Science (CIE), and Language (LEN) Chilean University Admission Exams scores, along with a score according to their high school grades (NEM) and class ranking (RKG). All these individual scores have a scale from 0 to 850. Finally, PING is the weight average admission score, computed as: 20 % NEM, 20 % RKG, 10 % LEN, 35 % MAT and 15 % CIE. These data is what we take as a reference of students' prior knowledge and skills. Lastly, in order to understand academically where the students that adopted the MOOCs platform before the diagnostic exam came from, we divided the cohort in quartiles according to their PING. The groups are Q1, Q2, Q3 and Q4; where Q1 is the group with the lowest PING and Q4 is the one with the highest scores.

In order **to address RQ1 about the students' adoption** of the MOOC initiative and their behavior in the platform, we first organize the students into "active" and "non-active" depending on their usage of the platform in two periods: (1) before the diagnostic exam (Before Diagnostic Phase, BDP), and (2) during remedial courses (During Remedial Phase, DRP). We classified the students into these two groups by analyzing the number of movements that each student registered on the different MOOCs in each phase.

After classifying the students into active and non-active, we plotted the number of movements in a bar graph from the beginning of the study until the end to analyze the activity patterns in the different periods. Also, we analyzed the students' interactions with both the video-lectures and the exercises (quizzes and other activities). We used this data to get an idea about whether the students used the MOOC for reviewing theoretical concepts through video-lectures or exercising.

In order **to address RQ2 about the students' learning outcomes** we conducted several statistical analyses and looked for correlations between the students' activity in the MOOCs with the scores they each obtained in the diagnostic exam and in the remedial course exams. These calculations allowed us to understand whether the interactivity levels have an influence on their results.

Then, in order to understand if the active students had more chances of passing the exams, we performed a t-test for the scores between the non-active and active students in both diagnostic exams and the remedial courses. Given that the results observed in this first analysis were significant for the diagnostic test, we applied a proportion test to the percentage of approval rates between active and non-active students. Thirdly, in order to understand the effect of the platform along with other variables that characterize the students' prior knowledge, we performed a stepwise multivariable regression analysis that related the scores of the diagnostic or the remedial exams using as initial predictors the national admission exam scores NEM (high school GPA score), MAT (mathematics score), CIE (science score), and RKG (ranking score), and the categorical variable "active" or "non-active" student, which represents the platform adoption strategy of the student. All statistical analyses were carried out using MINITAB 17 (www.minitab.com).

## 3 Results

This section reports on the results obtained from the analysis to address the two research questions. Subsect. 3.1 presents the results about the students' adoption of the MOOC initiative, and Subsect. 3.2 about the effects on students' learning outcomes.

### 3.1 Students' Adoption of the MOOC Initiative

**R1.1. Up to 16 % of the students were active in the MOOCs. Active students used the MOOCs more before the diagnosis exam than during the required courses**. Between 5 % and 16 % were active in the MOOCs. As shown in Table 3, M2 is the course that concentrated most of the activity, followed by M1, M4, and M3. M2 is a MOOC about trigonometry, a content that is no longer evaluated in the college admission test since 2014[2].

Figure 1 shows the activity of the students during the pilot study. The average number of interactions per day per MOOC during the three days before the diagnostic exam (from January 11$^{th}$, which is when the students found out they had been accepted,

[2] http://www.educarchile.cl/ech/pro/app/detalle?id=225229.

**Table 3.** Active MOOC students vs. Non-active

| Course | Before diagnostic Phase, BDP | | During remedial phase, DRP | |
|---|---|---|---|---|
| | Active | Non-active | Active | Non-active |
| M1 | 14 % (N = 84) | 86 % (N = 505) | 7 % (N = 42) | 93 % (N = 547) |
| M2 | 16 % (N = 97) | 84 % (N = 492) | 13 % (N = 79) | 87 % (N = 510) |
| M3 | 8 % (N = 48) | 92 % (N = 541) | 5 % (N = 29) | 95 % (N = 560) |
| M4 | 12 % (N = 73) | 88 % (N = 516) | 10 % (N = 56) | 90 % (N = 533) |

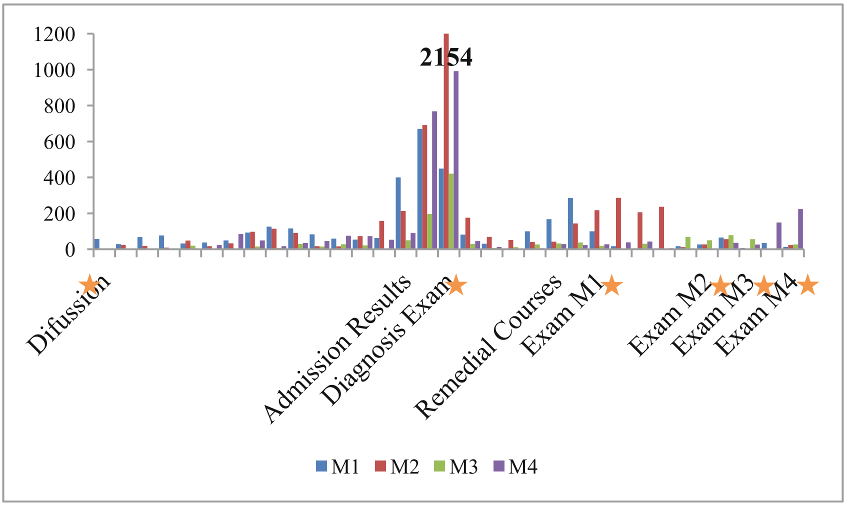



**Fig. 1.** Total amount of movements in the 4 MOOCs before the calculus exam and during the courses

to January 13th) is 591 (with a total of 7.095 learner actions traced), whereas there are only 61 daily interactions per MOOC during the face-to-face required courses (with a total of 3.701movements registered from January 14th through January 29th). Specifically, students interacted more with each MOOC during their participation in the required course. M1 and M2 were the MOOCs most used.

**R1.2. Students used the courses for exercising.** Table 4 shows that the exercise sections registered more interactions than the video sections. This result is observed in all courses and in both phases. By both phases, we mean before the diagnostic exam and during the remedial courses.

### 3.2 Effects of the MOOC Initiative on Students' Learning Outcomes

**R2.1. Students who were active in the MOOCs before the diagnostic exam showed better scores on this exam, but no significant effect was observed in the scores of students that were required to take final exams after traditional face-to-face courses**. Results in Table 5 indicate that there is no statistically significant difference in

**Table 4.** Interactions captured in each MOOC section Before the Diagnostic Exam Phase (BDE) and During Remedial Phase (DRP) and proportions of interactions per MOOC per phase

| | BDE | | DRP | |
|---|---|---|---|---|
| | Videos-lectures | Exercises | Video-lectures | Exercises |
| M1 | 503 (39 %) | 793 (61 %) | 194 (38 %) | 316 (62 %) |
| M2 | 439 (22 %) | 1.516 (78 %) | 240 (28 %) | 626 (72 %) |
| M3 | 37 (10 %) | 341 (90 %) | 44 (20 %) | 181 (80 %) |
| M4 | 248 (23 %) | 853 (77 %) | 40 (16 %) | 205 (84 %) |
| **Total** | **1.227 (26 %)** | **3.503 (74 %)** | **580 (28 %)** | **1.328 (72 %)** |

**Table 5.** Diagnostic exam scores and final exam results from required courses to the students' use of each MOOC.

| Course | Group | N | Mean | SD | P-value |
|---|---|---|---|---|---|
| ScoreDE-M1 | Non-active | 505 | 0.760 | 0.147 | 0.002 |
| | Active | 84 | 0.805 | 0.129 | |
| ScoreDE-M2 | Non-active | 492 | 0.383 | 0.273 | 0.000 |
| | Active | 97 | 0.536 | 0.215 | |
| ScoreDE-M3 | Non-active | 541 | 0.607 | 0.183 | 0.004 |
| | Active | 48 | 0.676 | 0.166 | |
| ScoreDE-M4 | Non-active | 516 | 0.585 | 0.260 | 0.000 |
| | Active | 73 | 0.720 | 0.194 | |
| ScoreRE-M1 | Non-active | 65 | 0.748 | 0.161 | 0.971 |
| | Active | 7 | 0.750 | 0.166 | |
| ScoreRE-M2 | Non-active | 232 | 0.701 | 0.158 | 0.621 |
| | Active | 50 | 0.713 | 0.166 | |
| ScoreRE-M3 | Non-active | 208 | 0.820 | 0.134 | 0.525 |
| | Active | 16 | 0.842 | 0.125 | |
| ScoreRE-M4 | Non-active | 147 | 0.644 | 0.192 | 0.040 |
| | Active | 25 | 0.556 | 0.220 | |

the final scores of the remedial exams (ScoreRE-M1…M4) between those students that were active in the MOOCs and those who were not active. The only exception corresponds to ScoreRE-M4, where active students obtained a lower mean score compared to the non-actives ones. In contrast, we found that the mean scores of the active users were significantly higher than the non-active students in all cases of the Diagnostic test (ScoreDE-M1…M4).

**R2.2. Students that were active users in the MOOCs before the diagnostic exam reported statistically higher approval rates in this test**. Results in Table 6 show that the percentage of active users passing the Diagnostic Exam is higher than those who were non-active. This result is especially different (with more than 17.3 points of difference) for the one that took the M2 MOOC, which corresponds to the MOOC that registered the higher amount of learner actions (see Fig. 1).

**Table 6.** Percentage of students that passed the diagnostic test, classified as Active and Non-active users.

| Course | Active users (n) | Non-active users (n) | Fisher's exact test P-value |
|---|---|---|---|
| M1 | 94 % (79) | 84.1 % (425) | 0.009 |
| M2 | 43.3 % (42) | 26 % (128) | 0.001 |
| M3 | 58.3 % (28) | 43.1 % (233) | 0.030 |
| M4 | 69.9 % (51) | 53.1 % (274) | 0.005 |

**R2.3. Being active in the MOOC platform appears to be a predictor variable for the score of the Diagnostic Exam, but not for the scores on the final exams of required courses, in which the only predictor variable is the math scores the students got on their University Admission Exams (MAT)**. Table 7 shows the results of the stepwise multivariable regression analysis. This analysis allowed us to have a better understanding of what variables explain better the approval rates in each of the phases. The results in Table 7 show that several of the predictors were statistically significant for the diagnostic exam phase, including the categorical variable "Active user" (taken as a measure of adoption). For the traditional remedial courses, only the MAT score was a statistically significant predictor of the final exam score in each course.

**Table 7.** Regression analysis of the different course scores.

| Course | Diagnostic exam | | Traditional remedial courses | |
|---|---|---|---|---|
| | Significant variables | P-value | Significant variables | P-value |
| M1 | NEM | 0.000 | MAT | 0.000 |
| | MAT | 0.000 | | |
| | CIE | 0.029 | | |
| | RKG | 0.043 | | |
| | Active user | 0.005 | | |
| M2 | MAT | 0.000 | MAT | 0.000 |
| | CIE | 0.002 | | |
| | RKG | 0.019 | | |
| | Active user | 0.000 | | |
| M3 | NEM | 0.018 | MAT | 0.000 |
| | MAT | 0.000 | | |
| | Active user | 0.021 | | |
| M4 | NEM | 0.000 | MAT | 0.000 |
| | MAT | 0.000 | | |
| | Active user | 0.000 | | |

**R2.4. The activity rates on the MOOCs do not depend on the PING (student's final admission score).** Table 8 shows the percentage of active students that fall in each of the quartiles by PING. The results show that the percentages of active students are similar independent to the quartile they belong to.

**Table 8.** Adoption rates according to PING quartiles before the diagnostic exam.

| | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| M1 | 33.3 % | 22.6 % | 29.8 % | 14.3 % |
| M2 | 19.6 % | 24.7 % | 28.9 % | 26.8 % |
| M3 | 25.0 % | 18.8 % | 22.9 % | 33.3 % |
| M4 | 21.9 % | 19.2 % | 24.7 % | 34.2 % |

# 4 Lessons Learned

The lessons reported in this section were obtained from reflecting on the pilot study results from both the student's adoption and the students' learning outcomes. In an effort to highlight those aspects of the study that could be applied to other contexts, we report on the limitations and analyze the issues that emerge from this work and would deserve further work.

**First, students are not yet enough prepared to adopt MOOCs if proposed as a complement to traditional courses and if they are not mandatory**. The results of our study show that between 8 % (the minimum) (N = 48) and 16 % (the maximum) (N = 97) of the students were active in the MOOCs under study for the diagnostic exam. The activity in the MOOCs decreased during the traditional remedial courses period to 5 % (the minimum) (N = 29) and 13 % (the maximum) (N = 79) of the students, depending on the MOOC. Considering how the online initiative was promoted within the students, these percentages are less than what we expected. Prior studies show that the adoption is higher when MOOCs are proposed as a mandatory course.

**Second, MOOCs are a good mechanism to help students refresh their previous knowledge on a particular topic regardless of not having any support, but they need to be carefully integrated with a traditional course in order to impact on students' learning outcomes**. The data of this study shows that those students that used the MOOC before the diagnostic exam had significantly more chances to pass this exam and skip the traditional required courses. Also, we observe through a regression analysis that passing the exam is not only dependent on the use of the MOOC, but also influenced by students' NEM, MAT, CIE and/or RKG scores. This last result is not surprising, since previous studies show the importance of the students' prior knowledge to succeed in a MOOC [12]. However, what it is interesting is that, when students participate in the MOOC as a complement to the traditional remedial course, no effects on the learning outcomes are observed and prior knowledge is the only variable able to predict the learning outcomes. Other case studies about blended learning approaches are especially useful when the MOOC is fully integrated as part of the traditional course [2, 7, 12]. These results suggest that service MOOCs that are not fully integrated with traditional courses might be not as beneficial for the students in terms of learning outcomes.

**Third, the study of students' adoption of MOOCs might signal what students are expecting to reinforce regarding the lack of opportunities to learn required skills and contents.** A curriculum narrowing effect has emerged from the fact that the national admission test is not evaluating trigonometry, a branch of mathematics that is required for succeeding in engineering calculus courses. Therefore, the availability of M2 might have raised student awareness of the importance of this topic for succeeding not only in the diagnostic test, but also in their first year of college. Further research on MOOCs used as a complement for improving academic preparation for college should be addressed.

**Fourth, the interactivity patterns show that students tend to be active in the MOOCs more intensively before the exams, but this activity is very different between the MOOCs' topics and the phase of the study.** The results of this study show that most of the movements on the course were registered before the diagnostic exam and before the exams of each remedial course. However, students show a better self-regulation pattern in the activity when the MOOC is aligned with the remedial face-to-face course. Several studies indicate that thanks to the work in virtual platforms, students can follow their own learning pace [4]. This is obvious, for example, when observing the different hours of the day that the students access the online course in our pilot study. But previous work has reported that although most of the participants in a MOOC tend to follow a linear path through the course content, these paths can vary depending on characteristics such as the age or the country of origin [10]. In addition, differences were observed on the activity patterns in each of the courses. Course M2 registered more movements than the other 3, followed by M1, then M4 and finally M3. Since all the courses where prepared by the same teachers and used the same resources, we suggest that this difference can be due to the needs of the students on the different course topics. For example, M2 and M3, which were the MOOCs registering a higher activity, work on topics that students do not practice in their previous studies before entering the university. But it could also be due to the quality of the MOOCs. Moreover, we need to take into account the students' diversity, since some students might be interested only in certain parts of the course. Also there are students that lose interest as they advance in the courses, because they feel unable to achieve the MOOCs' goals [7].

**And fifth, service MOOCs should be designed for diversifying learning activities and exercises**. We showed that most of the students' activity was registered in the exercises. Recent work shows the importance of including exercises for practicing, especially in topics related with sciences and technology [16]. The results of this study corroborate the importance of designing MOOCs that include activities for exercising.

## 5 Conclusions and Future Work

There is little empirical research that analyzes the effects of MOOC-based models in remedial courses in terms of students' adoption and learning outcomes. This pilot study serves to prove that promoting the use of MOOCs as a complement for remedial traditional courses gives those students better chances of succeeding in the corresponding exams. Also, their interactivity in the MOOCs varies greatly given that students can follow their own learning pace.

Future work includes further investigation of the results obtained. First, more information needs to be extracted to better understand the reasons that moved active students to participate in the MOOCs and the reasons of those who did not. For example, the course content could not have been interesting enough, so evaluations on the MOOCs' content would be needed to be able to judge this aspect. Second, we need to better understand how students' self-regulate in these type of courses and what type of support they need to encourage future freshmen students to use the MOOCs and obtain better results in the diagnostic exam and remedial courses. Also, we should consider analyzing the students' social learning aspects. Finally, and taking into account that the MOOCs are available also during the calculus courses of the first year, future work includes analyzing how is the adoption of these MOOCs during the first semester and what are the learning outcomes of those who used them more intensively.

**Acknowledgements.** This work was supported by FONDECYT (Chile) N 11150231, the MOOC-Maker Project (561533-EPP-1-2015-1-ES-EPPKA2-CBHE-JP), and the Comisión Nacional de Investigación Científica - CONICYT Ministry of Education, Chile, Ph.D.



## References

1. Kop R., Fournier H., Sitlia H. The value of learning analytics to networked learning on a personal learning environment. In: Long P., Siemens G., Conole G., Gasevic D. (eds.) Proceedings of the First International Conference on Learning Analytics and Knowledge, Banff, Alberta, Canada, February 27–March 01 2011. National Research Council Canada, Banff (2011). http://nparc.cisti-icist.nrccnrc.gc.ca/npsi/ctrl?action=rtdoc&an=18150452
2. Aronson, N., Arfstrom, K., Tam, K.: Flipped Learning in Higher Education Always Learning. Pearson, London (2013)
3. Collier, A., Caulfield, M.: What happens in distributed flips? Investigating students' interactions with MOOC videos and forums [Web log post], 2 May 2013. http://redpincushion.me/2013/05/02/what-happens-in-distributed-flips/
4. Fernandez, L.: B-learning as an alternative to lecturing for first semesters of engineering studies. II Congreso Internacional sobre Aprendizaje, Innovación y Competitividad (CINAIC) (2013)
5. Luján, S.: From the traditional lecture to the MOOC: twelve years of evolution of a subject about web application programming. Revista Docencia Universitaria **11**, 279–300 (2013)

6. SCOPEO—SCOPEO INFORME Nº2: MOOC: Estado de la situación actual, posibilidades, retos y futuro. Junio 2013. Scopeo Informe No. 2 (2013). http://scopeo.usal.es/wp-content/uploads/2013/06/scopeoi002.pdf
7. Yin, R.K.: Preparing to Collect Cases Study Evidence, Case Study Research: Design and Methods, pp. 96–98. Sage publications, London (2013)
8. Zhang, Y.: Benefiting from MOOC. In: World Conference on Educational Multimedia, Hypermedia and Telecommunications, vol. 2013, no.1, pp. 1372–1377 (2013)
9. Del Campo, B., Macià, M., Manjabacas, G.: ¿Qué podemos hacer para solventar las carencias en matemáticas de los alumnos de nuevo ingreso?. Actas de las XX Jornadas de Enseñanza Universitaria de Informática, Jenui 2014, pp. 295–302, Oviedo, Julio 2014 (2014)
10. Guo, P.J., Reinecke, K.: Demographic differences in how students navigate through MOOCs. In: Proceedings of the First ACM Conference on Learning @ Scale Conference - L@S 2014, pp. 21–30 (2014). doi:10.1145/2556325.2566247
11. Muñoz, P, Méndez, E., Delgado, C.: SPOCs for Remedial Education: Experiences at the Universidad Carlos III de Madrid (2014)
12. Amaya, A., Valles, M.: Beneficios de los MOOC en la Educación Superior. Memorias del encuentro internacional de educación a distancia, vol. 4. Universidad de Guadalajara (2015)
13. Israel, M.: Effectiveness of integrating MOOCs in traditional classrooms for undergraduate students. Int. Rev. Res. Open Distrib. Learn. **6**(5), 102–118 (2015)
14. Kellogg, S., Edelmann, A.: Massively Open Online Course for Educators (MOOC-Ed) network dataset. Br. J. Educ. Tech. **46**(5), 977–983 (2015). Special Issue: Open Data in Learning Technology
15. Soffer, T., Cohen, A.: Implementation of Tel Aviv University MOOCs in academic curriculum: a pilot study. Int. Rev. Res. Open Distrib. Learn. **16**, 80–97 (2015)
16. Alario-Hoyos, C., Kloos, C.D., Estévez-Ayres, I., Fernández-Panadero, C., Blasco, J., Pastrana, S., Villena-Román, J.: Interactive activities: the key to learning programming with MOOCs. In: Proceedings of the European Stakeholder Summit on Experiences and Best Practices in and Around MOOCs, EMOOCS 2016, p. 319 (2016)
17. Perez-Sanagustín, M., Hilliger, I., Alario-Hoyos, C., Delgado Kloos, C., Rayyan, S.: Describing MOOC-based Hybrid initiatives: The H-MOOC Framework. European MOOCs Stakeholders Summit EMOOCs 2016

# Are You Ready to Collaborate? An Adaptive Measurement of Students' Arguing Skills Before Expecting Them to Learn Together

Chrysi Rapanta(✉)

IFILNOVA, Universidade Nova de Lisboa,
Av. de Berna 26, 1069-061 Lisbon, Portugal
crapanta@fcsh.unl.pt

**Abstract.** This paper describes a novel instrument of assessing adolescent and adult students' perception of arguments structure, nature, and quality as a method of adapting teaching and designing of argumentation tasks to learners' epistemic knowledge of argumentation. The author's goal is to present the steps of the validation of the instrument discussing validity and reliability issues, and to discuss potential uses of the instrument as a way to diagnose students' status of argument quality perception before engaging them in collaborative tasks.

**Keywords:** Argumentation · Assessment · Computer-supported collaborative learning · Preparedness · Quality perception

## 1 Introduction

The link between collaboration and argumentation has been the focus of many studies, which in their total show a mutual quasi-causal relationship between the two. More precisely, argumentation has been considered as a main component of collaborative discussions that lead to learning of taught concepts. At the same time, collaborative dialogues have been shown to "contain" more and better argumentation than other types of interaction do.

More precisely, when students are asked to collaborate in order to solve a problem or arrive at a commonly shared point of view on an ill-defined topic, they engage in a series of discourse activities related to knowledge acquisition and construction. The construction of arguments, either individually or together with peers, is a main part of the process of interacting with other learners (Andriessen et al. 2003). Through argumentative knowledge construction, learning partners acquire knowledge about argumentation as well as knowledge of the content under consideration (Weinberger and Fischer 2006). This twofold approach learning to argue and arguing to learn established the relationship between collaborative learning and argumentation (Muller-Mirza and Perret-Clermont 2009).

In both cases, a gap has been observed in regards to the degree of adaptation of the participant students to the goals of the task, either when this goal is to reach a consensus about an issue or simply to learn together. An effort towards diagnosing learners' preparedness for collaboration tasks, using or not a computer tool, through pre-assessing their level of perception of argument quality is the focus of this paper.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 234–246, 2016.
DOI: 10.1007/978-3-319-45153-4_18

## 2 Literature Review

According to Jonassen and Grabowski (1993), the goal and need for adaptive instruction lies on three main assumptions: (a) that learning outcomes may be taught in many ways, (b) individuals will respond to different forms of instruction in different ways, and (c) learning outcomes are affected by the form of instruction. Moreover, adaptive learning has two main characteristics: (a) it can be performed in a number of equally valid and effective ways, and (b) the various functions can be initiated by either the instructional agent (e.g. teacher, textbook, computer, etc.) or by the student (Schuell 1992). Similarly, teaching-learning experiences are adaptive when they allow learners to initiate the learning functions by themselves (Schuell 1992). In the case of collaborative learning situations, either supported by computers or not, a great part of accomplishing the goals of collaboration discussed in the Introduction implies students' preparedness to perceive the task(s) and/or function(s) of arguing together in order to learn.

Some examples of lack of adaptivity between learners and tasks that require argumentation and collaboration emerge as secondary findings or considerations in the literature. From a Computer-Supported Collaborative Learning (CSCL) perspective, concerns are raised regarding issues of interactivity with the system, validity of methods for measuring the quality and nature of students' contributions, as well as the quantity of steps or moves regarding teachers' support during scaffolding (Clark et al. 2007). In general, when students are asked to engage in argumentative knowledge construction, they are expected to perform at least three kinds of epistemic activities: (a) to construct the problem space, through evaluating and relating single components of information regarding the issue; (b) to construct the conceptual space, through distinguishing concepts from each other, and (c) to construct the relations between the two spaces, through applying knowledge adequately and relating theoretical concepts to case information (Weinberger and Fischer 2006). Nonetheless, such epistemic operations do not always take place, as it can been shown from studies reporting low collaborative learning outcomes, even among adolescents and adults (Koschmann 2003).

From an educational psychology point of view, main considerations concern the adaptivity of instruments for assessing students' argumentation quality according to the participants' age and gender, and to the goal of the argumentation task. Regarding the latter, different goal prescriptions have been found to have different argumentation and learning outcomes. As Felton et al. (2015) overview, college students are more likely to cite arguments that originate from their peers and more likely to integrate arguments with counterarguments when they are asked to reach consensus in a chat rather than when the goal instruction is to persuade their peer. Similarly, argumentative discourse goals may have an impact not only on the quality of participants' arguments but also on their content learning. Regarding adolescents and adults another crucial factor influencing on their level of argumentation, and subsequently collaboration, is their awareness of the epistemological norms of argumentation (e.g. Weinstock et al. 2004) or, simply put, their level of preparedness to learn from argumentation (Duschl and Osborne 2002).

For the reasons mentioned above, several instruments have been thus far developed to assess students' argumentation skills before engaging them in an argumentative knowledge construction task. The majority of these instruments assess students' general epistemological understanding and beliefs about knowledge and knowing (see Mason and Scirica 2006, for an overview). Other studies have focused on the explicit epistemic norms of argumentation developing tasks of identifying fallacies as a way of assessing students informal reasoning skills (see Rapanta and Macagno 2016, for an overview). A common drawback of the existing assessment methods is that they tend to focus on the characteristics of the learners or the task leaving out the possibilities of comparison, re-adaptation, and re-use of the same in different contexts.

The present paper focuses on the generic skill of argument quality perception as the main layer for other skills to emerge and develop as result of collaborative argumentation tasks. Argument quality perception refers to the capacity of learners to perceive the quality of different arguments through identifying main argument elements and through producing parts of key argumentation schemes. A main criterion in both cases is that of relevance, due to its strong relation with the epistemic operations implied in argumentative knowledge construction as described in the Introduction.

## 3 Goal

This paper presents an instrument of assessing adolescent and adult students' perception of arguments structure, nature, and quality as a method of adapting teaching and designing of argumentation tasks to learners' epistemic knowledge of argumentation. The goal is twofold: first, to present the steps of the validation of the instrument discussing validity and reliability issues; second, to discuss potential uses of the instrument as a way to diagnose students' status of argument quality perception before engaging them in collaborative tasks.

## 4 Method

### 4.1 Participants

The participants were 80 University students in a public University in Lisbon area, Portugal. Fifteen of them were Masters' students whereas the rest 65 were undergraduates in the Faculty of Humanities and Social sciences; regarding gender, 23.75 % were males and the rest females. The average age was 21.4 years old. The big majority (95 %) were Portuguese. All participants voluntarily accepted to complete the questionnaire using a hard copy distributed to them by their instructors in three different classrooms at the beginning of the Spring semester of 2016.

### 4.2 Variables

The main variable of the present study is the variable defined as argument quality perception. To construct this variable, we draw on two main assumptions: first, that the capacity to argue in educational contexts is based on two different types of skills, namely production and interpretation of arguments (Rapanta et al. 2013); second, that relevance is an umbrella concept including the other two main argument assessment criteria (i.e. sufficiency and acceptability) as recently proposed by Macagno and Walton (under review). Considering the above, the following sub-variables emerge:

- Identify argument elements
- Judge on the relevance of different argument elements
- Produce relevant arguments

### 4.3 Instrument

The instrument was composed of 12 items, separated into three parts accordingly to the three variables mentioned above. For the first variable (i.e. Identify argument elements), we used a paragraph adapted from Stab and Gurevych (2014), on the everyday topic of pros and cons of living abroad. The paragraph was fairly short (9 lines long), with clear structure, and written in plain English, which was adequately translated into Portuguese by a native speaker (the whole instrument was translated). The argument elements, which the students were asked to identify, were: reason, evidence, counter-argument, and conclusion. By "reason" we mean the main premise on which the authors is based to support her claim (*Living and studying overseas is an irreplaceable experience*) in the paragraph. Evidence corresponds to the scientific data mentioned by the author in her effort to convince the readers about her opinion (*A study among Erasmus students showed that 93 % of young people who study abroad for the first time in their lives feel more capable of dealing with any type of problems than they were feeling before leaving their homes*). Counter-argument refers to the integrated contrary opinion that an opponent might have (*One who is living overseas will of course struggle with loneliness, living away from family and friends*). Finally the conclusion is the idea to which the author arrives after weighing the pros and the cons of the issue (*Being independent is more important than any difficulties*).

The second part of the questionnaire (Judge on the relevance of different argument elements) contained four binary items on which the participants had to decide about one being a stronger argument than the other. Three of the four items were adapted from Larson et al. (2009) whereas the fourth item was originally used in the study of Goldstein et al. (2009). More precisely, item Q5 represents a simple informal argument structure between a claim and a relevant premise, whereas items Q6 and Q7 emphasized on the role of claim predicates in determining relevance. Finally, the fourth item in this section (item Q8) required for the distinction between a valid justification (How do you know?) and an explanation (What do you mean?), which is a common theme in several studies (e.g. Brem and Rips 2000; Kuhn 2001).

The final part of the questionnaire (Produce relevant argument components) was constructed by the authors and it included four incomplete arguments, each one referring to a type of argumentation scheme. Item Q9 referred to argument from expert opinion, items Q10 and Q12 to argument from negative consequences, and item Q11 to argument from positive consequences. The three used argumentation schemes are presented in Table 1, whereas all the items of the questionnaire are found in the Appendix.

**Table 1.** Argumentation schemes from expert opinion and from positive/negative consequences (Walton et al. 2008).

| Argumentation schemes | From expert opinion | From positive/negative consequences |
|---|---|---|
| Major premise | Source E is an expert in subject domain S containing proposition A. | If A is brought about, consequence *a* will occur. |
| Minor premise | E asserts that proposition A is true/false | Consequence *a* is probably good/bad. |
| Conclusion | A is true (false). | Therefore I should/shouldn't do A. |

Eight of the items (Q1–Q8) were assessed as right/wrong answers and the last four as highly, medium, and poorly relevant as further explained in Sect. 5.3.

# 5 Findings

## 5.1 Factor and Items Analysis

First we performed an exploratory factor analysis (EFA) to determine potential components or latent variables as emerged from the responses received at this pilot phase of the study. Data were subjected to factor analysis using Principal Axis Factoring and orthogonal Varimax variation. During the EFA, we obtained two negative measures: (a) the Kaiser-Meyer-Olkin measure (KMO) was below 0.5, and (b) more than 50 % (71.0 %) of the nonredundant residuals after the extraction of components had an absolute value greater than 0.05. These were both indications that the sample was not adequate for extracting a factors' model with a good fit as explained by Yong and Pearce (2013). However, this initial factor analysis gave us a good approximation of which items were highly and positively correlated with the three factors emerged. Items 3, 5 and 12 had a negative correlation with the survey components as shown on Table 2.

Considering the negative correlation of items Q3, Q5, and Q12 and the fact that inverting them was not the case (all of the questions were positive), we decided to exclude these three items. We then looked at possible reasons for such negative correlation and we assumed that this was either due to high correct responses level (Q3 and Q12) or to high similarity among items. Regarding the latter, we identified a high similarity in the answers' pattern for items Q5 and Q6, which contributed into our decision to also exclude Q6. We then calculated the scale reliability of the remaining items, and we further discovered that: (a) the correlation of items Q8 and Q3 to the scale was negative, and (b) the exclusion of items Q2 would render a higher reliability. Excluding these additional elements, we re-calculated the scale reliability including

**Table 2.** Initial exploratory factor analysis results

| | Components | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Q11 | .67 | | |
| Q1 | .63 | | |
| Q4 | .63 | | |
| Q5 | –.45 | | |
| Q2 | .43 | | |
| Q7 | | .68 | |
| Q9 | | .66 | |
| Q10 | | .55 | |
| Q8 | | | .67 |
| Q12 | | | –.53 |
| Q3 | | | –.48 |
| Q6 | | | .43 |

only six items: Q1, Q4, Q7, Q9, Q10, and Q11. The Cronbach's alpha was 0.48, which may be considered unacceptable for a developing questionnaire that needs to exceed 0.70 (Rattray and Jones 2007). However, this medium scale reliability may also be considered as a positive indication of heterogeneity, meaning that different trains of the same skill are measured in the same test (Alderson et al. 1995). This might be true if we also consider the complex nature of argumentative competence as commented elsewhere (Rapanta et al. 2013). However, reaching above 0.60 is a reasonable goal for the internal consistency to be achieved in a subsequent version of the scale.

A second factor analysis was performed using only the six remaining items. The sampling adequacy measures were better than the ones obtained in the first EFA: KMO was above 0.5 and the Bartlett's test of Sphericity was significant at .005. Given that our sample was smaller than 100 participants, which is considered for many authors as the minimum for a factor analysis sample adequacy (Rattray and Jones 2007), the numbers obtained from the tests were acceptable. The principal components analysis yielded two factors with a cumulative variance of 52.2 %. Table 3 shows the factor loadings after rotation using a significant factor criterion of .4.

**Table 3.** Final exploratory factor analysis results

| | Factor 1 | Factor 2 |
|---|---|---|
| Q4 | .74 | |
| Q1 | .70 | |
| Q11 | .68 | |
| Q9 | | .76 |
| Q7 | | .76 |
| Q10 | | .56 |
| Eigenvalues | 1.60 | 1.53 |
| % of variance | 26.73 | 25.5 |

## 5.2 Initial Descriptive Findings

Regarding the first part of the questionnaire (identifying argument elements from a text), students showed a medium to high difficulty in identifying the main reason (29.1 % got it wrong), the evidence in support of the reason (21.3 % wrong answers), the counter-argument (30.8 %), and the conclusion (43 %). In the second part of items focusing on identifying the stronger version of an argument, it was fairly easy for the participants to identify a pertinent reason for a claim as 88.8 % of them got questions 5 and 6 right. However, question 7, which was again about pertinence, yielded different results, with more than half participants (52.5 %) responding wrongly. This failure might be related to the complexity of the reason given in the example of Q7, "DNA has been used to prove that many sentenced to death were innocent". The evidential weight is on the use of DNA and not on the fact that people were innocent, thus the correct claim-conclusion is the one characterizing the death penalty as "immoral" and not the one calling it "ineffective". Q8 also received a great number of wrong answers, with 67.5 % of the respondents not being able to distinguish an explanation from a justification. Table 4 shows the valid percentages of right and wrong answers for each one of the items Q1 to Q8.

**Table 4.** Percent frequencies of right and wrong answers for items Q1–Q8.

| | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 |
|---|---|---|---|---|---|---|---|---|
| % Right | 70.9 | 78.8 | 69.2 | 57.0 | 88.8 | 88.8 | 47.5 | 32.5 |
| % Wrong | 29.1 | 21.3 | 30.8 | 43.0 | 11.3 | 11.3 | 52.5 | 67.5 |

## 5.3 Relevance of Arguments Produced

A special attention needs to be given to the last four items of the questionnaire, regarding the filling-in of incomplete arguments in order for them to be meaningful. The construction of the items was based on the concept of argumentation schemes, as they are considered an adequate method for judging on the validity of the majority of arguments produced in everyday contexts (Walton et al. 2008; Rapanta and Walton 2016a). The elements produced as part of an argumentation scheme can be more or less relevant regarding the number of inferential steps an external judge should make in order to pass from one element to another (Macagno and Walton, under review). Two raters (the author and an external rater) have assessed the sentences (premises or conclusions) that the participants came up with to fill in the incomplete arguments (see Appendix, items Q9–Q12). The assessment was based on the valid form of the corresponding argumentation schemes (see Table 1) and the relevance-based distance rating from 1 as very close to an argument that makes sense from an argumentation point of view to 3 as the most irrelevant. The inter-rater reliability calculation gave a high percent of agreement (89.7 %) and an acceptable Krippendorff's Alpha of 0.76. The average of the scores obtained in items Q9–Q12 (producing relevant arguments) yielded a mean score of 1.4 with 29.9 % of the participants producing strongly relevant arguments as Table 5 shows.

**Table 5.** Frequencies, means and standard deviations of the average scores in the argument production task (items Q9–Q12).

| Average score on items Q9–Q12 | | Frequency | Percent | Valid percent |
|---|---|---|---|---|
| Valid | 1.00 | 20 | 25.0 | 29.9 |
| | 1.25 | 8 | 10.0 | 11.9 |
| | 1.50 | 20 | 25.0 | 29.9 |
| | 1.75 | 15 | 18.8 | 22.4 |
| | 2.00 | 4 | 5.0 | 6.0 |
| Missing | | 13 | 16.3 | |
| Total | | 80 | 100 | 100 |
| Mean | 1.4 | | | |
| St. Deviation | 0.32 | | | |

## 6 Discussion

A common effort among argumentation and collaborative learning scholars focuses on how to bridge the gap between assessment and practice in what regards the quality of peer interactions oriented towards the argumentative construction of knowledge. Research shows that the design and scaffolding of the task may influence significantly on the type and quality of argumentation (e.g. Clark et al. 2007; Felton et al. 2015); however, how the preparedness of students in regards to learn together influences on the poor quality of argumentation often emerging in collaborative settings has been understudied.

The present paper introduces a variable related to a generic arguing skill, namely the capacity to perceive the quality of arguments produced by other and by oneself. A 12-item questionnaire was constructed with three types of tasks, including the identification of argument elements from a text, the distinction between relevant and irrelevant argument components, and the production of premises or conclusions that make sense in an argumentation scheme context. The two exploratory principal components analyses yielded two factors, each one including items from different tasks. More precisely, Factor 1 comprised two argument element identification items (Q1 and Q4) and one argumentation scheme production (Q11), whereas Factor 2 covered two argumentation scheme items (Q9 and Q10) together with an item from the relevant reasons identification task (Q7). Regarding the use of the questionnaire as a method of diagnosing students preparedness to argue effectively in simple collaboration tasks, primary analysis shows that the level of argument quality perception of young educated adults is fairly low. This result agrees with other studies reporting reasoning flaws in simple argumentation tasks engaging university students (e.g. Rapanta and Walton 2016b).

More precisely, the present study revealed a series of weaknesses in identifying and constructing arguments among both undergraduate and graduate students, including: the identification of counter-arguments and conclusion, the assessment of complex reasons, and the production of relevant premises. Based on these findings, subsequent tasks including collaborative argument activities may be adapted to the level of epistemic preparedness of the students in the following ways: (a) to help students be able to identify relevant elements in a text before engaging in an argumentative interaction, some guiding questions or text highlights may be used; (b) to increase the capacity of reason assessment, students may be confronted with various pieces of information that may be used for evidence in support of their own or their peers point of view; a task of combining right pieces of evidence with the most adequate theory before engaging in argumentative dialogue may be useful as other studies have shown (e.g. Berland and Reiser 2011); (c) finally, to increase the possibility that students produce valid arguments when engaging in discussion with each other, some prior exercises with the use and application of argumentation schemes either in the form of a map (Rapanta and Walton 2016a) or through matching them with appropriate critical questions (Walton et al. 2008) may be helpful.

In sum, this paper showed that even basic skills of argument quality perception should not be taken for granted when argumentation tasks are designed. An instrument akin to reveal the main skills, and subsequently flaws, of argument identification and production may be used as a diagnostic method and basis for the setting of activities that require dialogical argument skills. An understanding of the importance of developing the argument epistemic skill is necessary for instructors to be able to design collaborative activities adapted to participants' level of preparedness to argue in a more or less skilled way. Moreover, an assessment instrument like the one presented here can also be used as a method of pre and post task comparison of student' general capacity to argue. Although the full complexity of argumentative competence cannot be grasped into one simple measurement, identifying which skills may be more related to the quality of argumentative performance in collaboration tasks is possible through the instrument presented here. Future testing of its actual implementation as a pre-post assessment method will further confirm not only the instrument's validity but also the mutual relation between argumentation and collaboration.

Our next step will be to complete the questionnaire's reliability assessment through its re-distribution to the same population. Moreover, more participants will be included from different age and education backgrounds to be able to confirm our assumption that the argument quality perception skill is generic, meaning that it is not limited to specific age groups or subject domains. The use of argumentation schemes as a baseline for such type of assessment seems appropriate, as the current findings have shown. Identifying arguing profiles based on the ability of people to identify and complete valid and relevant argumentation schemes will be the primary outcome of this type of research. A subsequent matching of complementary profiles among participants and the scaffolding of one type of skill at the time will be the contribution of the proposed diagnosing method in orchestrating more successful collaborative tasks from an argumentative knowledge construction point of view.

## 7 Conclusion

The present paper was based on the already proven relationship between argumentation and collaborative learning as discussed elsewhere (e.g. Nussbaum 2008). Under the assumption that if learners are expected to collaborate they are also expected to engage in argumentative knowledge construction, the proposal of an assessment instrument of learners' perception of argument quality was made. The paper described the pilot phase of a study in progress in which learners' pre-assessment based on the instrument presented here will be used to further evaluate their preparedness to engage in collaborative argumentation. Issues of reliability and validity of the proposed instrument as well as some initial descriptive findings of the participants' level of argument quality perception skill were presented. Future research will further validate the instrument as well as its use as a diagnostic method of students' capacity to learn together in both oral and written tasks.

## Appendix

**Dear Student:**

In the margins of a research in argumentation in higher education, we are conducting this small "exam" in order to see your current ways of reasoning about everyday issues. The goal of this "exam" is for us to understand what are some major difficulties which Portuguese pre-graduates face when they deal with simple arguments. Current education systems all over the world require for all University graduates to be critical thinkers, no matter what is their disciplinary area. Your answers will help us understand how far or near we as educators are from this goal. Please dedicate the necessary time for your answers to be more complete and well-thought possible. We thank you in advance for your attention.

**Gender:**
**Age:**
**Nationality:**

**(A) Read the following paragraph carefully and answer the questions that follow based on the text.**

Living and studying overseas is an irreplaceable experience when it comes to learn standing on your own feet. One who is living overseas will of course struggle with loneliness, living away from family and friends, but those difficulties will turn into valuable experiences in the following steps of life. Mainly, she will learn how to be independent and self-motivated. A study among Erasmus students showed that 93 % of young people who study abroad for the first time in their lives feel more capable of dealing with any type of problems (administrative, practical, including personal) than they were feeling before leaving their homes. At the end of the day, being independent is what matters most in the life, isn't it?

1. What is the author's main reason to believe that a?
   ………………………………………………………………………………………
2. Where is (s)he based on to believe that this is the main reason for a?
   ………………………………………………………………………………………
3. What is main counter-argument that opposes to his/her belief?
   ………………………………………………………………………………………
4. What is the author's conclusion?
   ………………………………………………………………………………………

**(B) From every set of sentences, circle the option that you think represents a stronger argument.**

5a. Handguns encourage criminal behavior, so handguns should be banned.
5b. Ninety percent of handgun purchases are now subject to instant FBI criminal background checks, so handguns should be banned.
6a. Recycling is very beneficial because it helps protect the environment.
6b. Recycling is cost-effective because it helps protect the environment.
7a. The death penalty is immoral because DNA has been used to prove that many innocent people have been sentenced to death.
7b. The death penalty is ineffective because DNA has been used to prove that many sentenced to death were innocent.
8. Why do teenagers start smoking? Which is the strongest argument?
8a. Smith says it's because they see ads that make smoking look attractive. A good-looking guy in neat clothes with a cigarette in his mouth is someone you would like to be like.
8b. Jones says it's because they see ads that make smoking look attractive. When cigarette ads were banned from TV, smoking went down.

**(C) Complete the following blanks with a sentence that you think is appropriate for the argument to make meaning.**

9. Professor Coleman is an experienced scientist in earthquakes. He predicted that a big earthquake is going to take place in the northern part of Portugal towards the end of this year. ……………………………………………………………….. Therefore, it is very possible that Professor Coleman is true.
10. You have been saying for years that you want to lose weight. Chocolate is very bad for your health plus it has a lot of empty calories. ……………………………………………………………………………………… Therefore, you shouldn't eat chocolate every day.
11. Studying hard for the final exams increases the possibility of success. The grade of the final exam counts a lot for the final grade. Therefore, ……………………… ……………………………………………………………………
12. I had a cousin who died from drugs abuse, and she was very young. …………… ……………………..…………………………………………..…………………So I suppose I shouldn't smoke marijuana.

## References


Alderson, J.Ch., Clapham, C., Wall, D.: Language Test Construction and Evaluation. Cambridge University, Press Cambridge (1995)

Andriessen, J., Baker, M., Suthers, D.: Argumentation, computer support, and the educational context of confronting cognitions. In: Andriessen, J., Baker, M., Suthers, D. (eds.) Arguing to Learn: Confronting cognitions in Computer-Supported Collaborative Learning Environments, pp. 1–25. Springer, Amsterdam (2003)

Berland, L.K., Reiser, B.J.: Classroom communities' adaptations of the practice of scientific argumentation. Sci. Educ. **95**(2), 191–216 (2011). doi:10.1002/sce.20420

Brem, S.K., Rips, L.J.: Explanation and evidence in informal argument. Cogn. Sci. **24**(4), 573–604 (2000). doi:10.1016/S0364-0213(00)00033-1

Clark, D.B., Stegmann, K., Weinberger, A., Menekse, M., Erkens, G.: Technology-enhanced learning environments to support students' argumentation. In: Erduran, S., Jiménez-Aleixandre, M.-P. (eds.) Argumentation in Science Education, pp. 217–243. Springer, Amsterdam (2007)

Duschl, R.A., Osborne, J.: Supporting and promoting argumentation discourse in science education. Stud. Sci. Educ. **38**, 39–72 (2002). doi:10.1080/03057260208560187

Felton, M., Garcia-Mila, M., Villarroel, C., Gilabert, S.: Arguing collaboratively: argumentative discourse types and their potential for knowledge building. Br. J. Educ. Psychol. **85**(3), 372–386 (2015). doi:10.1111/bjep.12078

Goldstein, M., Crowell, A., Kuhn, D.: What constitutes skilled argumentation and how does it develop? Informal Logic **29**(4), 379–395 (2009)

Jonassen, D.H., Grabowski, B.L.: Handbook of Individual Differences, Learning, and Instruction. Lawrence Erlbaum Associates, Hillsdale (1993)

Koschmann, T.: CSCL, argumentation, and Deweyan inquiry. In: Andriessen, J., Baker, M., Suthers, D. (eds.) Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments, pp. 261–269. Springer, Amsterdam (2003)

Kuhn, D.: How do people know? Psychol. Sci. **12**(1), 1–8 (2001). doi:10.1111/1467-9280.00302

Larson, A.A., Britt, M.A., Kurby, C.A.: Improving students' evaluation of informal arguments. J. Exp. Educ. **77**(4), 339–366 (2009). http://www.ncbi.nlm.nih.gov/pmc/articles/pmc2823078/

Mason, L., Scirica, F.: Prediction of students' argumentation skills about controversial topics by epistemological understanding. Learn. Instr. **16**(5), 492–509 (2006). doi:10.1016/j.learninstruc.2006.09.007

Muller-Mirza, N., Perret-Clermont, A.-N. (eds.): Argumentation and Education: Theoretical Foundations and Practices. Springer, New York (2009)

Nussbaum, E.M.: Collaborative discourse, argumentation, and learning: preface and literature review. Contemp. Educ. Psychol. **33**(3), 345–359 (2008). doi:10.1016/j.cedpsych.2008.06.001

Rapanta, C., Garcia-Mila, M., Gilabert, S.: What is meant by argumentative competence? an integrative review of methods of analysis and assessment in education. Rev. Educ. Res. **83**(4), 483–520 (2013). doi:10.3102/0034654313487606

Rapanta, C., Macagno, F.: Argumentation methods in educational contexts. Introduction to the special issue. Int. J. Educ. Res. (2016) doi:10.1016/j.ijer.2016.03.006

Rapanta, C., Walton, D.: The use of argument maps as an assessment tool in higher education. Int. J. Educ. Res. (2016a). doi:10.1016/j.ijer.2016.03.002

Rapanta, C., Walton, D.:. Identifying paralogisms in two ethnically different contexts at university level. Infancia y Aprendizaje: Journal for the Study of Education and Development **39**(1), 119–149 (2016b)

Rattray, J., Jones, M.C.: Essential elements of questionnaire design and development. J. Clin. Nurs. **16**, 234–243 (2007)

Schuell, T.J.: Designing instructional computing systems for meaningful learning. In: Jones, M., Winne, P.H. (eds.) Adaptive Learning Environments: Foundations and Frontiers, pp. 19–54. Springer-Verlag, Heidelberg (1992)

Stab, Ch., Gurevych, I.: Annotating argument components and relations in persuasive essays. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 1501–1510, Dublin Ireland, 23–29 August 2014

Walton, D., Reed, C., Macagno, F.: Argumentation Schemes. Cambridge University Press, New York (2008)

Weinberger, A., Fischer, F.: A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Comput. Educ. **46**(1), 71–95 (2006). doi:10.1016/j.compedu.2005.04.003

Weinstock, M., Neuman, Y., Tabak, I.: Missing the point or missing the norms? epistemological norms as predictors of students' ability to identify fallacious arguments. Contemp. Educ. Psychol. **29**(1), 77–94 (2004). http://doi.org/10.1016/S0361-476X(03)00024-9

Yong, G., Pearce, S.: A beginner's guide to factor analysis: focusing on exploratory factor analysis. Tutorials Quant. Methods Psychol. **9**(2), 79–94 (2013)

# Examining the Effects of Social Media in Co-located Classrooms: A Case Study Based on SpeakUp

María Jesús Rodríguez-Triana(✉), Adrian Holzer, Luis P. Prieto, and Denis Gillet

École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
{maria.rodrigueztriana,adrian.holzer,luis.prieto,denis.gillet}@epfl.ch

**Abstract.** The broad availability of mobile computing devices has prompted the apparition of social media applications that support teaching and learning. However, so far, there is conflicting evidence as to whether the benefits such applications provide in terms of engagement and interaction, outweigh their potential cost as a source of distraction. To help in clarifying these issues, the present paper presents a case study on the impact of using SpeakUp (an app aimed at promoting student participation through anonymous chatrooms) in an authentic face-to-face learning scenario. Concretely, we focus on the connection between SpeakUp and the student engagement, distraction, social interaction, and the influence of the teachers' style. Our findings highlight that SpeakUp favored students' engagement and social interaction, but they also point towards its limitations in keeping students communicating about content relevant to the course.

**Keywords:** Social media · Engagement · Attention · Interaction · Learning · Teaching

## 1 Introduction

Multiple social media applications are appearing to support teaching and learning, leveraging the broad access to mobile devices (e.g., in "bring your own device" approaches). However, there is conflicting evidence on whether the use of mobile technologies in the classroom is positive (e.g., improving student participation) [33] or negative (e.g. distracting students due to multitasking) [28,35].

In this context, we are interested in studying how to use social media effectively in the classroom. This paper focuses on the pedagogical use of SpeakUp, a mobile app aimed to promote student participation in face-to-face sessions. In SpeakUp, students can anonymously join chatrooms, post messages and vote on them.

Since the mere introduction of social media in educational contexts does not ensure a positive effect, this paper analyses the impact of SpeakUp in an authentic learning scenario carried out with first-year (bachelor) university students.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 247–262, 2016.
DOI: 10.1007/978-3-319-45153-4_19

In particular, this paper explores the following research question: *does SpeakUp favor situations that lead to learning?* To answer this question, we structured the study according to the following topics: active participation [25] (i.e., engagement), attention [17] (i.e., remaining on-task), and social interaction [6] (on relevant content).

The CSCL-EREM framework [19] guided the formalization of this case study, as well as the data gathering and analyses, leading us to use multiple informants (students, teachers, researchers and the technology used), different data gathering techniques (observations, questionnaires, SpeakUp logs, and user comments in the app), and mixed methods analyses, including: student attendance to the session, teacher and student participation face-to-face and via SpeakUp, content of the comments, as well as teacher and student perceptions about the impact on engagement, attention and interaction.

The paper is structured as follows: Sect. 2 reviews previous research on the usage of social media for educational purposes; Sect. 3 introduces the SpeakUp app and its main functionalities; Sect. 4 describes the research methodology followed in the present case study, while Sect. 5 details the main results of the data analyses and that are later discussed in Sect. 6 together with the main conclusions and the future work.

## 2 Related Work

**Historical Overview**. Social interaction in the classroom is considered by numerous researchers as a *conditio sine qua non* for learning [8,21]. Providing learners with a digital channel for interaction can be traced back to the 80's when IBM started to experiment with student interaction systems [16]. Many of these systems are based on reactive interaction where teachers can conduct live polling by asking multiple choice questions and students answer by pressing a button on a clicker. Studies on clickers show that they can foster more participation in the classroom, and that students generally have a positive attitude towards them (e.g., [3,9,32,34]). On top of the reactive channel, some systems provide a proactive channel, where students can post questions and comments. With the rise of mobile devices, systems also started relying on the students' own devices. An early effort in this direction was the TXT-2-LRN [29] mobile system, with which students could send free-form SMSs to the teachers.

**Students' Perceptions**. More recently, systems also include a social media layer, where students can vote and comment on each other's contributions (e.g., ClassCommons [7], Fragmented Social Mirror [2], Pigeonhole Live [11], Backchan.nl [13], or SpeakUp [14,15]). Mainstream social media, such as Twitter [26,27] and Reddit, are also popular when attempting to foster interaction between speakers and their audience in both conferences and classrooms. Research investigating the use of such social media applications in the classroom generally concludes that students perceive such systems as positive and that they feel it increases interactivity [1,2,9,13–15,29]. Furthermore, students often prefer to use a digital channel to interact instead of raising their hand [29].

**Teachers' Opinions**. The Pearson education service company conducted a survey with 7969 U.S. higher education teachers to better understand the bigger picture of the social media usage by teachers [30]. The survey finds that teachers are generally aware of social media and they are using it in their private lives (70.3 % of faculty use it at least once per month). The use of social media in the classroom lags behind the usage in their personal lives (41.0 %) but is increasing every year. Teachers see social media and technology as having a "considerable potential" for learning. However, 56.0 % of teachers also consider that social media in class can be more distracting than helpful.

**Potential Shortcomings**. The issue of distraction and multitasking in education is receiving increased attention, with conflicting results so far. Certain research suggests that laptop multitasking hinders learning for both users and nearby peers [28], and that providing slides to students can affect performance adversely [22,35]. On the other hand, researchers also argue that it is possible to take advantage of social media in the classroom by embracing multitasking, which students seem to able to effectively do in the classroom [20,36]. A recent meta-analysis on the use of mobile devices in the classroom nuances these claims and shows a moderate positive learning effects [33].

This paper aims at better understanding whether, and under what circumstances, social media usage in the classroom may have a positive impact.

## 3 SpeakUp

SpeakUp is a social media app designed to foster participation in co-located situations where such interaction is difficult, either within the audience or between the speaker and the audience (e.g., a university lecture with a large number of students, or a conference). In a typical usage scenario with SpeakUp, teachers create a chatroom that students can join by typing its number as shown in Fig. 1.1. Note that any user can join such rooms without login or registration (enabling an immediate use of the app).

Inside the chatroom, any user can post text messages, comment on existing messages, and vote them (up or down, see Fig. 1.3). Each message has a score, which shows the difference between the number of upvotes and downvotes. For instance, the top message in Fig. 1.3 has a score of −1 and the bottom message a score of +3. The chatroom creator, i.e., the teacher, can create multiple choice messages (Fig. 1.2) for students to answer. Inside the chatroom messages are sorted either by time or by score.

Furthermore, in the chatroom all users are anonymous, thus fostering the expression of more uninhibited points of view. This implies that users interact, not directly with one another, but rather on the basis of the content posted by the different anonymous users.

Classroom interaction in a lesson using SpeakUp can occur along the face-to-face (f2f) channel (i.e., teachers and students interacting orally), as well as along a digital channel (i.e., posting comments and voting on SpeakUp). There can also be transitions from one channel to the other. For instance, a teacher can

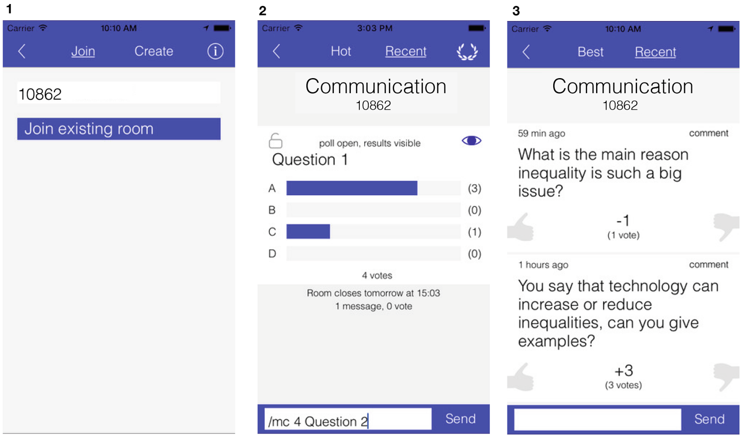



**Fig. 1.** Screenshots of the SpeakUp mobile app. (1) joining a chatroom. (2) creating a multiple choice question in the chatroom. (3) viewing messages in the chatroom ordered by time or score.

instruct students to answer a poll on SpeakUp, or explicitly ask students to post messages on SpeakUp. Conversely, questions posed by students on SpeakUp, can be answered by teachers orally.

## 4 Methodology

The present study is framed within a wider research effort whose general goal is to understand how social media can be used effectively in the classroom. Towards this aim, several exploratory studies have been performed in the past on the use of SpeakUp in classrooms [14, 15], in which SpeakUp was deemed easy to use, and motivating for students to participate more in lectures. In turn, the present study is the first of a series in which we aim to *evaluate* the effectiveness of SpeakUp to foster learning in more details using authentic educational settings [5]. We also aim to assess its potential role in the distraction of students, and its relations with various teaching strategies and styles.

This concern with deep evaluation of a social learning tool has led us to use a case study methodology [31], structured using the Computer Supported Collaborative Learning – Evaluand Oriented Responsive Evaluation Model (CSCL-EREM, see [19]) framework. This framework was designed specifically to evaluate the impact of TEL interventions, especially in authentic settings. Hereafter, we discuss the research issue and topic, the data sources, and the data analyses (see Fig. 2).

**Research Issue and Topics.** Guided by this framework, we organised the perspective of the case study around the definition of an *issue*. An issue can be understood as a troubling choice, a tension, an organizational perplexity or a problem. In this case study, the main issue is defined as: *does SpeakUp favor situations that*

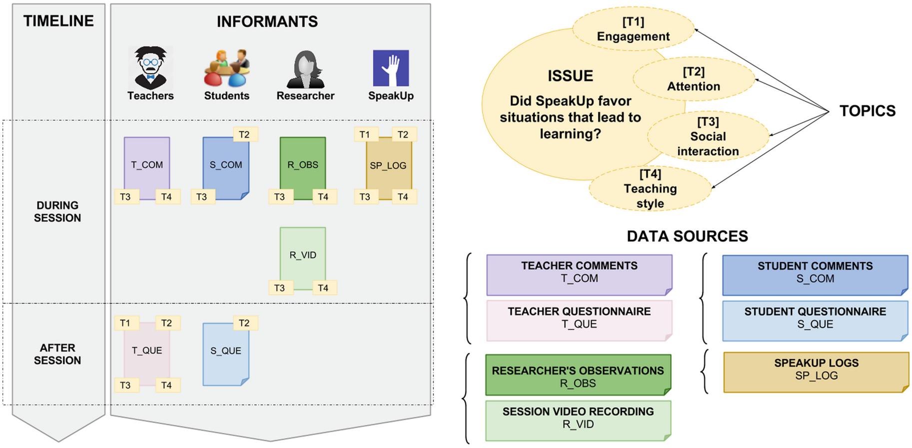



**Fig. 2.** Diagram representing the issues, topics, data sources and informants used in the case study

*lead to learning*, such as active participation (i.e., engagement), attention (i.e., focus on-task), and social interaction (on relevant content)? Then, following an anticipated data reduction procedure (common in qualitative data analysis [24]), this issue is illuminated by answering a number of informative questions, clustered around four topics (see Fig. 2). These topics are related to the users' active participation and engagement with the tool (T1), its effects on attention (i.e., focusing vs. distracting from the lesson topic, T2), and on the classroom social interactions (T3). Finally, another topic explores the interactions of teacher actions and style with the different kinds of SpeakUp usage (T4).

**Data Gathering, Informants and Data Sources.** We use a mixed method approach [4, 10] combining quantitative and qualitative data coming from four types of informants (two teachers, 145 students, one observer, plus the SpeakUp logs) using different data gathering techniques: questionnaires, logfile analysis, observations, video recordings, and student contributions in SpeakUp. This mixed methods approach is commonly used in TEL research [18] and promoted by the CSCL-EREM methodology in order to obtain different perspectives about the evaluand (the object of the evaluation, in our case the use of SpeakUp in the lesson), thus enriching the evaluation process.

**Data Analysis.** Different quantitative (descriptive statistics and exploratory computational analyses) and qualitative analyses (manual coding of the messages generated by the users, see below) have been performed on the data. Then, the results from these analyses were *triangulated* [12] to increase the trustworthiness of our findings.

In order to better understand the aforementioned aspects of engagement, attention and social interaction, we manually coded all the messages and comments generated during the lesson, into two main categories: Messages that are *relevant* for learning and messages not relevant for learning, similarly to previous studies on SpeakUp [14]. We further divided these main categories in four sub-categories, inspired by those proposed by McCarthy [23]: the relevant message were divided into content related messages (i.e., questions or comment about the content of the course), organisation related messages (i.e., messages related to team and course organisation), SpeakUp related messages (i.e., messages discussing SpeakUp itself) and miscellaneous messages (i.e., messages such as greetings and policing). Non-relevant messages were also divided into content-related (i.e., messages that discuss course content but are not relevant to learning), SpeakUp related (i.e., not relevant messages related to the use of SpeakUp) and miscelleanous messages. We also added a social message category (i.e., non-relevant messages about people) and a bullying message category (i.e., non-relevant messages with negative social connotations). Figure 3 shows examples of messages in each category. Furthermore, each message was also labeled as comment, answer or question, and tags were also added about the direction of the interaction: students to teachers, students to students, students to all, and teachers to students.

| **Relevancy** | **Category** | **#** | **Examples** | |
|---|---|---|---|---|
| **Relevant** | Content | *42* | *I don't remember anything about probabilities* | *Bays Formula* |
| | SpeakUp | *34* | *Are we anonymous here?* | *This app is ruining my battery* |
| | Organisation | *17* | *Do they know that there is only 10 min left in the class?* | *Looking for a group* |
| | Misc | *13* | *Hello* | *some respect please!* |
| **Non relevant** | Content | *60* | *earth is flat* | *LOVE CNN* |
| | SpeakUp | *17* | *If I get 100 likes, I will take off my clothes* | *Telegram channel : comA ?* |
| | Social | *83* | *He endorses the socialist party's initiative* | *I m sure he's French* |
| | Misc | *59* | *Let's ask Jawed* | *IamYourMother* |
| | Bullying | *9* | *homo* | *answer of a blond* |

**Fig. 3.** Examples of SpeakUp message categories.

In a similar way, and in order to understand these topics as they occurred in the face-to-face channel of the classroom, the video recording of the lesson was also coded, according to the following categories: Which *actor* was speaking at each moment during the lesson (e.g., each of the three teachers present, or one of the students); what action was being performed at that moment (e.g., presentation/lecturing, asking questions, providing answers, noting technical or other kinds of problems); who was the target of the interaction, if any (e.g., a teacher, students, or all the class); and finally, what supporting resources were being used, if any (e.g., slides, videos, SpeakUp).

# 5 Case Study

The different quantitative and qualitative sources detailed in Sect. 4 were analysed and triangulated to illuminate the issue and topics addressed in the case study. This section presents the results obtained after presenting the context in more details.

## 5.1 Context

The case study took place in the first lecture of a Communication course at the École Polytechnique Fédérale de Lausanne in Switzerland. In this, which lasted for 90 min, 145 students (38 female) were present. This Communication course, which discusses different kinds of communication channels, social media platforms and technology-enhanced learning, is part of the Global Issues program, which aims at introducing first-year undergraduate engineering students to interdisciplinary topics and soft skills. A particularity of the programme is that each course is taught by an interdisciplinary research team covering engineering and social science expertise. In this communication course, the teaching team was composed of three lecturers with expertise in social media, information systems, behavioral sciences and management.

The lecturers were familiar with the usage of social media in the classroom, as they had already used social media apps such as Twitter or SpeakUp in their practice. To understand the attitude of students towards technology, we conducted a voluntary questionnaire at the beginning of the session (based on 7-point Likert scale questions). The respondents ($N = 140$) considered that technologies are useful in the classroom (average Likert score $\mu = 5.57$) and there should be more interaction in the courses ($\mu = 4.62$). Many students asserted that they feel quite free to express what they think in class ($\mu = 4.60$), but also that they often have questions that they do not ask ($\mu = 4.52$). Furthermore, students had a variety of opinions on whether anonymity could be important in order to express what they think during the courses ($\mu = 4.09$).

During this course, SpeakUp was introduced as a communication channel with students to increase interaction, but it also had another pedagogical purpose: since the course deals with communication channels, social media and TEL, SpeakUp would provide students with hands-on experience of many of the subjects studied in class.

## 5.2 Student Engagement (Topic 1)

Teachers, via the [T_QUE] questionnaire, perceived the app as engaging for the students ($\mu = 5$ in a 5-point Likert scale). The teachers pointed out that the main aspects triggering the high engagement could be the possibility of getting responses quickly without being exposed to the whole audience, the anonymity, the potential to know and react on what others think, as well as the opportunity to interact with everyone.

As an overview, if we compare the number of students attending the session (145) [R_OBS] with those joining the SpeakUp chatroom (147) [SP_LOG], we may infer that almost everyone used the tool, even though such use was not compulsory. The number of students registered in SpeakUp was higher than the actual students participating in the face-to-face session, due to the fact that some students started using the app from their phone, and then switched to using it from their computer [R_OBS].

Figure 4 shows how much teachers and students participated face-to-face and via SpeakUp throughout the session (from 16:15 to 18:00). Face-to-face activity is measured in minutes of active participation extracted from the video [R_VID]. Concretely, in the face-to-face channel, teachers were speaking for about 77 min and students 11 min. In the case of SpeakUp, the participation is measured according to the number of actions [SP_LOG], obtaining a total of 51 and 3841 actions carried out by teachers and students respectively. Looking at Fig. 4 we can identify a certain connection with the events happening face-to-face [R_OBS, R_VID]. For example, although the teachers used SpeakUp from the very beginning of the lesson (e.g., adding welcome messages), the app was presented to the students around 16:35, reason why the students started using it later. Then, there was a break of 15 min in the session at 17:10, but students continued using the app during this period. Besides, the main peaks of activity correspond to moments in which teachers asked explicitly to use the app in order to answer a poll (e.g., around 17:15) or to write down some ideas about certain topics.

Based on the user activity (e.g., number of posted messages, number of likes and dislikes, etc.) [SP_LOG], we have carried out a bottom-up clustering analysis

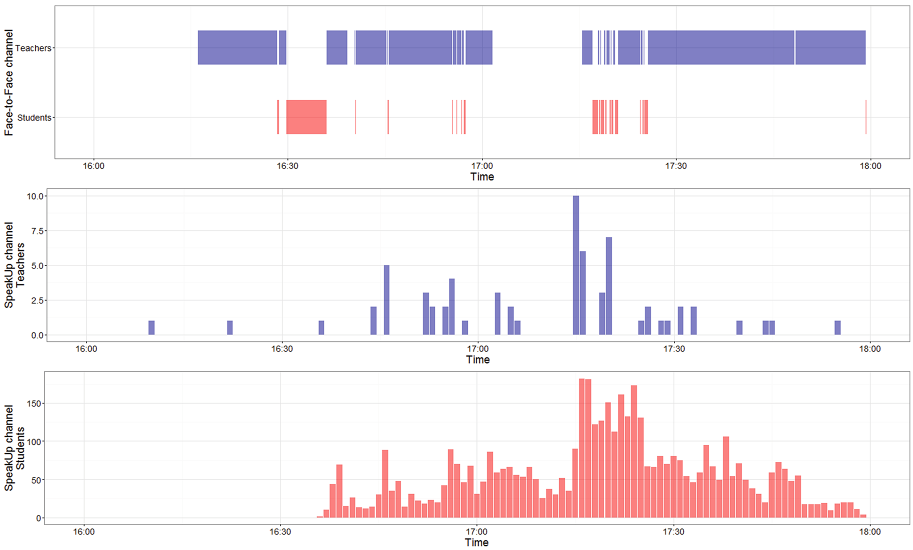



**Fig. 4.** Face-to-face and SpeakUp-mediated participation during the session.

(using a k-means clustering algorithm with k = 6, chosen in terms of within-groups sum of squares), leading to the kinds of users detailed in Table 1. These clusters include large groups of students with low amounts of active usage of SpeakUp (e.g., "Passive"), but also smaller clusters of students with very peculiar engagement patterns (e.g., "Very pro-active", which create a large number of messages and votes; or "Super-active voters", who do an unusual amount of voting – especially dislikes –, and very little else).

**Table 1.** Types of students based on their interaction with SpeakUp. The action values represent the average values for the cluster.

| Clusters | # Students | # Actions | # Answers to polls | # Posted messages | # Replies to messages | # Likes | # Dislikes | # Spam reports |
|---|---|---|---|---|---|---|---|---|
| "Passive" | 77 | 7 | 0 | 0 | 0 | 4 | 2 | 0 |
| "Semi-passive" | 36 | 14 | 1 | 1 | 0 | 8 | 5 | 0 |
| "Pro-active/reactive" | 6 | 38 | 1 | 7 | 2 | 17 | 10 | 1 |
| "Mildly pro-active" | 22 | 63 | 1 | 5 | 0 | 32 | 25 | 0 |
| "Very pro-active" | 3 | 143 | 1 | 19 | 1 | 78 | 44 | 0 |
| "Super-active voters" | 4 | 190 | 1 | 0 | 0 | 58 | 130 | 0 |

### 5.3 Student Attention (Topic 2)

From the teachers' perspective [T_QUE], SpeakUp had no clear impact on the student attention ($\mu = 3$ in a 5-point Likert scale). They found that although SpeakUp enabled an open channel for topics which might not be related to the course, the app took up one screen of the students' devices, increasing the chances of gathering focused and distracted students. Besides, teachers considered that it might be hard for students to pay attention to both the face-to-face and SpeakUp channels simultaneously.

A minority of students considered that the app did distract them (18.4 %, N = 65, [S_QUE] see Fig. 5, left). However, among student comments [S_COM], only 30.7 % of the messages (out of a total of 322) were categorised as relevant (relevance ratio[1] =-0.38). These relevant messages were related to the learning content presented during the lesson, the course details, the organisers (i.e., teachers and teacher assistants). It should be noted that the mean scores provided by students – sum of likes and dislikes – are slightly higher for relevant ($\tilde{x} = 1.16$) than for non-relevant messages ($\tilde{x} = 0.89$).

[1] Calculated as: $(relevant\_posts - non\_relevant\_posts)/(relevant\_posts + non\_relevant\_posts)$. Hence, ranging from -1 (all messages irrelevant) to +1 (all messages relevant).

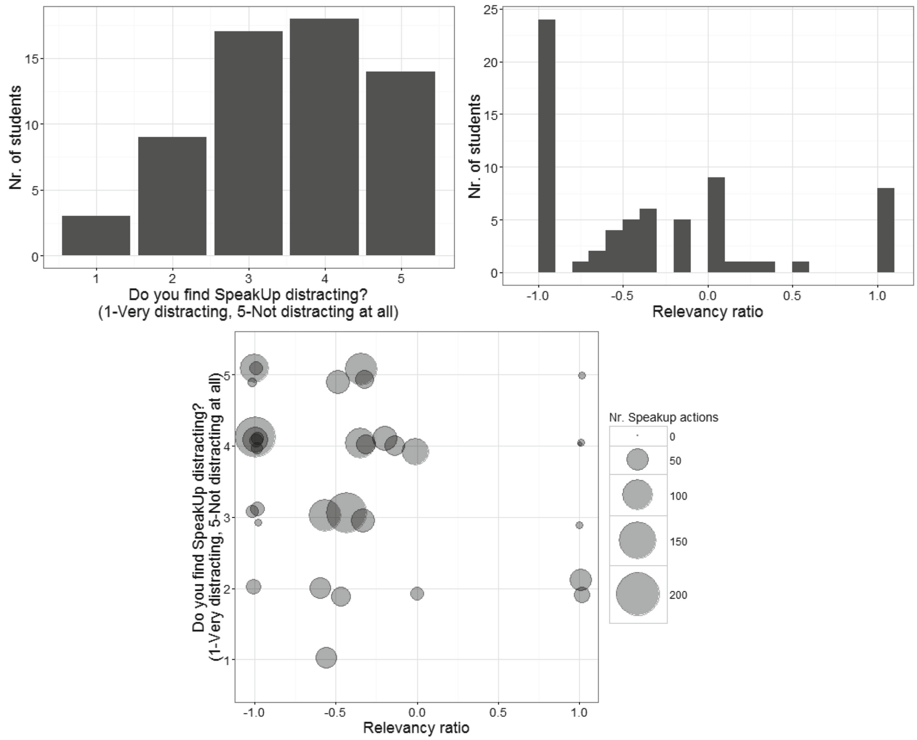



**Fig. 5.** Students' subjective opinions on whether SpeakUp is distracting them (top-left), and distribution of values of relevancy ratio of messages per student (top-right). At the bottom, both values are represented along each student's SpeakUp participation (size of the circles).

To get a picture of the quality of the contribution of each student, Fig. 5 (right) depicts the distribution of students in terms of their relevancy ratio. This diagram shows that many of the students sent mainly non-relevant messages, while just 12 students (out of 68 students who generated any kind of message) sent mostly relevant messages; a significant amount of students sent both relevant and not relevant messages.

Figure 5 (bottom) puts both graphs in perspective and relates the students perception about SpeakUp and their behaviour using the app. Those students that considered the app less distracting (4 and 5 in the Likert-scale) where the ones who created more non-relevant messages. On the other hand, those that perceived the SpeakUp as more distracting, contributed with less messages but, in some cases, more relevant ones.

### 5.4 Social Interaction (Topic 3)

The teachers [T_QUE] perceived SpeakUp as a mechanism that promoted interaction between them and students ($\mu = 5$ in a 5-point Likert scale) and among students themselves ($\mu = 4.5$). Regarding the interaction between teachers and students, the app helped teachers discover and handle important questions and comments. Among the drawbacks, the main concern was that SpeakUp messages required supervision, e.g., to avoid bullying and other interactions detrimental to the class dynamic.

In order to better understand how users interacted during the session, Fig. 6 shows the amount of interaction registered in the face-to-face and SpeakUp channels. For face-to-face interaction, we have taken into account the amount of time spent in the communication (extracted from the video observation [R_VID]). For SpeakUp interactions we have counted the number of messages and votes generated by the users that were registered in the logs [T_COM, S_COM, SP_LOG]. Figure 6 reveals that the face-to-face channel supported mainly the interaction going from teachers to students, while SpeakUp concentrated most of the interactions between students.

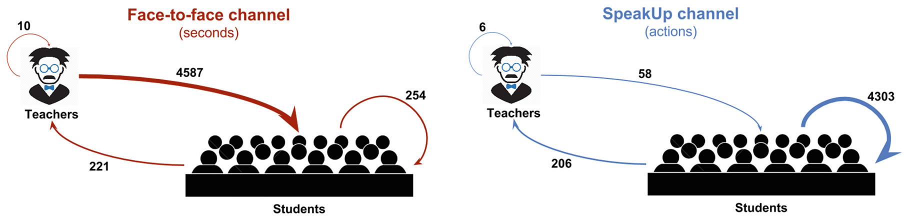



**Fig. 6.** Analyses of the communication direction in the face-to-face and SpeakUp channel.

The social network analysis shown in Fig. 7 reveals that, far from existing multiple separate groups that interacted mostly among themselves (a common pattern in social networks), the network of interactions was rather dense. This may be caused by the fact of using anonymous users. Since it is not possible to know who sends the message, the user cannot decide to answer or follow just specific people, and mainly reacts to the content post by other users. As it is shown in Fig. 7 (right), although 20 students were isolated, the interaction degree ($\mu = 59.4$, $\tilde{x} = 20$, $\sigma = 89.2$) is much higher than the number of students that could interact in a physical environment (e.g., 8 peers sitting around). Note that many of the students did not received any vote or comment (in-degree: $\mu = 29.7$, $\tilde{x} = 0$, $\sigma = 60.9$), while, on the other hand, most of the students comment, answer or vote at least once (out-degree: $\mu = 29.7$, $\tilde{x} = 10.5$, $\sigma = 45.5$).

### 5.5 Teaching Style (Topic 4)

As a general schema, the teachers of this course switch often during a same lecture to keep the course dynamic. Figure 8 shows which parts of the session were

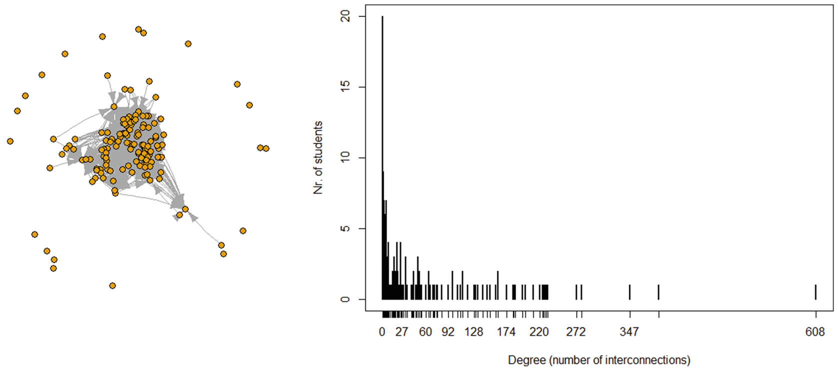



**Fig. 7.** Social network of SpeakUp interactions (left), and degree (number of interconnections) of the different SpeakUp users (right).

led by each one of the three teachers [R_VID] and the amount of relevant activity during such periods [T_COM, S_COM, SP_LOG]. Although, at first sight there was more activity in SpeakUp during the parts of the session led by Teacher1, it would be necessary to analyse more sessions in order to clarify if there is a dependence with the presentation style of the teacher (e.g., voice level, inflections, and physical language, duration), the support material (e.g., slides, videos, questionnaires, specific apps, etc.) or the specific content of the presentation. What seems to be more obvious is that high levels of relevant activity correspond to those moments when the teachers explicitly asked the students to use SpeakUp for specific learning purposes.


**Fig. 8.** Overview of actors, actions and resources used by the teachers during the session.

Regarding the way teachers used SpeakUp [T_QUE], before the session (see Fig. 4), Teacher1 created the chatroom to be shared with the rest of the users. Then, during the session, while one teacher was presenting, the others checked SpeakUp to identify emerging questions or problems, vote (dis-like) non-relevant comments, and delete inappropriate ones. A significant difference between teachers styles [T_QUE, R_OBS, R_VID] is the way they interact with the tool. On the one hand, Teacher2 and Teacher3 did not use SpeakUp while they were lecturing. On the other hand, Teacher1 used it during his slots to satisfy his own teaching needs (e.g., he had a quick look to the messages when there was some noise, and checked in case of questions at the end of the presentation), and to support some learning activities (e.g., he asked students to answer some questions and give their opinions using the app).

As already mentioned in this section, the teachers found several benefits using SpeakUp that supported them in their practice. This tool provided them with awareness of a students back channel, and informed the interventions. However, they also pointed out that managing two simultaneous channels is demanding, specially difficult if teaching alone. Therefore, there is a need for finding an adequate scheme to handle face-to-face and computer-mediated interactions.

## 6 Discussion, Conclusions and Future Work

In our way towards understanding how to use social media effectively in the classroom, this paper analyses the use of SpeakUp in a face-to-face session with 3 teachers and 145 university students. In particular, we have explored to what extend SpeakUp favored situations that lead to learning, such as active participation (topic 1 - engagement), remaining on-task (topic 2 - attention), and social interaction (topic 3). Besides, we have explored the impact of the teaching style on the SpeakUp usage (topic 4).

The engagement results reveal that, even though the use of SpeakUp was optional, all students attending the session at least accessed the tool once. The clustering of users reveals that there is a gradient of involvement from passive to active users in terms of posting and voting. It should be noted that for most clusters there is usually a 2 to 1 ratio between the number of upvotes and the number of downvotes. Interestingly there is a cluster that we could dub the "SpeakUp police", who are the most active voters of all, and are mostly assigning negative votes in the opposite proportion.

Whereas many students wrote mainly non-relevant messages, compared to the 12 students who contributed mostly relevant messages, there was a significant amount of students who posted both relevant and not relevant messages, which means that using the tool for something else than learning is not just the activity of some bad apples. The results showing that the students with the lowest relevancy scores find the interaction in SpeakUp not distracting, whereas the students with the highest relevancy score find it the most distracting indicates a potential risk for the app usage if the high relevancy students start turning off their app.

One of the SpeakUp advantages highlighted by the teachers, and supported by the data analyses, refers to social interaction. First, students could not only share (doubts, problems, resources) but also comment and vote others contributions, favoring to get answers from peers without waiting for the teachers. Second, the app complemented the face-to-face channel. While most of the time teachers interacted orally with the students, the interaction between students was supported mostly via SpeakUp. Additionally, comparing the number of students reachable in the physical environment (e.g., 8 peers sitting around) versus the interaction degree in SpeakUp ($\mu = 59.4$, $\tilde{x} = 20$), we may conclude that the tool contributed to increase the social network.

One of the aspects to be discussed is the twofold effect that anonymity might have on engagement, attention and interaction results. On the one hand, the anonymity could increase the usage of SpeakUp, since the students embrace the idea of not disclosing their identity (see Sect. 5.1). The flip side was that the anonymity brought more non-relevant messages and required teachers to monitor the activity and intervene in case of inappropriate interaction (e.g. bullying).

Regarding the teacher impact on the student use of SpeakUp, it is noteworthy that when teachers asked the students to use the app in a certain way, the relevancy of the user activity increased significantly. Thus, the teacher role as scaffolding provider could contribute to a more effective use of the app.

Going back to the issue addressed in this case study, we can conclude that SpeakUp favored situations that led to learning, especially in terms of active participation (i.e., engagement) and social interaction. However, dealing with the attention, alternatives should be found in order to foster the appearance of relevant content (e.g., with teacher guidance). Nevertheless, the fact that the case study only covered the first session when the app was used by the students, could have introduce some additional distraction (novelty factor). Thus, it would be necessary to analyse the use of SpeakUp during the whole course, to see how student engagement, attention and interaction evolve. This study is currently under way, and is our most immediate avenue for future research.

## References

1. Anderson, R.J., Anderson, R., Vandegrift, T., Wolfman, S., Yasuhara, K.: Promoting interaction in large classes with computer-mediated feedback. Designing for Change in Networked Learning Environments: Proceedings of the International Conference on Computer Support for Collaborative Learning 2003. Computer-Supported Collaborative Learning, vol. 2, pp. 119–123. Springer, Amsterdam (2003)
2. Bergstrom, T., Harris, A., Karahalios, K.: Encouraging initiative in the classroom with anonymous feedback. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part I. LNCS, vol. 6946, pp. 627–642. Springer, Heidelberg (2011)
3. Blood, E., Neel, R.: Using student response systems in lecture-based instruction: does it change student engagement and learning? J. Technol. Teach. Educ. **16**(3), 375–383 (2008)

4. Creswell, J.W., Plano Clark, V.L., Gutmann, M.L., Hanson, W.E.: Advanced mixed methods research designs. In: Tashakkori, A., Teddlie, C. (eds.) Handbook of Mixed Methods in Social and Behavioral Research, pp. 209–240. Sage Publications, Thousand Oaks (2003)
5. Dewan, P.: An integrated approach to designing and evaluating collaborative applications and infrastructures. Comput. Support. Coop. Work (CSCW) **10**(1), 75–111 (2001)
6. Dillenbourg, P.: What do you mean by collaborative learning. In: Collaborative-learning: Cognitive and Computational Approaches, vol. 1, pp. 1–15. Elsevier Publishing (1999)
7. Du, H., Rosson, M.B., Carroll, J.M.: Augmenting classroom participation through public digital backchannels. In: Proceedings of the 17th ACM International Conference on Supporting Group Work, GROUP 2012, NY, USA, pp. 155–164. ACM, New York (2012)
8. Erickson, J., Siau, K.: Education. CACM **46**(9), 134–140 (2003)
9. Fies, C., Marshall, J.: Classroom response systems: a review of the literature. J. Sci. Educ. Technol. **15**(1), 101–109 (2006)
10. Greene, J.C., Benjamin, L., Goodyear, L.: The merits of mixing methods in evaluation. Evaluation **7**(1), 25–44 (2001)
11. Grotenbreg, G., Wong, S.B.J.: Using Pigeonhole ® Live to elicit feedback, questions & reinforce learning during lectures. CDLT Brief **16**(2), 2–7 (2013)
12. Guba, E.G.: Criteria for assessing the trustworthiness of naturalistic inquiries. J. Theor. Res. Devel. Educ. Commun. Technol. **29**(2), 75–91 (1981)
13. Harry, D., Green, J., Donath, J.: Backchan.nl: integrating backchannels with physical space. In: CHI 2009, pp. 2751–2756. ACM (2008)
14. Holzer, A., Govaerts, S., Vozniuk, A., Kocher, B., Gillet, D.: Speakup in the classroom: anonymous temporary social media for better interactions. In: Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI EA 2014, NY, USA, pp. 1171–1176. ACM, New York (2014)
15. Holzer, A., Govaerts, S., Ondrus, J., Vozniuk, A., Rigaud, D., Garbinato, B., Gillet, D.: SpeakUp – a mobile app facilitating audience interaction. In: Wang, J.-F., Lau, R. (eds.) ICWL 2013. LNCS, vol. 8167, pp. 11–20. Springer, Heidelberg (2013)
16. Horowitz, H.M.: Student response systems: interactivity in a classroom environment. In: Proceedings of the Sixth Annual Conference on Interactive Instruction Delivery. Salt Lake City, UT, February 1988
17. Jensen, E.: Teaching with the brain in mind. Association for Supervision & Curriculum Development (1998)
18. Johnson, R.B., Onwuegbuzie, A.J.: Mixed methods research: a research paradigm whose time has come. Educ. Researcher **33**(7), 14–26 (2004)
19. Jorrín-Abellán, I.M., Stake, R.E., Martínez-Monés, A.: The needlework in evaluating a CSCL system: the evaluand oriented responsive evaluation model. In: International Conference on Computer Supported Collaborative Learning, CSCL 2009, pp. 68–72, International Society of the Learning Sciences, Rhodes, Greece (2009)
20. Kinzie, M.B., Whitaker, S.D., Hofer, M.J.: Instructional uses of instant messaging (IM) during classroom lectures. Educ. Technol. Soc. **8**(2), 150–160 (2005)
21. Kreijns, K., Kirschner, P.A., Jochems, W.: Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. Comput. Hum. Behav. **19**(3), 335–353 (2003)

22. Kuznekoff, J.H., Munz, S., Titsworth, S.: Mobile phones in the classroom: examining the effects of texting, twitter, and message content on student learning. Commun. Educ. **64**(3), 344–365 (2015)
23. McCarthy, J.F., Boyd, D.M.: Digital backchannels in shared physical spaces: experiences at an academic conference. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2005, pp. 1641–1644. ACM, New York (2005)
24. Miles, M.B., Huberman, A.M.: Qualitative Data Analysis: An Expanded Sourcebook. Sage Publications, Thousand Oaks (1994)
25. Prince, M.: Does active learning work? a review of the research. J. Eng. Educ. **93**(3), 223–231 (2004)
26. Reinhardt, W., Ebner, M., Beham, G., Costa, C.: How people are using twitter during conferences. In: Proceedings of 5th EduMedia Conference, p. 145 (2009)
27. Retelny, D., Birnholtz, J.P., Hancock, J.T.: Tweeting for class: using social media to enable student co-construction of lectures. In: Poltrock, S.E., Simone, C., Grudin, J., Mark, G., Riedl, J. (eds.) CSCW (Companion), pp. 203–206. ACM (2012)
28. Sana, F., Weston, T., Cepeda, N.J.: Laptop multitasking hinders classroom learning for both users and nearby peers. Comput. Educ. **62**, 24–31 (2013)
29. Scornavacca, E., Huff, S., Marshall, S.: Mobile phones in the classroom: if you can't beat them, join them. CACM **52**(4), 142–146 (2009)
30. Seaman, J., Tinti-Kane, H.: Social media for teaching and learning. Technical report, Pearson Learning Solutions (2013)
31. Stake, R.: The Art of Case Study Research. Sage Publications, Thousand Oaks (1995)
32. Stowell, J.R., Nelson, J.M.: Benefits of electronic audience response systems on student participation, learning, and emotion. Teach. Psychol. **34**(4), 253–258 (2007)
33. Sung, Y.T., Chang, K.E., Liu, T.C.: The effects of integrating mobile devices with teaching and learning on students' learning performance: a meta-analysis and research synthesis. Comput. Educ. **94**, 252–275 (2016)
34. Trees, A.R., Jackson, M.H.: The learning environment in clicker classrooms: student processes of learning and involvement in large university-level courses using student response systems. Learn. Media Technol. **32**(1), 21–40 (2007)
35. Worthington, D.L., Levasseur, D.G.: To provide or not to provide course powerpoint slides? the impact of instructor-provided slides upon student attendance and performance. Comput. Educ. **85**, 14–22 (2015)
36. Yardi, S.: The role of the backchannel in collaborative learning environments. In: Proceedings of the 7th International Conference on Learning Sciences, ICLS 2006, pp. 852–858. International Society of the Learning Sciences (2006)

# Enhancing Public Speaking Skills - An Evaluation of the Presentation Trainer in the Wild

Jan Schneider(✉), Dirk Börner, Peter van Rosmalen, and Marcus Specht

Welten Institute, Open University of the Netherlands, Heerlen, The Netherlands
{jan.schneider,dirk.boerner,peter.vanrosmalen, marcus.specht}@ou.nl

**Abstract.** The increasing accessibility of sensors allows the study and development of multimodal learning tools that create opportunities for learners to practice while receiving feedback. One of the potential learning scenarios addressed by these learning applications is the development of public speaking skills. First applications and studies showed promising empirical results in laboratory conditions. In this article we present a study where we explored the use of a multimodal learning application called the *Presentation Trainer,* supporting learners with a real public speaking task in the classroom. The results of this study help to understand the challenges and implications of testing such a system in a real-world learning setting, and show the actual impact compared to the use in laboratory conditions.

**Keywords:** Evaluation in the wild · Sensor-based learning support · Public speaking · Multimodal learning application

## 1 Introduction

Experiencing a great presenter delivering a novel idea is an inspiring event. Therefore, at least for the last 2500 years humans have been studying the art of the oratory [1]. Currently the ability to present effectively is considered to be a core competence for educated professionals [2–5]. This relevance in learning how to communicate effectively is reinforced by the thought that ideas are the currency of the twenty first century [6]. Research on how to develop public speaking skills is a topic that has already been extensively studied. One of the conclusions to be drawn out of these studies is that practice and feedback are key aspects for the development of these skills [7]. Whereas it is possible to attend different courses and seminars on public speaking, opportunities to practice and receive feedback from tutors or peers under realistic conditions are limited.

Sensors have lately become increasingly popular [8], showing to be a technology with great potential to enhance learning, by providing users with feedback in scenarios where human feedback is not available or to give access to data sources to enhance learning [9]. This has led to the development and research of new sensory technologies


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 263–276, 2016.
DOI: 10.1007/978-3-319-45153-4_20

designed to support users with the development of their public skills [10–13]. These technologies have not been widespread yet, and so far their impact has not been tested outside from controlled laboratory conditions. One of these technologies is the *Presentation Trainer* (PT), a multimodal tool designed to support the development of basic public speaking skills, by creating opportunities for learners to practice their presentations while receiving feedback [13]. This paper describes a field study where we took the PT outside of the laboratory and tested it in a classroom. The paper discusses the implications of using such a system in the wild, and identifies which of the findings in a lab setting [13] also hold in the real world.

## 2 Background Work

Educational interventions such as feedback are needed to develop public speaking skills [14]. Having a human tutor available to give feedback on these skills is neither always feasible nor affordable. Therefore, technological interventions designed to provide this feedback are desirable. Public speaking skills require from presenters a coherent use of their verbal and nonverbal channels. Timely measurement of these multimodal performances with an acceptable accuracy is challenging. However, in recent years driven by the rising availability of sensors, research on multimodal learning applications designed to support the development of public speaking skills has been undertaken.

During a presentation, the presenters communicate their messages using their voice together with their full body language, e.g., body posture, use of stage, eye contact, facial expressions, hand gestures, etc. Multimodal learning applications supporting the development of public speaking skills [10–16] generally use a depth sensor such as the Microsoft Kinect[1] in order to capture the body language of the user, and microphone devices to capture the user's voice.

Studies on applications designed to support public speaking skills have been exploring effective strategies to provide feedback to users. In [11] feedback indicating whether the energy, body posture and speech rate is correct or not, is displayed on a Google Glass[2]. Another feedback strategy employed in [10, 15] is the use of a virtual audience. Members of the virtual audience change postures and behaviors depending on the nonverbal communication of the user. Besides the display of the virtual audience the prototype in [10] also provides the user with direct visual indications regarding her own body posture. The applications in [12, 16] provide the user with a dashboard interface that displays a mirrored image of the user together with modules indicating the use of nonverbal communication aspects such as use of gestures, voice, etc. In line with that, the feedback interface of the PT shows a mirror image of the user and displays at maximum one instruction to the user regarding her nonverbal communication at a given time (see Fig. 1). This instruction is communicated to the user through a visual and a haptic channel [13].

---

[1] https://dev.windows.com/en-us/kinect/hardware.

[2] https://www.google.com/glass/start/.

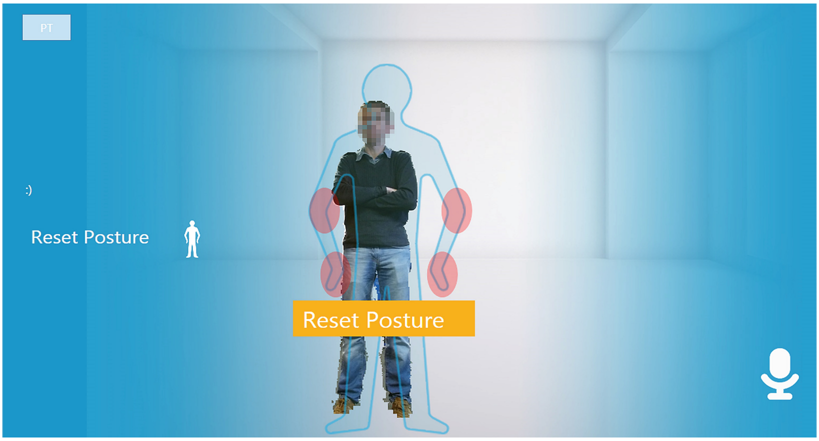



**Fig. 1.** PT telling the user to correct the posture.

The impact of this type of applications on learners has also been studied, showing positive results in laboratory conditions. In the study of [10] the feedback of the system, regarding the closeness or openness of the learner's body posture, helped learners to become more aware of their body posture. The impact of the PT's feedback on learners has also been studied in controlled setups. The study in [13] showed, through objective measures made by the system, that after five practice sessions receiving feedback from the PT learners on average reduced 75 % of their nonverbal mistakes.

## 3 Purpose

In this study we tested the PT in a classroom setting following an exploratory research approach [17], focusing on three main objectives:

*Objective 1*: The first objective of this study is to explore the implications of investigating the use of a tool such as the PT in a regular learning scenario outside of a laboratory setup.

*Objective 2:* Studies on multimodal learning applications for public speaking have shown promising results in laboratory conditions according to quantified and timely machine measurements [10, 13]. However, the purpose of a presentation is to transmit the desired message and provide the desired impact to a human audience, in contrast of improving a machine-based score. Studies showing evidence that an improved performance according to machine measurements is reflected in a better presentation according to a human audience are still missing. Therefore, the second objective of this study is to gain insights on how the improvements obtained by a learner using the PT to practice for a presentation relate to the impact that this trained presentation has on the audience. In other words, to what extent does an audience agree with the PT that a presentation improved.

*Objective 3:* A core competence for current professionals is having good public speaking skills [2–5]; therefore teaching these skills has become a common target for different courses. Feedback is a key aspect for learning and developing public speaking skills [7], therefore current courses in public speaking include well-established feedback practices to help learners with the development of these skills. The effectiveness of this feedback depends on various variables. One of these variables concerns the source where the feedback comes from. Feedback provided by a tutor in combination with feedback provided by peer students has proven to be more effective than feedback provided only by a human tutor [18]. The third objective of this study, researches the introduction of the PT to the already established practices for teaching public speaking skills, exploring whether its use and feedback contribute to the creation of more comprehensive learning scenarios for students.

# 4 Methodology

## 4.1 Study Context

We conducted this study in the setting of a course in entrepreneurship for master students in a university. In this course students were divided in two teams, where each team is represented as an entrepreneurial business. During the course the teams have to develop and present their project. Thus, the students of the course receive some presentation training guidance. The teams have to give a presentation about their projects twice, at the middle and at the end of the course. The middle term presentations are recorded and in following sessions these recordings are used to give feedback to the students regarding their presentation skills, both by tutors and peers.

## 4.2 Study Procedure

This study was conducted some sessions after the students have already presented their project and received feedback. Nine participants, seven males and two females between the age of 24 and 28 years old took part in the study. A sketch of the study is shown in Fig. 2. To prepare for the study, students got the homework to individually prepare a 60–120 s long pitch regarding their project. One week later the study was conducted during a two-hour session slot.

The study started with students individually presenting their pitch in front of their peers and course teachers. The objective of this first pitch was to obtain a baseline of the students' performance. Peers evaluated the pitch by filling in a presentation assessment questionnaire.

After presenting the pitch each student moved to another room for the practice sessions. Before the practice sessions, students received a small briefing regarding the PT's feedback. The purpose of this small briefing was to reduce the exploration time needed to understand the feedback given by the PT. After this short briefing time, participants were supposed to know how to correctly react to the feedback given by the PT. The practice sessions consisted delivering the pitch two consecutive times while receiving feedback from the PT. During the practice session students stood between

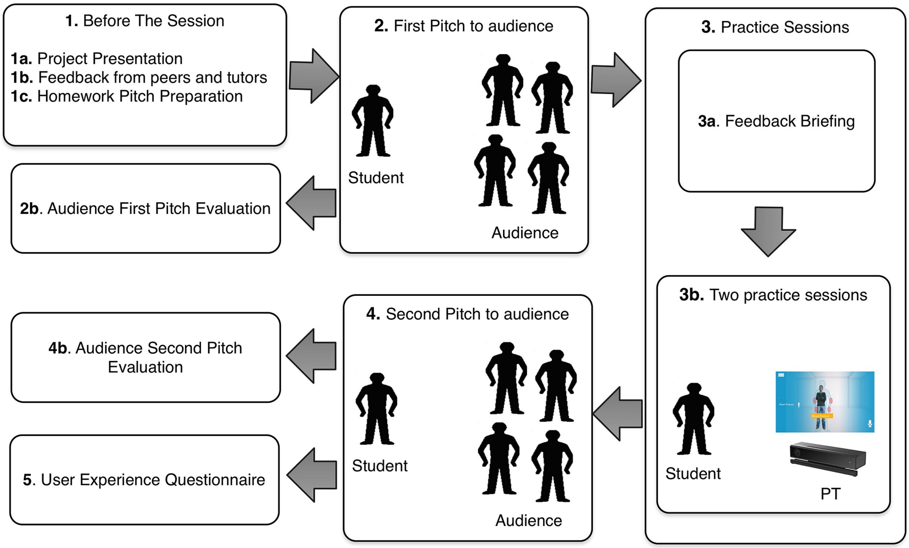



**Fig. 2.** Study procedure

1.5 and 3 m in front of the Microsoft Kinect sensor and a 13-inches display laptop running the PT.

For the next phase of the study, the student returned to the classroom and presented the pitch once more to their peers. The objective of this second pitch was to explore the effects of the practice sessions. To observe these effects, peers evaluated this final presentation once more by filling in the presentation assessment questionnaire. The PT was also used to assess these pitches. However, due to a technical failure only the pitches given by the last three participants were assessed by the PT. After delivering this final pitch, students were asked to fill in a questionnaire regarding the experience of using the PT to practice.

### 4.3 Apparatus and Material

To evaluate the pitches done by the students, peers filled in a presentation assessment questionnaire. The questionnaire consists of eleven Likert-scaled items. The first seven items refer to a general assessment of the presentation including: the overall quality of the presentation, delivery of the presentation, speaker knowledge about the topic, confidence of the speaker, enthusiasm of the speaker, understandability of the pitch, and fun factor of the pitch. The last four items consisted of some of the specific nonverbal behaviors that can be trained using the PT: posture, use of gestures, voice quality, and use of pauses.

To practice for the second presentation of the pitch students used the current version of the PT. This version of the PT uses the immediate feedback mechanism described in [13], providing users with the maximum of one corrective feedback at the time regarding their body posture, use of gestures, voice volume, phonetic pauses or filler sounds, use of pauses, and facial expressions (45 s without smiling). The PT logs all the recognizable behaviors (mistakes and good practices) as events. It displays these events at the end of each practice the session a timeline (see Fig. 3) allowing learners to get an overall picture of their performance. These logs are stored into files that can later be used for data analysis.

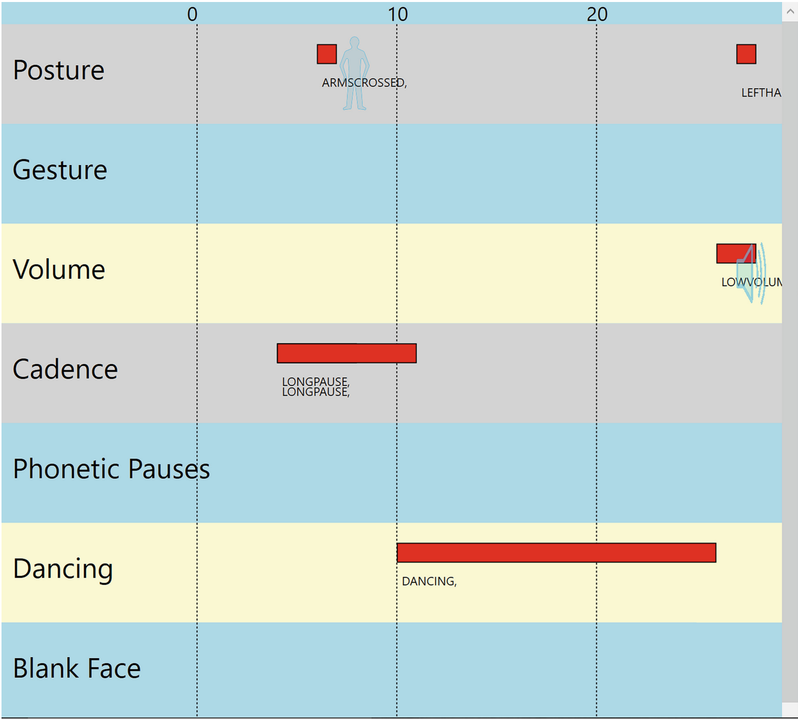



**Fig. 3.** Timeline displaying all tracked events, showed to the user after the presentation.

A user experience questionnaire was used to capture the impressions of the students regarding the use of the PT. This questionnaire consists of seven items in total, five Likert-scale items and two open questions. The purpose of this questionnaire was to inquire the learning perception, usefulness of the system, and comparison between human assessment and system assessment.

## 5 Results

The peer evaluation of the first pitches is shown in Fig. 4. Regarding the general aspects of the pitch, the item with the best score was the knowledge about the topic displayed by the presenter with an average score of 3.76 and the item with the lowest score was the entertaining factor of the pitch with an average score of 3.1. The non-verbal communication behavior with the highest score was the voice quality of the presenter with an average score of 3.73 and the behavior with the lowest score was the proper use of pauses during the pitch with an average score of 3.21.

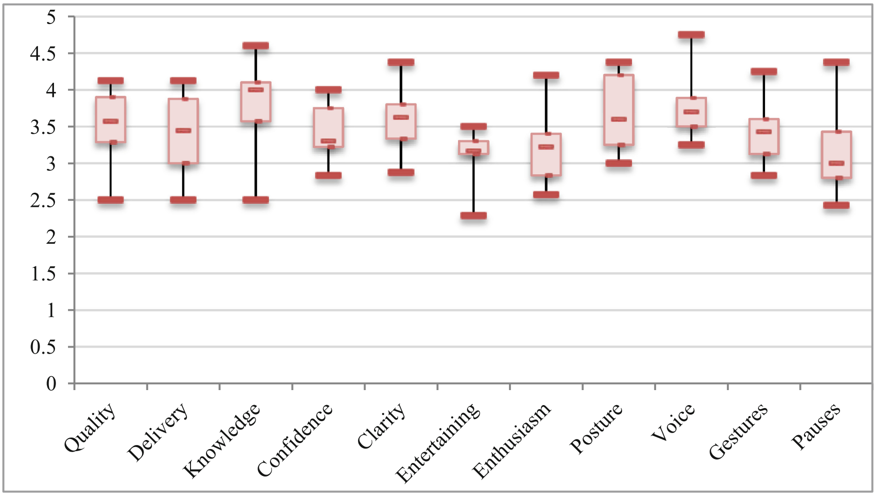



**Fig. 4.** Evaluation scores of the first pitches.

After giving the first pitch, students practiced it two times using the PT. We analyzed these practice sessions using the logged files created by the PT. To evaluate the impact of each of the identified behaviors captured by the PT, we used the percentage of time that this behavior was displayed during the training session (pTM). The pTM value for each behavior has a range from 0 to 1, where 0 indicates that the behavior was not displayed at all and 1 indicates that the behavior was identified throughout the whole presentation. The average pTM values for all the tracked behaviors are displayed in Table 1. Results indicate that participants on average during the second practice session show an improvement in all trained aspects. The behavior that on average received the worst assessment for the first practice session was the use of gestures, followed by the voice volume and then posture. The pTM value for the other tracked behaviors was very low. In the second practice session voice volume received the worst assessment, followed by gestures and then posture. The area showing the biggest improvement was the use of gestures.

The peer evaluation of the pitches presented after the practice sessions is shown in Fig. 5. Regarding the general assessment of the pitches the item with the highest score was the knowledge about the topic displayed by the speaker with an average score of 3.96. The item with the lowest score having an average of 3.55 was the entertaining factor of the pitch. Regarding the nonverbal communication aspects, the one with the highest score was the voice quality of the presenter with and average of 4.14 and the correct use of pauses was the lowest with and average of 3.71.

**Table 1.** pTM scores capture during the practice sessions. Mean and standard deviation.

| | Posture pTM | Volume pTM | Pauses pTM | Blank F. pTM | Gestures pTM | Dancing pTM | Phonetic P. pTM | Total pTM |
|---|---|---|---|---|---|---|---|---|
| Session 1 | 0.132 (0.22) | 0.179 (0.16) | 0.040 (0.41) | 0.083 (0.14) | 0.217 (0.18) | 0.026 (0.08) | 0.020 (0.01) | 0.697 (0.31) |
| Session 2 | 0.078 (0.11) | 0.167 (0.11) | 0.010 (0.17) | 0.019 (0.02) | 0.123 (0.12) | 0 (0) | 0.017 (0.01) | 0.414 (0.22) |
| Mean difference | 0.054 | 0.012 | 0.030 | 0.064 | 0.094 | 0.026 | 0.004 | 0.284 |

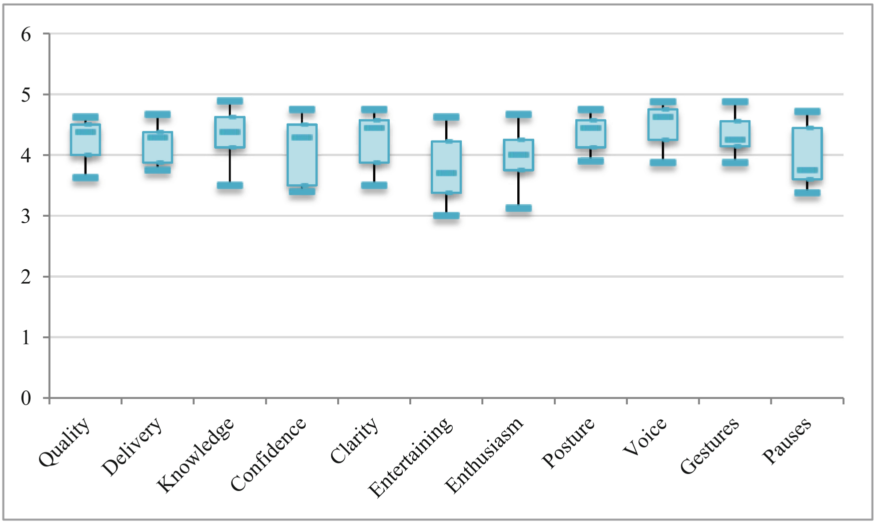



**Fig. 5.** Evaluation scores of the second pitches.

To explore the relevance of having a tool designed to practice specifically the delivery of the pitch, we used Pearson's r to measure the correlation between the scores of the overall quality of the pitch (content + delivery) and the scores of its delivery. These measurements show a correlation of [$r = 0.94$, $n = 18$, $p < 0.01$]. We also used Pearson's r on the scores of the pitches to measure the correlation between the behaviors that can be trained using the PT and the overall quality of the presentations (see Table 3). This with the objective to explore the relevance of training these behaviors. The behavior displaying the strongest correlation was the use of pauses, followed by posture, voice quality and use of gestures.

Figure 6 shows the comparison in the evaluations between the first and second pitches. These comparisons show and improvement in all evaluated items. The general quality of the pitches increased on a 21.94 %. We calculated the significance of this difference using a t-test. The result of this t-test was $t(14) = 3.6$, $p < .01$. This indicates that the improvement observed is statistically significant. Regarding the general aspects of a presentation the delivery of the pitch was the item displaying the biggest

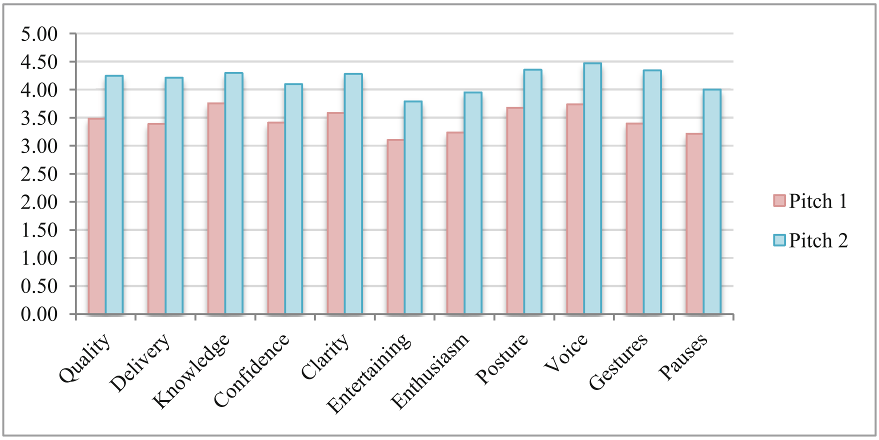



**Fig. 6.** Comparison between first and second pitch

improvement showing an increment of 24.27 %. The item showing the lowest improvement was the knowledge about the topic displayed by the presenter. This item had an improvement of only 14.37 %.

By examining the improvements on the nonverbal communication behaviors, the area that displayed the biggest improvements was the use of gestures with an increment of 27.89 %.

The PT's assessment the second pitch for the last three speakers is shown in Table 2[3]. Results from these tracked performances show that all of them had a total pTM value lower than 1.

**Table 2.** pTM values for the last three speakers on their final pitches.

| Speaker # | Posture pTM | Volume pTM | Pauses pTM | Blank F. pTM | Gestures pTM | Dancing pTM | Phonetic P. pTM | Total pTM |
|---|---|---|---|---|---|---|---|---|
| 7 | 0.160 | 0.088 | 0.054 | 0.104 | 0.000 | 0.000 | 0.026 | 0.427 |
| 8 | 0.148 | 0.063 | 0.153 | 0.000 | 0.000 | 0.000 | 0.026 | 0.390 |
| 9 | 0.142 | 0.105 | 0.112 | 0.243 | 0.000 | 0.015 | 0.039 | 0.656 |
| Average | **0.150** | **0.085** | **0.106** | **0.115** | **0.000** | **0.005** | **0.030** | **0.491** |

**Table 3.** Pearson's linear correlation. Mean and standard deviation.

| Aspect trained | Overall quality |
|---|---|
| Posture | $r = 0.86$, $n = 18$, $p < 0.01$ |
| Voice | $r = 0.85$, $n = 18$, $p < 0.01$ |
| Gestures | $r = 0.76$, $n = 18$, $p < 0.01$ |
| Pauses | $r = 0.89$, $n = 18$, $p < 0.01$ |

[3] 'A technical failure prevented data capture of the first six participants' (See Sect. 4.2).

Results from the user experience questionnaire are listed in Table 4. These scores show that students would likely use the PT to prepare for future presentations. Results show that students perceived an increment of their nonverbal communication awareness. Students felt that the feedback of the PT is more useful as an addition rather than as a reinforcement of the feedback that peers and tutors can provide.

**Table 4.** Results from the user experience questionnaire. Mean and standard deviation.

| Item | Likert-scale scores (1 Strongly disagree - 5 Strongly Agree) |
|---|---|
| My nonverbal communication awareness increased | 3.89 (0.93) |
| I learned something while using the PT | 3.67 (1.12) |
| I see myself using PT in the future | 4.11 (0.78) |
| The PT reinforced the feedback of peers and tutor | 3.56 (0.88) |
| The PT complements the feedback of peers and tutor | 3.78 (0.83) |

When asking students about the similarities between the PT's and the feedback received in previous sessions by tutors and peers all students mentioned the correct use of pauses while presenting. Two of them also mentioned the use of gestures. Four students mentioned that, previously, they received the feedback of not given enough eye contact to the audience by their tutors and peers and that this aspect is missing in the PT's feedback. Three students commented that receiving immediate feedback by the system makes it much more easy to identify and correct their behavior. One student mentioned that the PT gave feedback regarding the phonetic pauses while peers and tutors did not. One student mentioned a contradiction between the feedbacks regarding the use of voice. Peers and tutors in a previous presentation told the participant to speak louder, and during the training sessions the PT told the participant to speak softer.

## 6 Discussion

Studying the use of the PT outside of the laboratory in a real life formal learning scenario has several implications. In studies conducted in the lab, the setup of the experiment is carefully designed, allowing experimenters to have full control of variables such as time of each experimental session, location and instruments. This control allows the acquisition of reliable and replicable results. For this study we had to adapt our setup according to the restrictions of the ongoing course followed by the students. We encountered two main challenges while designing and conducting our study: time and location.

Regarding time, in previous laboratory studies participants had individual timeslots of sixty minutes, where they received all the briefing necessary and had five practice sessions with the PT. Moreover, experimenters had the chance to conduct their study with a large enough control and a treatment group, allowing them to assess significant results [13]. For this study we had two hours to conduct the whole experiment without knowing beforehand the amount of students that would show up that day for the course. Therefore, we reduced the training sessions from five to three and adapted to only two training sessions during the flow of the experiment. The act of training with the PT is individual and designed to be performed in a quiet room where the learner can focus on the task. That forced us to use a separate room where one student could do the practice session while the others waited in the lecture room. The room used for the practice sessions was not designed for the setup of the PT. The location of the power plugs, lighting conditions, place to position the Kinect and laptop screen running the PT were far from ideal. This problem of not having the ideal practice setup partially explains the difference between the average pTM values obtained in this study and the ones obtained in laboratory conditions [13]. In lab conditions the average values from the first and second training sessions were 0.51 and 0.32 respectively, while in this study they were 0.69 and 0.41. Nevertheless, despite the differences the values did show a similar trend displaying similar improvements in a less than ideal setting.

Previous studies showed that using the PT to practice for presentations improves the performance of the learner according to the measurements tracked by the PT [13]. The second objective of this study was to investigate whether using the PT to practice a presentation has also an influence in the way that the audiences perceives it. Results from this study showed that according to a human audience, all participants performed better in all aspects after having two practice sessions with the PT. The restricted time slot and restricted number of participants, did not allowed us to make use of a controlled and a treatment group. Therefore it is not possible to directly determine whether the improvements perceived by the audience are the results of practicing with the PT or just practicing. The results, however, revealed three key aspects suggesting the influence of the PT on this perceived improvement. The first key aspect is revealed by the assessed improvements regarding the general aspects of a presentation. The item showing the least improvement between the first and the second pitch is the knowledge that the presenter displayed regarding the topic. While on the other hand the item showing the biggest improvement was the delivery of the pitch. This aligns with the fact that the focus of the practice sessions using the PT was purely on the delivery of the pitch.

The second key aspect pointing out the influence of the PT has to do with the use of gestures. Use of gestures exhibited the biggest improvement from the first human assessed pitch to the second. This aligns with the computer assessment from the two practice sessions, where the aspect exhibiting the biggest improvements was also the use of gestures.

The third key aspect suggesting the influence of the PT is the PT's assessment of the three of the nine final pitches. In previous studies the average total pTM for presentations of people who did not practice with the PT was close to 1.0, in contrast with the results shown in this study where all the three measured final pitches had total pTM below 0.67. Unfortunately, as mentioned before, due to technical and logistical difficulties we were not able to assess all pitches using the PT.

For the third objective of this study we investigated whether the introduction of a tool such as the PT can contribute to the creation of more comprehensive learning scenarios for the acquisition of public speaking skills. Results from our study support this. As seen in the evaluations of the first pitch, the highest evaluated aspect was the knowledge of the topic displayed by the presenter. This gives us a hint that when preparing for a presentation or a pitch, a common practice is to focus efforts on preparing only its content. This practice does not seem optimal according to the strong correlation measured in this study between the overall quality of a pitch and the quality of its delivery. The results illustrate how by practicing the pitch two times using the PT, students significantly improved the overall quality of it. The students also reported benefits regarding their experience of using the PT to practice. They affirmed that the practice sessions helped them to learn something about public speaking and increase their nonverbal communication awareness. It is interesting to note that according to the students the feedback of the PT complements the feedback received by tutors and peers. Three students stated that the immediate feedback received by the PT helped them to exactly identify and correct their behavior. One more important aspect to note is that students expressed the intention to use the PT in the future.

This study showed some benefits of using of a tool such as the PT to support common practices for learning public speaking skills. However, the introduction of such a tool is still a challenge. The Microsoft Kinect is not a product owned by many students, and it is not feasible to provide each student with a Kinect in order to train some minutes for their presentations. However, Intel is already working in the miniaturization of depth cameras that can be integrated to laptop computers[4]. Therefore, in a medium term it will become more feasible for students to have access to tools such as the PT and use them for home practice. In the meantime the introduction for dedicated places to practice the delivery of presentations would be needed in order to introduce the support of these types of tools to the current practices for teaching and learning public speaking skills.

## 7 Conclusion and Future Work

The creation of multimodal learning technologies to support the development of public speaking skills has been driven in recent years by the advances and availability in sensor technologies. In laboratory settings some of these technologies have already started to show promising results. In this study we took one of these technologies, the *Presentation Trainer*, outside of the lab and conducted some tests with students following an entrepreneurship course as part of the course agenda. The main purpose of this study was to start the exploration of the support that these technologies can bring to a formal learning scenario.

Studying the use of the PT for a real classroom task revealed that location and time constrains interfere with the straightforward conduction of research. Due to location constrains it was not possible to set up the PT in ideal conditions for its use. Due to time constrains it was not possible to have the students follow all the expected training sessions, and we were not able to use the PT to measure all the first and second pitches presented to the audience. These constrains do not allow us to determine the causes for

[4] http://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html.

some of the obtained results in this study. However, results from this study align to a large extend with results obtained in the lab [13].

Regarding the support that the use of a tool such as the PT can bring to the established practices of teaching and learning public speaking skills, results from this study show the following:

- Students see themselves willingly using a tool such as the PT to practice for future presentations.
- Students find the feedback of the PT to be a good complement to the feedback that peers and tutors can give.
- Practicing with the PT leads to significant improvements in the overall quality of a presentation according to a human audience.

For future work we plan to show the results obtained in this study indicating the advantages of using the PT to coordinators of public speaking courses. This comes with a plan to deal with environmental constraints impeding the setup of PT and, hence, its use in the wild. Furthermore we plan to continue improving the PT. The purpose of the PT is to help humans give better presentation to humans. Hence, we plan to explore the relationship between human-based and machine-based assessment, and study how this information can later be used to provide learners with better feedback.

To conclude, there is still a lot of room for improvement for multimodal learning applications designed to support the development of public speaking skills. Introducing them to formal and non-formal educational scenarios still has some practical challenges. Though the application of the PT in a practical setting may not require equally strict conditions as in our research. In any case, studying the use of the PT in the wild has shown promising results regarding the support that such tools can bring to current practices for learning public speaking skills, indicating how courses on developing public speaking skills can be enhanced in the future.

**Acknowledgment.** The underlying research project is partly funded by the METALOGUE project. METALOGUE is a Seventh Framework Programme collaborative project funded by the European Commission, grant agreement number: 611073 (http://www.metalogue.eu).

## References


1. DeCaro, P.A.: Origins of public speaking. In: The public speaking project (2011). http://www.publicspeakingproject.org/psvirtualtext.html. Chapter 2
2. Parvis, L.F. The Importance of Communication and public-speaking skills. J. Environ. Health 35–44 (2001)
3. Campbell, K.S., Mothersbaugh, D.L., Brammer, C., Taylor, T.: Peer versus self-assessment of oral business presentation performance. Bus. Commun. Q. **64**(3), 23–42 (2001)
4. Hinton, J.S., Kramer, M.W.: The impact of self-directed videotape feedback on students' self- reported levels of communication competence and apprehension. Commun. Edu. **47**(2), 151–161 (1998)
5. Smith, C.M., Sodano, T.M.: Integrating lecture capture as a teaching strategy to improve student presentation skills through self-assessment. Act. Learn. High Educ. **12**(3), 151–162 (2011)
6. Gallo, C.: Talk Like TED: The 9 Public Speaking Secrets of the World's Top Minds. Pan Macmillan (2014)
7. Van Ginkel, S., Gulikers, J., Biemans, H., Mulder, M.: Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education. Educ. Res. Rev. **14**, 62–80 (2015)
8. Swan, M.: Sensor mania! the internet of things, wearable computing, objective metrics, and the quantified self 2.0. Journal of Sensor and Actuator. Networks **1**(3), 217–253 (2012)
9. Schneider, J., Börner, D., Van Rosmalen, P., Specht, M.: Augmenting the Senses: a re-view on sensor-based learning support. Sensors **15**(2), 4097–4133 (2015)
10. Barmaki, R., Hughes, C.E.: Providing real-time feedback for student teachers in a virtual rehearsal environment. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 531–537 (2015)
11. Damian I., Tan C.S.S., Baur T., Schöning J., Luyten K., André E.: Augmenting social interactions: realtime behavioural feedback using social signal processing techniques. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, pp. 565–574 (2015)
12. Dermody, F., Sutherland, A.: A multimodal system for public speaking with real time feedback. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 369–370 (2015)
13. Schneider, J., Börner, D., Van Rosmalen, P., Specht, M.: Presentation trainer, your public speaking multimodal coach. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 539–546 (2015)
14. Kerby, D. Romine, J.: Develop oral presentation skills through accounting curriculum design and course- embedded assessment. J. Educ. Bus. **85**(3). (2009)
15. Wörtwein, T., Chollet, M., Schauerte, B., Morency, L. P., Stiefelhagen, R., Scherer, S.: Multimodal public speaking performance assessment. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 43–50 (2015)
16. Schneider, J., Börner, D., Van Rosmalen, P., Specht, M.: Stand tall and raise your voice! a study on the presentation trainer. In: Proceedings of the tenth European Conference on Technology enhanced learning, EC-TEL2015, pp. 311–324 (2015)
17. Shields, P.M., Rangarajan, N.: A playbook for research methods: Integrating conceptual frameworks and project management. New Forums Press, Stillwater (2013)
18. Mitchell, V.W., Bakewell, C.: Learning without doing enhancing oral presentation skills through peer review. Manage. Learn. **26**(3), 353–366 (1995)

# How to Quantify Student's Regularity?

Mina Shirvani Boroujeni(✉), Kshitij Sharma, Łukasz Kidziński, Lorenzo Lucignano, and Pierre Dillenbourg

EPFL-CHILI, Lausanne, Switzerland
{mina.shirvaniboroujeni,kshitij.sharma,lukasz.kidzinski, lorenzo.lucignano,pierre.dillenbourg}@epfl.ch

**Abstract.** Studies carried out in classroom-based learning context, have consistently shown a positive relation between students' conscientiousness and their academic success. We hypothesize that time management and regularity are main constructing blocks of students' conscientiousness in the context of online education. In online education, despite intuitive arguments supporting on-demand courses as more flexible delivery of knowledge, completion rate is higher in the courses with rigid temporal constraints and structure. In this study, we further investigate how students' regularity affects their learning outcome in MOOCs. We propose several measures to quantify students regularity. We validate accuracy of these measures as predictors of students' performance in the course.

**Keywords:** Regulation · Self-regulation · Time management · Massive open online courses · Procrastination · Engagement

## 1 Introduction

Massive Online Open Courses allow millions of students from all over the world to participate in top quality courses on-line. Due to a great number of distractions in the environment where MOOCs are usually watched, it is more difficult to grasp learners' attention in a MOOC than in a classroom.

In this paper we present a quantitative framework which simplifies analysis of time-related behaviours. From the full spectrum of variables reflecting consciousness, we focus on regularity of a student. We investigate three key dimensions of regularity: intra-course, intra-week and intra-day as well. The intra-course regularity refers to the repetitive participation in the lectures and responsiveness to course-related events, intra-week corresponds to participation on the same day(s) of the week whereas intra-day corresponds to daily behavioural pattern.

We hypothesize that there are two strategies for participating in MOOCs. First, regular scheduling of learning activities; and second adaptive scheduling of the learning activities based on the daily work or study schedule. The learners affirming to the first strategy will have higher values for our definitions of regularity than the ones following the later strategy. In the current work we investigate if the regularity is a predictive of performance in MOOCs context.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 277–291, 2016.
DOI: 10.1007/978-3-319-45153-4_21

Our study is motivated by previous results on engagement. Behaviours inducing a habit are considered as a key to success of many on-line platforms [4]. Similarly, inducing a habit of participation in an on-line course can indicate a success of the course and of the platform. Second, in our previous studies we found that time management is dependent on employment status [20]. Analysis of regularity can allow us to further understand student's employment needs and opportunities. In this context employment can be seen as an external factor as described in a hypothetical model in Fig. 1.

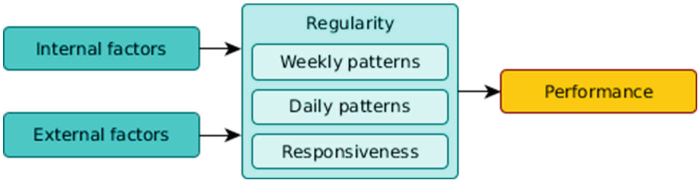



**Fig. 1.** We analyze regularity as a factor explaining performance, influenced by external and internal variables.

We hypothesize that regularity is one of the key factors related to student's success. In particular, we will answer following research questions:

**Question 1.** How can we quantify regularity of a student?
**Question 2.** Is regularity related to performance?

The key contribution of this paper is the definition of different measures of regularity and analysis of their properties. These measures can serve as indicators for quantifying to what extent certain features of a course or platform influences regularity and engagement of participants, or can be used to compare the courses and MOOC platforms regarding their habit inducing properties. Moreover, as we show in Sect. 6.4 the regularity features can be employed to predict users' performance.

## 2 Related Work

The importance of time management for succeeding in MOOC is highlighted in previous studies [3,15]. Recent studies show that difficulty with keeping up to deadlines is the main obstacle for engaging in a course [8]. In this section, we analyze regularity in the context of consciousness, review measures of regularity which can potentially be used in MOOCs and analyze the link between regularity and performance.

### 2.1 Conscientiousness and Self-Regulation

Early educational psychologists hypothesized that self-regulation is a key contributor to the academic success of students and it has since been verified [26]. Students' personalities also affect their academic success. The main factor that has been found to be correlated with students' performance is conscientious [16,19,23]. [16] in a review showed that from 33 different studies, examining the relation between the personality factors and academic success (GPA, course grade, average grade, exam score, thesis success), 21 found a significant correlation between conscientiousness and academic success. In two different meta analyses, [23] and [19] showed that the correlation between conscientiousness and academic success is also significant at the university level education.

Procrastination, defined as the tendency to delay of the task completion [11] has also been found to be correlated with the academic success. Klassen and colleagues [9] found that the students with negative procrastination had significantly lower GPA scores. Solomon and Rothblum [22] found that the students who reported higher levels of procrastination attempted significantly lower number of self-paced quizzes. Moreover, Ferrari and Ware [5] found that the task aversiveness was correlated to the self-reported procrastination of the students.

The main feature of both the self-regulative learning strategy and conscientious (in learning context) is organizing and planning learning goals. Time management and regularity are the key constituents for both the aforementioned factors. Thus, we hypothesize that there might exist a correlation between the MOOC performance and students' regularity.

### 2.2 Time Series Analysis

Time series analysis provides us with technical tools to assess regularity. Our main reference for elementary time series techniques is [2]. We can consider regularity as a seasonal component of a time series and take advantage of tools designed for quantifying seasonality. In classical time series analysis researchers often remove this seasonal pattern and focus on modeling the remaining behaviour of the process. In our case, since the pattern varies between the subjects, it becomes a characteristic of interest discriminating students.

We focus on two key approaches, time domain methods and frequency domain methods. To use time domain methods we slice the time series into segments of the length of interest (e.g. day, week) and compare repeatability of the slices [6,24]. In particular, we use Jensen-Shannon divergence to analyse a histogram of a segmented signal [13]. Frequency domain methods are based on the fact that inner product of a signal with and a periodic function is large if the signal has the same period [21]. Statistical tools have been developed to analyze if the signal on a given frequency is significant [18].

### 2.3 Performance Prediction

Student's performance is one of the key metrics analyzed in MOOCs. Many studies chose performance as an indicator for showing the value of the categorization methods. Massive datasets allow us to discover relation between performance and even the smallest factors like the number of pauses during watching a MOOC video or ratio of a video replayed [12]. Performance is also a crucial indicator for policy makers and MOOC practitioners. Reports focus on performance of MOOCs as a function of performance of students [14].

In previous studies, measures such as time spent on lecture, homework, forum, quiz and assignments were used to predict students' learning gain [10,25]. Lauria et al. [10] used the amount of content viewed, forum read, number of posts, assignments and quizzes submitted, to predict the performance and the engagement of the students. Other attempts to predict the performance root from the Social Network Analysis of the forum actions of the students. For instance, [17] used the network density, efficiency, individual student's contribution, in- and out- degrees, richness of the content, to find the correlations with engagement and performance. Regarding the analysis of timing patterns, Wolff et al. [25] used the temporal clickstream data to predict students' performance; similarly, Kennedy et al. [7] used number of submissions and active days (submitting days) to predict the final grades of the students in a programming MOOC. Likewise, we focus on the temporal regularity of students' activities, contributing in defining novel measurements for the regularity and showing their link with performance.

## 3 Methodology

The main steps towards assessing the regularity level of a student are defining what is considered as a regular behaviour and providing methods to capture such behaviour. Regularity in the context of MOOCs can be defined in two domains, actions and time, or a combination of the two. Regularity in actions is evident as repeating patterns in user's actions sequence (e.g. a student who watches the lecture and views the forum before doing an assignment), whereas regularity in time corresponds to repeating patterns in timing of study sessions (e.g. student who studies MOOCs on particular days or times). Regularity in the combined domain on the other hand is reflected by the dependencies between action types and their occurrence time (e.g. student who watches the lecture on Mondays and works on the assignments on Fridays).

In this work, we focus on time regularity. We aim to provide methods for quantifying regularity level of students considering the timing of their activities throughout the course. Regularity in time may emerge in different patterns. We consider six patterns of regularity listed in Table 1 and in Sect. 4 introduce measures to capture these patterns.

Note the difference between P3 and P4 in Table 1, which is the focus on relative (P3) and absolute (P4) amount of participation time on different weekdays. An example for P3 is a student who spends relatively more time on the course on Mondays compared to Tuesdays and Wednesdays, while example of P4 is a student who spends six hours on Mondays, four hours on Tuesdays and two hours on Wednesdays. Therefore P5 and P4 are subsets and more restricted forms of P3.

**Table 1.** Regularity patterns in time domain

| ID | Description |
|---|---|
| P1 | Studying on certain hours of the day |
| P2 | Studying on certain day(s) of the week |
| P3 | Studying on similar weekdays, over weeks of the course |
| P4 | Same distribution of study time among weekdays, over weeks of the course |
| P5 | Particular amount of study time on each weekday, over weeks of the course |
| P6 | Following the schedule of the course |

## 4 Design of Measures

Table 2 provides an overview of our proposed measures and the regularity patterns they reflect. In the following we present problem formulation and detailed description of the measures.

**Table 2.** Regularity measures and corresponding regularity patterns

| Measure | Description | Dimension | Pattern |
|---|---|---|---|
| PDH | Peak on day hour | Intra-day | P1 |
| PWD | Peak on week day | Intra-week | P2 |
| WS1 | Weeks similarity measure 1 | Intra-week | P3 |
| WS2 | Weeks similarity measure 2 | Intra-week | P4 |
| WS3 | Weeks similarity measure 3 | Intra-week | P5 |
| FDH | Periodicity of day hour | Intra-day | P1 |
| FWH | Periodicity of week hour | Intra-week | P1 |
| FWD | Periodicity of week day | Intra-week | P2, P3 |
| DLV | Delay in lecture view | Intra-course | P6 |

### 4.1 Problem Formulation

Let $n$ be the number of events by the user and $T = \{t_1, t_2, ..., t_n\}$ be the set of timestamp of events. We assume minutes as a unit of time and set $t = 0$ when the course starts. Let $L_m$, $L_d$ and $L_w$ be the course length (time from course release till the deadline of the final assignment) in minutes, days and weeks respectively. We can treat user's activity time series as a binary signal defined as (examples in Fig. 4)

$$F_W(x) = \begin{cases} 1 & \text{if } \exists t_i \in T : x = \left\lfloor \frac{t_i}{W} \right\rfloor \\ 0 & \text{otherwise} \end{cases}, \text{ where } x \in \{1, 2, ..., L_m/W, \}$$

where $W$ is the length of a time window in minutes.

Based on this definition, $F_{60}(x) = 1$ implies that user had at least one action at hour $x$ after the course start and $F_{60\times24}(x) = 1$ indicates at least one action at day $x$ of the course.

### 4.2 Time Based Measures

We define two measures, **PDH** and **PWD**, based on the entropy of the histogram of user's activitiy over time. PDH identifies if user's activities are concentrated around a particular hour of the day and PWD determines if activities are concentrated around a particular day of the week.

We define function $D(h)$ on every hour of a day, and function $W(d)$ on on every day of a week as

$$D(h) = \sum_{i=0}^{L_d-1} F_{60}(24i + h), \text{ where } h \in \{0, 1, ..., 23\}.$$

$$W(d) = \sum_{i=0}^{L_w-1} F_{60\times24}(7i + d), \text{ where } d \in \{0, 1, ..., 6\}.$$

Therefore $D(h)$ corresponds to the number of days in which user was active at hour $h$ of the day, and $W(d)$ represents the number of weeks in which user was active at day $d$. See examples of these two functions in Fig. 2.

Although resulting histograms are already informative, they still distinguish the time on which regularity appears. In order to define a measure invariant to the time of regularity, we focus on spikes. The popular measure which identifies if given distribution is uniform or has a spike is entropy. Based on its definition, we suggest daily and weekly entropy as

$$E_D = -\sum_{h=0}^{23} \hat{D}(h)\log(\hat{D}(h)), \qquad E_W = -\sum_{d=0}^{6} \hat{W}(d)\log(\hat{W}(d)),$$

where $\hat{D}$ and $\hat{W}$ are normalized histograms.

A small entropy value encodes presence of spikes in the distributions. However, since entropy is computed on the normalized histogram, it does not reflect the magnitude of the spike in the original histogram. To overcome this limitation, we define two regularity measures, PDH and PWD as

$$PDH = (log(24) - E_D) \max_h D(h), \qquad PWD = (log(7) - E_W) \max_d W(d).$$

Therefore PDH is bounded in $[0, log(24).L_d]$ and PWD is bounded in $[0, log(7).L_w]$. A high value of PDH or PWD measure respectively implies a strong spike in $D(h)$ or $W(d)$.

### 4.3 Profile Similarity

We define three measures **WS1**, **WS2** and **WS3** based on the similarity between weekly profiles of user's activities. WS1 measures if the user works on the same weekdays. WS2 compares the normalized profiles and measures if user has a similar distribution of workload among weekdays, in different weeks of the course. Whereas, WS3 compares the original profiles and reflects if the time spent on each day of the week is similar for different weeks of the course. In the following we describe the construction of weekly profiles and the three similarity functions used to compare them.

We define activity profile of a user during week $k$ as the following vector (examples in Fig. 3).

$$P(k) = [P(1,k), P(2,k), ..., P(7,k)]^T, \text{ where } k \in \{0, 1, ..., L_w\},$$

where $P(d,k)$ represents the number of hours user was active in day $d$ of week $k$ and is defined as

$$P(d,k) = \sum_{i=0}^{23} F_{60}(24(d+7k)+i), \text{ where } d \in \{0,1,...,6\},\ k \in \{1,2,...,L_w\}.$$

**Similartiy Measure 1:** Let $Active(k)$ be the set of days in week $k$, on which the user had some activity. We define the first profile similarity measure as

$$Sim1(P(i), P(j)) = \frac{\|Active(i) \cap Active(j)\|}{max(\|Active(i)\|, \|Active(i)\|)}$$

Therefore for two weeks in which the user is active on exactly same days, this similarity measure returns the maximum value (1).

**Similarity Measure 2:** The second profile similarity measure compares the normalized profiles ($\hat{P}(k)$) of two weeks based on Jensen-Shannon divergence (JSD) as

$$Sim2(\hat{P}(i), \hat{P}(j)) = 1 - \frac{JSD(\hat{P}(i), \hat{P}(j))}{log(2)}$$

$$\mathcal{JSD}(P_1, P_2, ...P_n) = H\Big(\sum_{i=1}^{n} \pi_i P_i\Big) - \sum_{i=1}^{n} \pi_i H(P_i),$$

where $\pi_i$ is the selected weight for the probability distributions $P_i$ and $H(P)$ is the entropy for distribution $P$. We consider uniform weights for all weeks, hence $\pi_i = 1/n$. The value of Sim2 is bounded in $[0,1]$ and high value of this measure reflects similar shapes of activity profiles in the weeks of comparison.

**Similartiy Measure 3:** In order to capture the similarity in shape and magnitude of weekly profiles, we define the third similarity function, based on $\chi^2$ divergence as

$$Sim3(P(i), P(j)) = 1 - \frac{1}{\|Active(i) \cup Active(j)\|} \sum_{d=1}^{7} \Big(\frac{P(d,i) - P(d,j)}{P(d,i) + P(d,j)}\Big)^2$$

Therefore the highest similarity value (1) is achieved if the two profiles are identical. Finally we define three regularity measures WS1, WS2 and WS3 as the average of pairwise similarity of weekly profiles computed by $Sim1$, $Sim2$ and $Sim3$ respectively.

### 4.4 Frequency Based Measures

One common approach to detect seasonal components of a signal is to convert the signal ($X(t)$) from its original domain (often time or space) to a representation in the frequency domain ($\mathcal{F}(\theta)$) by applying Fourier transform. Fourier transform of a signal $X(t)$ is defined as

$$\mathcal{F}(\theta) = \sum_{t=-\infty}^{\infty} X(t)e^{(-2\pi i\theta t)}$$

The function $\mathcal{F}(\theta)$ is referred to as spectral density or periodogram, and is used to detect any periodicity in the data, by observing peaks at the frequencies corresponding to these periodicities. For the purpose of detecting weekly or daily regularity, we compute spectral density of user's time signals ($F_{60}(x)$ and $F_{24\times 60}$ defined in Sect. 4.1) and in the resulting periodogram, extract values corresponding to daily and weekly periods. We expect a high value for the resulting measures in case there is a daily or hourly repeating pattern in user's activities over time.

We propose three frequency based measures, **FDH**, **FWH** and **FWD** as

$$FDH = \mathcal{F}_h(1/day), \qquad FWH = \mathcal{F}_h(1/week) \qquad FWD = \mathcal{F}_d(1/week)$$

$$\mathcal{F}_h(\theta) = FFT(F_{60}(x)), \qquad \mathcal{F}_d(\theta) = FFT(F_{24\times 60}(x))$$

FDH measures the extent to which the hourly pattern of user's activities is repeating over days (e.g. the user is active at 8 h–10 h and 12 h–17 h on every day). FWH identifies if the hourly pattern of activities is repeating over weeks (e.g. in every week, the user is active at 8 h–10 h on Monday, 12 h–17 h on Tuesdays, etc.). FWD captures if the daily pattern of activities is repeating over weeks (e.g. the user is active on Monday and Tuesday in every week).

### 4.5 Adherence to Course Schedule

Some students watch the lecture right after it is released whereas others postpone watching lectures or submitting assignments. Therefore some users are regular not because of a weekly routine, but they follow the schedule of the course. To capture adherence to the course schedule, we define **DLV** measure as the average delay in viewing video lectures

$$DLV = \frac{1}{m}\sum_{i=1}^{m}(FirstView(i) - Release(i)),$$

where $m$ is the number of video lectures user has watched. We then normalize DLV by the length of the course to get a value in $[0, 1]$.

## 5 Dataset

Our analysis is based on an undergraduate engineering MOOC offered in Coursera entitled *"Functional Programming Principles in Scala"*. Total duration of the course was 10 weeks and lectures were released on a weekly basis. The final grade was calculated based on six graded assignments and passing grade was 60 out of 100. The initial dataset contained events by a total of 28,002 participants. In the data preparation phase, we removed inactive users, namely those who had less than two weeks with at most four actions of any type (13,102 users). Users who did not submit any assignments were also considered as inactive and hence removed from the dataset (4,644 users). Some participants, never watched a video on the platform, instead they downloaded the lectures and probably watched them offline. Since activity traces for such users is not available, we removed them from the dataset as well (225 users). Therefore, in our analysis we considered all events by remaining 10,031 participants. Their average grade was 55.7 and 51 % scored higher than the passing threshold (60).

## 6 Results

We computed the proposed regularity measures for participants in the dataset. Table 3 provides an overview of the computed values.

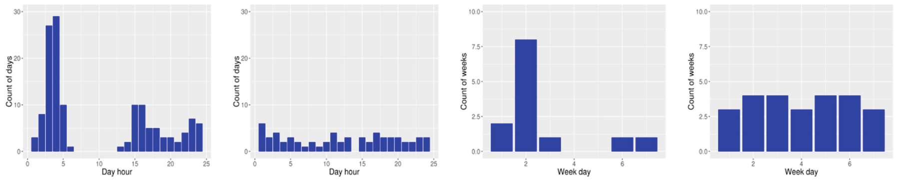
**Table 3.** Overview of regularity measures in the dataset

| Measure | Mean | Max | SD | Measure | Mean | Max | SD |
|---|---|---|---|---|---|---|---|
| PDH | 4.65 | 49.92 | 3.65 | FDH | 0.34 | 14.65 | 0.64 |
| PWD | 1.12 | 13.62 | 1.08 | FWH | 0.17 | 4.2 | 0.25 |
| WS1 | 0.14 | 0.90 | 0.13 | FWD | 0.36 | 4.64 | 0.35 |
| WS2 | 0.17 | 0.88 | 0.15 | DLV | 0.14 | 0.95 | 0.11 |
| WS3 | 0.11 | 0.74 | 0.10 | | | | |

### 6.1 Regularity Measures Examples

In the following we present examples of proposed features to verify if they capture the regularity patterns as expected.

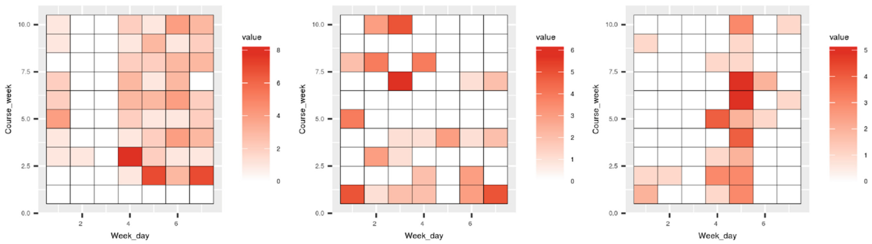
**PDH and PWD:** Figure 2 illustrates examples of users with high and low value of PDH and PWD measures. Histograms in Fig. 2a and b represent the number of days at which user was active on a particular hour, and Fig. 2c and d show the number of weeks at which user was active on a particular day. Clearly, high value PDH an PWD, represent peak of activity in partciular hour(s) or day(s) and hence they capture regularity patterns P1 and P2 respectively.




(a) PDH=20 (b) PDH=0.7 (c) PWD=6.13 (d) PWD=0.04

**Fig. 2.** PDH and PWD measures: examples of two users with high and a low values. Clearly a high value reflects a spike in the signal.

**WS1, WS2 and WS3:** Figure 3 provides examples of weekly activity profiles of three students. In the profile matrix, columns represent weekdays, rows represent week of the course and color intensity encodes amount of study time (hours) on a particular day. As it can be perceived form the profile in Fig. 3a, the activities of first user are clearly concentrated on the second half of the week, whereas no regular pattern is evident in weekly activities of the second user in Fig. 3b. All three profile similarity measures return a high value for the first case (regular) and obtain a low value for the second (not-regular). Figure 3c provides an example highlighting the difference between these three measures. The third user dedicates relatively more time on day five compared to the other days (high value of WS2), but the amount of study hours on this day varies between weeks (relatively lower value of WS3).




(a) 0.68, 0.66, 0.57 (b) 0.17, 0.2, 0.13 (c) 0.5, 0.69, 0.37

**Fig. 3.** WS1, WS2 and WS3 measures: weekly activity profiles of users with high and low values. Values below each chart correspond to WS1, WS2 and WS3 respectively.

**FDH, FWH, FWD:** Figure 4 illustrates examples of users with high and low value of FWD measure. As it can be inferred from the time signal (left) in the first row, user's activities follow a periodic weekly pattern which is also reflected by a large value (3.64) at the frequency corresponding to one-week period on the frequency domain chart (right). On the contrary, no seasonal pattern is evident in user's time signal in the second row and consequently FWD obtains a small value (0.04). FDH and FWH measure also follow the same principle.

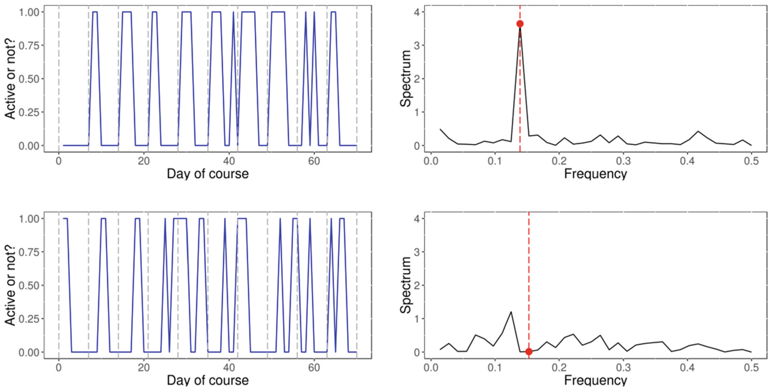



**Fig. 4.** FWD measure: Examples of activities of two users in time (left) and frequency domain. FWD=3.64 for the first row and FWD=0.04 for the second.

### 6.2 Correlation Between Measures

The profile similarity measures WS1, WS2 and WS3, although sensitive to different activity profiles (Fig. 3c), result to have strong correlation in pairwise comparison ($r = 0.9$, $p < 0.01$). FWD measure is also moderately correlated with profile similarity measures ($r = 0.57$, $p < 0.001$). The remaining set of measures are not strongly correlated with each other inferring that they capture orthogonal patterns of regularity.

### 6.3 Clustering Users Based on Regularity Measures

Based on calculated regularity measures, we clustered users into three categories using hierarchical clustering method with euclidean distance metric. Number of clusters was chosen based on the resulting dendogram. Figure 5 presents an overview the three clusters and average grade of users in each group (values were scaled to [0,1] for visualization). The three clusters clearly differ in terms of average grade. Users in the second cluster have the highest regularity according to all measure (except PWD and DLV) and score higher as well. The first and third cluster have very similar regularity values; however users in the third cluster have relatively longer delays in watching video lectures which could explain their lower average grade. Another possible explanation could be that the third cluster contains late-comers in the course who fail to meet the course deadlines. Further investigation of the users activities is required verify these hypothesis.

### 6.4 Predictive Power of Regularity Measures

In this section we analyze the link between regularity and performance, as presented in Fig. 1. Analysis of correlations between final grade and regularity measures, reveal that final grade is strongly correlated with WS2 ($r = 0.70, p < 0.001$),

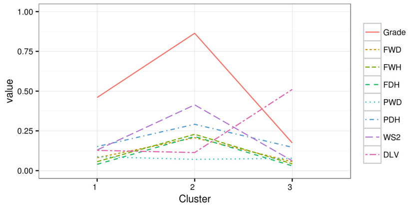



**Fig. 5.** Average value of regularity measures in each cluster.

FWD ($r = 0.46$, $p < 0.001$), moderately correlated with FWH ($r = 0.37$, $p < 0.001$) and FDH ($r = 0.32, p < 0.001$), slightly correlated with PDH ($r = 0.25$, $p < 0.001$), DLV ($r = -0.25$ , $p < 0.001$) and not correlated with PWD measure.

In order to analyze predictive power of the regularity features we build a linear model including all of them and we use penalized regression to improve the model by removing features of low importance. In our dataset, linear model with variables FDH, WS2 and DLV has $R^2 = 0.52$, which assures us about predictive potential of designed variables.

### 6.5 Other Applications of Regularity Measures

As an example of another application, we investigate the link between regularity and external factors, as presented in Fig. 1. Motivated by our previous results [20], we analyze the employment status. The database contains employment information for about 9.6 % of the participants. Based on these information we extract two categories of users: *full-employed* and *full-students* (559 v.s. 113 users). We assume that users in both categories have a daily or weekly routine imposed by their occupation or school schedule. Considering the time regularity, employed participants have higher regularity in weekly and daily basis. This is reflected by significantly higher value of WS2 measure for employed users ($m = 0.17$ v.s $m = 0.14$, $F[1, 670] = 4.8$, $p = 0.02$), higher value of FWD measure ($m = 0.38$ v.s. $m = 0.3$, $F[1, 670] = 4.2$, $p < 0.05$) and higher values for PDH measure ($m = 0.4.8$ v.s. $m = 3.6$, $F[1, 670] = 9.16$, $p < 0.01$).

## 7 Conclusions

The key objective of this study was to quantify students' regularity **(Question 1)**. By employing time domain [6,24] and frequency domain [21] techniques, we defined nine measures corresponding to regularity patterns on three dimensions: intra-day, intra-week and intra-course. Investigation of students' activities corresponding to low and high values of these measures illustrates their behaviour. We showed that a subset of the measures are not strongly correlated with each other, providing high predictive power.

We find that regularity is related to performance **(Question 2)**. The predictive power of suggested variables is encouraging for four reasons. First, our proposed measures are general and can be defined outside MOOCs' context. Second, they explain over 50 % of the grade variability, so they can be included in existing performance models. As in previous studies we verify that temporal patterns have significant predictive potential [25]. Third, features are not strongly correlated with each other. Fourth, although our analysis is a posteriori, features which we propose can be estimated throughout the course.

Positive correlation between the defined regularity measures and the performance of the students, supports the hypothesis that students who plan their learning activities in a regular manner have better chances of succeeding in the MOOC [3,15]. There are two plausible explanations for the fact that regularity is predictive of performance in the MOOC. First, regular student follows the structure of the course and therefore attains higher achievement. Second, having high regularity is related to certain factors internal to the students, i.e., motivation, commitment or learning strategies [1,26]. In the future work emerging from this contribution, we will attempt to capture the different factors influencing regularity in the students who have higher values of regularity measures.

Finally, the regularity measures we defined, allowed us to confirm the impact of external factors on regularity patterns [20]. We found that employed learners are more regular both on weekly and daily scales than the unemployed or university students. This application of the measures supports our claim that they can be used in practice to measure effects of interventions on user habits and to compare engagement between courses or platforms.

One limitation of the regularity measure we proposed is that, using our measures one cannot distinguish between the different strategies used by those students who adaptively plan their learning activities. Moreover, as any projections, our measures can only discriminate patterns that they were designed for and should be combined for accurate assessment of regularity. These limitations also enlighten the future work of this contribution.

## References

1. Blair, C., Diamond, A.: Biological processes in prevention and intervention: the promotion of self-regulation as a means of preventing school failure. Dev. Psychopathol. **20**(03), 899–911 (2008)
2. Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods. Springer Science & Business Media, New York (2013)
3. Dillenbourg, P., Li, N., Kidziński, Ł.: The complications of the orchestration clock. In: From Books to MOOCs? Emerging Models of Learning and Teaching in Higher Education. Portland Press (2016)
4. Eyal, N.: Hooked: How to Build Habit-Forming Products. Penguin Canada, Toronto (2014)
5. Ferrari, J.R., Ware, C.B.: Academic procrastination: personality. J. Soc. Behav. Pers. **7**(3), 495–502 (1992)
6. Jönsson, P., Eklundh, L.: Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans. Geosci. Remote Sens. **40**(8), 1824–1832 (2002)

7. Kennedy, G., Coffrin, C., de Barba, P., Corrin, L.: Predicting success: how learners' prior knowledge, skills and activities predict mooc performance. In: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, pp. 136–140. ACM (2015)
8. Kizilcec, R.F., Halawa, S.: Attrition and achievement gaps in online learning. In: Proceedings of the Second ACM Conference on Learning@Scale, pp. 57–66. ACM (2015)
9. Klassen, R.M., Krawchuk, L.L., Rajani, S.: Academic procrastination of undergraduates: low self-efficacy to self-regulate predicts higher levels of procrastination. Contemp. Educ. Psychol. **33**(4), 915–931 (2008)
10. Lauría, E.J., Baron, J.D., Devireddy, M., Sundararaju, V., Jayaprakash, S.M.: Mining academic data to improve college student retention: an open source perspective. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 139–142. ACM (2012)
11. Lay, C.H.: At last, my research article on procrastination. J. Res. Pers. **20**(4), 474–495 (1986)
12. Li, N., Kidziński, Ł., Jermann, P., Dillenbourg, P.: MOOC video interaction patterns: what do they tell us? In: Conole, G., Klobučar, T., Rensing, C., Konert, J., Lavoué, E. (eds.) Design for Teaching and Learning in a Networked World. LNCS, vol. 9307, pp. 197–210. Springer, Heidelberg (2015)
13. Lin, J.: Divergence measures based on the shannon entropy. IEEE Trans. Inf. Theor. **37**(1), 145–151 (1991)
14. McAuley, A., Stewart, B., Siemens, G., Cormier, D.: The MOOC model for digital practice (2010)
15. Nawrot, I., Doucet, A.: Building engagement for MOOC students: introducing support for time management on online learning platforms. In: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, pp. 1077–1082 (2014)
16. OConnor, M.C., Paunonen, S.V.: Big five personality predictors of post-secondary academic performance. Pers. Individ. Differ. **43**(5), 971–990 (2007)
17. Paredes, W.C., Chung, K.S.K.: Modelling learning & performance: a social networks perspective. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 34–42. ACM (2012)
18. Percival, D.B., Walden, A.T.: Spectral Analysis for Physical Applications. Cambridge University Press, Cambridge (1993)
19. Poropat, A.E.: A meta-analysis of the five-factor model of personality and academic performance. Psychol. Bull. **135**(2), 322 (2009)
20. Boroujeni, M.S., Kidziński, Ł., Dillenbourg, P.: How employment constrains participation in MOOCS? In: Proceedings of the 9th International Conference on Educational Data Mining, pp. 376–377 (2016)
21. Christopher, A.: Sims: seasonality in regression. J. Am. Stat. Assoc. **69**(347), 618–626 (1974)
22. Solomon, L.J., Rothblum, E.D.: Academic procrastination: frequency and cognitive-behavioral correlates. J. Couns. Psychol. **31**(4), 503 (1984)
23. Trapmann, S., Hell, B., Hirn, J.-O.W., Schuler, H.: Meta-analysis of the relationship between the big five and academic success at university. Zeitschrift für Psychologie/J. Psychol. **215**(2), 132–151 (2007)
24. Vetterli, M., Kovačević, J., Goyal, V.K.: Foundations of Signal Processing. Cambridge University Press, Cambridge (2014)

25. Wolff, A., Zdrahal, Z., Nikolov, A., Pantucek, M.: Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In: Proceedings of the Third International Conference on Learning Analytics and Knowledge, pp. 145–149. ACM (2013)
26. Zimmerman, B.J.: Investigating self-regulation and motivation: historical background methodological developments and future prospects. Am. Educ. Res. J. **45**(1), 166–183 (2008)

# Nurturing Communities of Inquiry: A Formative Study of the DojoIBL Platform

Ángel Suárez(✉), Stefaan Ternier, Fleur Prinsen, and Marcus Specht

Welten Institute, Open University of the Netherlands, Heerlen, The Netherlands
{angel.suarez, stefaan.ternier, fleur.prinsen, marcus.specht}@ou.nl

**Abstract.** This formative study introduces DojoIBL, a web-based platform to support collaborative inquiry-based learning processes. By supporting communication and collaboration with new technological affordances, DojoIBL aims at nurturing communities of inquiry. The study elaborates on the theoretical underpinning of DojoIBL, describes its added value and presents a detailed explanation about the functionalities supported. Thereafter, an evaluation about how users perceived DojoIBL has been performed. Besides, the positive acceptance of participants, the results also showed that DojoIBL seems to be a suitable tool to support essential components of communities of inquiry. The study concludes anticipating the integration of role support as future developments of DojoIBL.

**Keywords:** Inquiry-based learning · Community of Inquiry · Collaborative knowledge building · Context-awareness · Informal learning

## 1 Introduction

In recent years, there has been an increasing interest in socio-constructivist learning methods e.g., (mobile) inquiry-based learning (IBL) [1], as well as the technological tools that support them [2]. IBL is often characterized as a collaborative process, in which informal and formal learning activities are socially interconnected. These activities need to be seamlessly supported in order to provide an effective and complete experience to the students. The collaborative inquiry process was aptly defined in the 'Community of Inquiry' approach [3], which emphasizes that creation of knowledge requires social interactions of individuals with different background knowledge.

However, there is still a lack of research on the technological affordances to enhance the IBL process and nurture a community of inquiry. For instance, the power of cloud based services in combination with instant communication or notifications have not been entirely explored in the context of inquiry-based learning. Previous studies conducted in the context of the weSPOT European project [4][1], a three-year project in which experience and knowledge about IBL have been acquired, showed that there were issues integrating and using technology in collaborative IBL processes. These issues were related to the lack of adequate technological affordances nurturing

[1] http://inquiry.wespot.net/.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 292–305, 2016.
DOI: 10.1007/978-3-319-45153-4_22

the communities of inquiry as also the educational settings in which inquiry-based learning often is implemented. Certainly, teachers faced difficulties to encourage and to help students explore topics as a community also due to the complex different variations of inquiry-based learning from confirmation inquiry to open inquiries.

In our effort to study an affordable solution that combines the essential elements to support IBL with the added potential of new technological affordances, this research study contributes DojoIBL, a platform that focuses on supporting 'Community of Inquiry' (CoI).

In the first two sections we will elaborate the theoretical underpinnings of DojoIBL; existing IBL solutions and social collaborative tools are discussed, and the rationale to develop DojoIBL is explained. Next, the design principles of the DojoIBL are described. The added value of DojoIBL, as compared to other IBL solutions, is argued in section four. Thereafter, in section five and six, the research design of the study is introduced and the results of a study into DojoIBL user experiences are described. Section seven elaborates on the interpretation and discussion of the results. Finally, the conclusion and the future work of the DojoIBL platform are outlined.

## 2 Theoretical Framework

Inquiry-based learning is defined on the premise that learning is more than memorizing information, rather it is a process of understanding, developing inquiry skills and constructing knowledge sparked by curiosity [5]. Often, inquiry processes incorporate elements of collaboration, which was defined in [6] as the engagement of students in a common endeavor. Collaboration transforms the inquiry activities into processes of co-construction of knowledge around shared understandings or concepts. Collaborative inquiry learning has also been defined in [7] in its *Knowledge building* approach, as an unpredictable, holistic process of creative development of ideas within a community of learners [5]. Moreover, socio-constructivist learning theories stated that knowledge is materialized when people, with different background knowledge, collaborate to find answers to a problem.

**Community of Inquiry.** The definitions of collaborative inquiry-based learning, anticipated the concept of community in IBL. [3] coined the term 'Community of Inquiry' (CoI) to refer to a group of individuals (facilitators and students) transacting with the specific purposes of facilitating, constructing, validating understanding and developing capabilities leading to further learning. In other words, the CoI framework is concerned with the nature of knowledge formation in IBL. [8] already defined it as a continuous exploration of a topic of students' interest, where community members (students) engage in social interactions to generate shared understanding. It has been shown in the literature that text-based communications have a considerable potential to facilitate the creation of communities of inquiry (CoI) [9, 10]. As already mostly evident in the definition given in [11], CoI comprises three essential components to any educational transaction: *cognitive presence*, which is defined as the capability of each participant in the CoI to construct meaning through sustained communication [9], *social presence* that relates to the ability of students to positioned themselves socially and affectively in the CoI [12] and *teaching presence*, which is characterized as the design, facilitation and direction of cognitive and social processes in order to produce meaningful co-creation of knowledge [13].

[14] emphasized the need to establish a common ground and perform in a community of practice (even broader than CoI) in order to work and learn efficiently. Notifications and awareness in collaborative activities can contribute to achieve this common ground [15]. [15] defined the three following types of collaboration awareness. *Social* awareness, relates to the presence of others working in parallel and it involves motivational or attitudinal aspects like timing, frequency or intensity. *Action* awareness copes with the idea that social awareness is not enough. Besides knowing who is around, students must be informed about what is happening. The last type, *activity* awareness, advices organizational and structural changes that helps students to understand the context of the inquiry activity.

**Social Collaboration Supported with Technology.** Research has shown that technology can support inquiry-based learning [16–18]. We attribute this to advancements in technology and its capacity to offer new possibilities for scaffolding the inquiry-based learning process. Premised on the theoretical framework of social constructivism, inquiry-based learning supports co-creation of knowledge through social interactions, between students-students and students-facilitators. Co-Lab [18], an online desktop environment offering an integrated approach for collaboration, modeling and inquiry, already addressed this to promote scientific discovery learning. Other developments such as nQuire [19][2], a software application to guide personal inquiry learning, or Go-Lab[3] [20] (through *Graasp*[4]) a project that provides guided experimentation that helps students acquiring inquiry skills, addressed collaboration. However, these platforms have not yet fully exploit emerging technological affordances. More recently, educational platforms like Edmodo[5] or ClassDojo[6] have enabled students to connect and to collaborate using cloud-based and social functionalities similar to the affordances of most popular social network platforms. Edmodo, is a social learning community where students, teachers and parents form communities or groups of their interest. It uses the timeline metaphor to display the latest posts in the communities or groups the user is following. The user's contributions are based on the following four types; *notes*, *assignments*, *quiz* or *polls*, which allow participants to connect around shared ideas. Comparable, ClassDojo is a communication platform that aims at encourage students to learn in a happier way engaging parents on the process. ClassDojo has three visualizations for the classroom; *class story*, a timeline visualization of the latest contributions, a *classroom* visualization where all the students are displayed facilitating students' rewarding and *messages* visualization to easily connect with others. Both initiatives provide resources to increase students' awareness and communication.

---

2 http://www.nquire.org.uk/home.

3 http://www.golabz.eu/.

4 http://graasp.eu/.

5 https://www.edmodo.com/.

6 https://www.classdojo.com/.

Group awareness has been an emerging topic in Computer Supported Collaborative Learning (CSCL) research [21]. Three types of awareness can be extracted from the above research studies; process, social and activity awareness [14, 15, 22]. Each of the studies focuses on helping students to visualize and manipulate social processes in order to understand how the group moves forward. Moreover, regarding communication, it has been proven in literature that text-based communication has a considerable potential to facilitate the creation of communities of inquiry (CoI) [3, 10].

To sum up, current platforms [19, 20] have sought to support the IBL process. These platforms have yet to fully harness the affordances of educational and social network platforms (e.g. ClassDojo and Edmodo) and emerging technological tools to support social collaboration and to nurture community of inquiries. Hence, based on existing initiatives and studies, this research explores the affordances of emerging technologies in the design of DojoIBL to foster communities of inquires. Essentially, it investigates how DojoIBL can facilitate social interactions and raise students' awareness of collaborative IBL processes.

## 3 Research Design

This research study introduces DojoIBL, a multi-device Learning Content Management System[7] (LCMS) to scaffold and to support students' collaborative knowledge co-construction process in IBL. Rather than delivering course content material, DojoIBL provides the designers and the facilitators with the tools to structure IBL processes about any meaningful topic from students' curiosity. Therefore, it focuses on the process rather than in the content itself. DojoIBL has been developed following a design-based research approach [23] in which teachers, designers and researchers collaboratively generate feedback feeding the iterative and incremental development process. Results of the weSPOT European project [4], showed that it is important to involve teachers in the early stages of the design and development process; giving us a broader perspective on the flexibility that the platform should have. The weSPOT project experiences and knowledge encouraged our team to develop DojoIBL, following several design principles that will be summarized.

The weSPOT project showed that students can be overwhelmed if the cognitive requirements demanded by our system are too high. Therefore, one of our aims was to reduce *extraneous cognitive load*, by ensuring that all elements included in DojoIBL add value to the learning experience. Thus, unnecessary information or elements that distract students from learning have been avoided in the interface, and visual representations of the inquiry process have been used to make the system more intuitive. Moreover, research studies on IBL [24, 25] exemplify the need to scaffold the inquiry learning process hence, DojoIBL breaks down the inquiry process into phases [25], and the phases into activities, in order to provide implicit guidance on the inquiry process.

Inquiry based Learning is a collaborative process [5, 7, 26] where students also learn from their peers by reflecting and building on top of one another's ideas.

[7] https://en.wikipedia.org/wiki/Content_management_system accessed on March 2016.

Hence, DojoIBL implements an instant messaging system supporting *cognitive presence*, *social presence* and *teaching presence* [11–13] which contributes to generate a Community of Inquiry [3, 10]. Yet, students per se are not skilled on acting as a community. Consequently, teachers' orchestration [27] and scaffolding remain essential [28], especially at early stages of the inquiry process. In addition to instant messaging, DojoIBL implements a notification system and an inquiry timeline, which facilitates asynchronous collaboration and raise awareness among students [15].

In short, DojoIBL focuses on adding value to the authentic inquiry experiences, providing an intuitive, simple and flexible tool that enables collaborative self-directed learning for students and just in context - time and place - orchestration for teachers (Fig. 1).

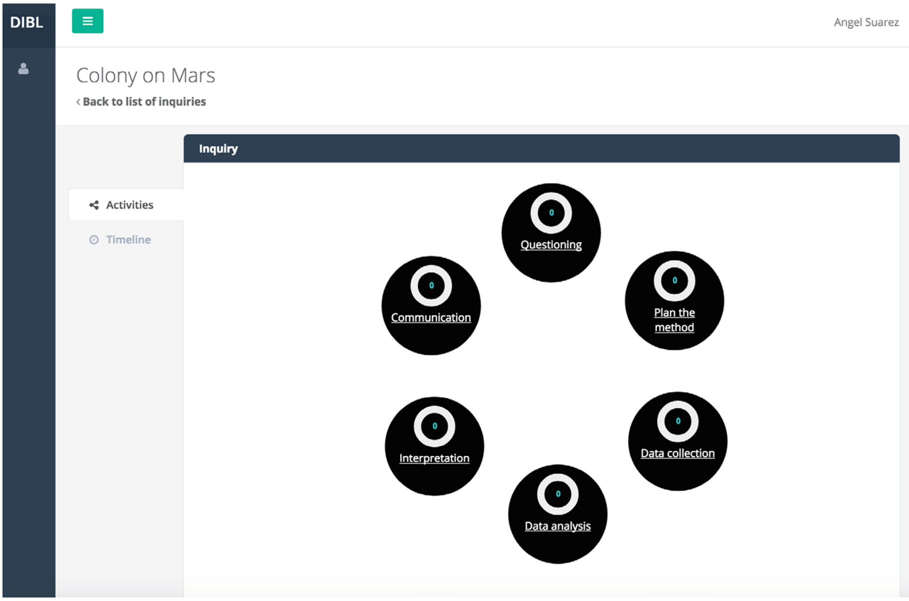



**Fig. 1.** Visualization of the inquiry process on the Colony on Mars activity

## 4 Affordances of DojoIBL

DojoIBL is an open source platform that builds on the ARLearn framework [29], a PaaS cloud based architecture deployed in Google App Engine (GAE). DojoIBL is a Learning Content Management System that provides *atomic* inquiry elements to structure collaborative inquiry processes. This section illustrates how the design challenges are addressed in DojoIBL, as well as discusses the added value of DojoIBL as compared to existing IBL solutions.

One of the main characteristic of DojoIBL is that users are able to design *blueprints* or *templates* for an inquiry structure. That means, several inquiries can be created based on the same *blueprint* or *template* of an inquiry structure. As a consequence, students can work in groups on different topics using a common inquiry structure. In addition, similar to what other educational platforms like Spiral.ac[8] or Edmodo[9] do, DojoIBL generates unique codes for each inquiry group. Consequently, managing and organizing students in inquiry groups can be reduced to share the specific codes with them. This functionality addresses one of the design requirements introduced before, simplicity.

Another design requirement highlights the necessity to work with intuitive designs and platforms that help students understand the inquiry process. The opportunity to practice, understand and master the steps needed to answer any given question helps students to be more self-directed learners and to be less dependent on facilitators' scaffolding. For instance, existing solutions like nQuire, uses visual representations of the inquiry cycle. In DojoIBL, inspired by those existing solutions, an interactive visualization of the inquiry structure is used (Fig. 1). This visualization builds on the IBL model [4] and represents every inquiry phase as a cycle, that when clicked opens the activities related to this phase.

DojoIBL aims at supporting authentic and transformative [30] inquiry learning processes. Rather than teachers providing the conceptual knowledge, IBL relies on teachers orchestrating and scaffolding the process using different strategies or structures [30]. To help students achieve higher order thinking and to create opportunities for students to develop their inquiry skills and their own understanding around questions, DojoIBL uses atomic inquiry elements. An atomic inquiry element is defined as the smallest re-usable type of activity that can be added to an inquiry phase. Currently, there are six types of activities available in DojoIBL, and each type provides a specific pedagogical affordance:

- The *research question* is an essential part of IBL where students collaboratively work around a shared question or topic. It aims at developing critical thinking skills [9, 11, 31], and it must be supported with tools to generate individual discussions, which enables self-directed learning as each student can create his/her own question, and other can contribute to it.
- *Discussion* forms the simplest type of activity which is based on plain text. Students can find a description, a story or a definition that inspire them about the specific topic. Activities are flexibly enabling any kind of activity design. For example, activities inform the student about the criteria (i.e. rubrics) that the teacher will use to evaluate in that particular activity. This will help students to work towards a save direction (Fig. 2).
- *Data collection* enables the visualization and uploading of data to DojoIBL. Every piece of research contains some sort of data collection, which very often consist of collecting existing information on the internet or in their environment.
- *Concept mapping* helps students to represent and organize knowledge and concepts around a topic [32, 33]. We have developed a type of activity that stores the

[8] https://spiral.ac/student.

[9] https://www.edmodo.com/.

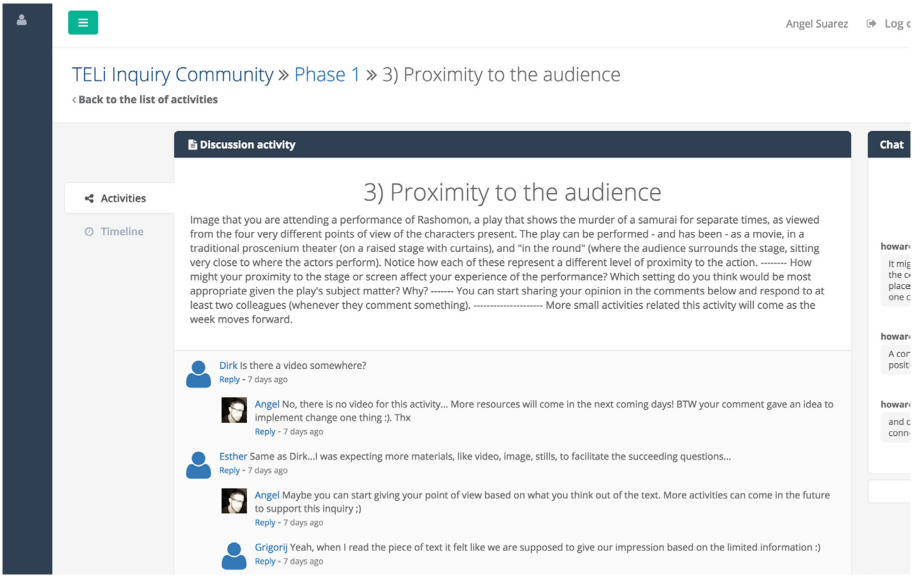



**Fig. 2.** Example of activity type: *discussion*.

information on the server, rather than relying on services like Mindmeister[10] that stores the concept map data externally.

- *External plugin* enables the integration of external widgets repositories like GoLabs [20]. Those widgets provide the possibility to conduct scientific experiments in a virtual environment.
- *Multimedia* are similar to discussion activity but it adds the possibility to incorporate a multimedia element to inspire students. The multimedia can be used to support the description of the activity.

The activities are provided with an individual section for comments or explanations. Students can, for example, share, negotiate or compare their ideas. Actually, they can experience what the study [34] defined as the five phases of negotiation and knowledge co-construction: sharing and comparing, dissonance, negotiation, co-construction, testing and application. In addition, in order not to increase *extraneous cognitive load* for students, the design is inspired on existing social network platforms. The idea is to help students to get confidence with system quickly to speed up the adaptation phase.

The last requirement in the design section was the support of collaboration. The *instant messaging system* (right side of Fig. 3) offers a communication channel that is contextualized to the inquiry topic, therefore discussions through the chat system are embedded in a context which helps to focus the discussions. The instant messaging facilitates the support of the three essential components of any educational transaction;

[10] https://www.mindmeister.com/ accessed on March 2016.

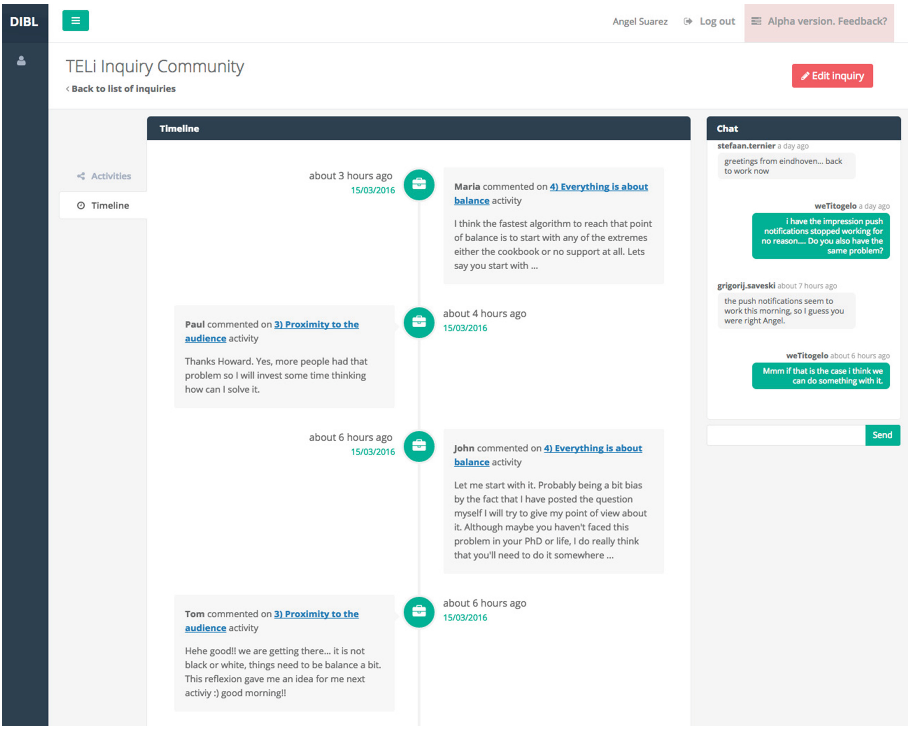



**Fig. 3.** Inquiry timeline

cognitive, social and teacher presence [11–13]. In addition, using an integrated communication channel external ways of communication are not needed anymore. This avoids the organizational burden of collecting students and teachers phone numbers or accounts to have a shared channel to communicate.

Additionally, DojoIBL implements a *notification system* and an *inquiry timeline* as is shown in Fig. 3. The timeline metaphor [35] works as a common ground where teachers and students have a high-level overview of the inquiry progress. Both the timeline and the notification system, promote collaboration awareness based on social, action and activity awareness described in [14]. Many social networks like Facebook® and Twitter® and also educational platforms like ClassDojo and Edmodo provide excellent patterns for communication that are used everyday by a large number of users. Inspired by these patterns, DojoIBL integrates several functionalities to facilitate students' collaboration and communication combined with atomic inquiry elements.

## 5 First Formative Study

DojoIBL will be used in already planned interventions in Dutch schools. In order to address any potential problems with the platform, a formative study was undertaken. The goal of this formative study was to get an understanding of how users perceived the integration of IBL functionalities with social collaborative tools.

For this experiment we had a total number of 11 experts in the field of Technology Enhanced Learning. Participants were invited to take part in the experiment voluntarily. To get an understanding of how the users perceived DojoIBL, a standardized User Experience Questionnaire (UEQ) [36] was used. The UEQ was designed to obtain a fast and immediate measurement of the user experience of interactive products [37]. It consists of 26 items that measure the perception of a user interface regarding pragmatic, hedonic and attractiveness dimension. *Attractiveness* represents the overall impression of the product, whereas pragmatic and hedonic are defined as follows.

Pragmatic dimensions include:

- *perspicuity*: How easy is to get familiar with the product?
- *efficiency:* Can users solve their tasks without unnecessary effort?
- *dependability*: Does the user feel control of the interaction?

Hedonic dimensions include:

- *stimulation:* Is exciting and motivating to use the product?
- *novelty:* Is the product innovate or creative? Does the product catch the interest of the users?

*Attractiveness* is represented by 6 items whereas pragmatic and hedonic by four items each. Next to the UEQ, the users perceived usability of DojoIBL was measured using the System Usability Scale (SUS) [38]. SUS is a reliable tool for measuring usability, which consists of 10 items with five possible answers. Both UEQ and SUS are quantitative analysis, therefore to complement the evaluation a semi- structured interview was used. This interview consists of three open questions for collecting more qualitative feedback.

**Experimental Design.** This formative study lasted for one and a half week. To inform and exhort participants to take part in the experiment, two emails were sent to them. The first one was sent a couple of days before the experiment started and it explained the goal and described the activity. The second email, sent on the same day where the activity started, provided the credentials for the participants to access DojoIBL. Participants were instructed to login DojoIBL, to join one inquiry using an inquiry code and to follow the activities created within the inquiry.

As the goal of the experiment was to know how users perceived the tool, we provided the participants a series of activities based on open ending questions to engage them with DojoIBL. During the time that the activity was running, participants talked in parallel about the topics discussed in DojoIBL. To collect feedback about the user experience (UX) participants were invited to answer questionnaires.

## 6 Results

The 11 participants generated in DojoIBL 260 messages in the chat and 92 responses for the 5 activities created for the inquiry. From those 92 responses, 31 were generated in the concept map and 61 were comments to activities (43 were initial comments and 18 replies to other's comments). The means (ranging from −3 to 3) and standard deviations (in parenthesis) of the UEQ dimensions for the 11 participants were: *attractiveness* 2.04 (0.51), *perspicuity* 1.84 (0.55), *efficiency* 1.82 (0.51), *dependability* 1.43 (0.82), *stimulation* 1.77 (0.61) and *novelty* 1.61 (0.67). According to these results, participants were equally satisfied with the judgment of hedonic and pragmatic quality dimensions and slightly more satisfied with the *attractiveness* dimension. For testing the reliability of the dimensions, Conbrach's Alpha was calculated for each dimension. *Attractiveness* 0.85, *perspicuity* 0.7, *dependability* 0.69 and *stimulation* 0.71 showed a satisfactory reliability. Comparing the results to a benchmark based on data from 163 studies, DojoIBL scored in the 10 % best results in all the scales besides *dependability*.

The overall usability of DojoIBL was rated high by the participants. The mean score for the SUS was 78.0 (12.6). The confidence interval, with confidence level on 95 %, ranged from 69.46 to 86.45. For testing reliability Conbrach's Alpha was calculated obtaining 0.81, which shows a satisfactory reliability. According to what SUS suggests, both the mean and the confidence interval are above 68 which is considered above the average.

From the semi-structured interviews, a number of issues were identified. In five cases, the participants reported problems while navigating back to the phase from the activities. Respondents stressed that going back to the phase overview was not intuitive enough. Also three participants noted problems positioning nodes in the concept maps. The suggestions for improving included a better way to qualify and label the links in the concept map, default inquiry templates while creating new inquiries following existing inquiry models and the integration of learning analytics.

The results, as shown in Fig. 4, confirmed that participants liked DojoIBL and it can be appreciated in several comments like "*I really like the social functionality*" or "*I like the timeline*" found in the chat.

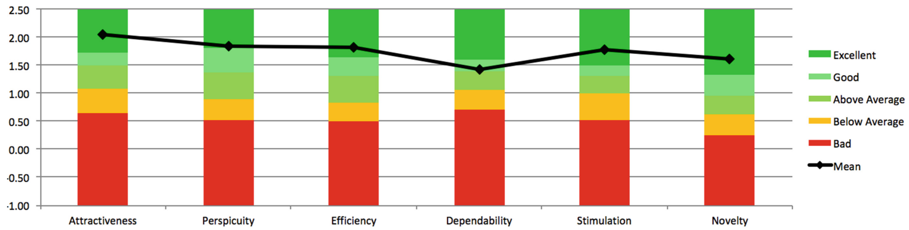



**Fig. 4.** DojoIBL scores comparison to benchmark

## 7 Discussion

DojoIBL has been developed through a process of design-based research, which promotes progressive refinement of the design [23]. Our conception of social collaborative inquiry learning and its support using DojoIBL motivated the conceptual basis for DojoIBL design, development and refinement leading to the impending interventions in the schools.

Our goal in this formative study was to gain a better understanding of the way in which the users perceived DojoIBL. In particular, how they perceived the integration of social collaborative tools into an IBL platform. The UEQ scales *efficiency*, *perspicuity* and *dependability*, which measured classical usability, showed that participants perceived DojoIBL as a suitable platform to elaborate and hold discussions around open ended questions. In addition, log data also supported this perception. Participants contributed 8 times on average to activities and they sent on average 23 messages to the chat. The 11 participants were merely instructed to read the description of the activities, having the freedom to contribute or not. Their levels of engagement in social interactions shows that DojoIBL supports social collaborative processes. These interpretations can be confirmed by the SUS questionnaire, where participants, with a high reliability, found the system easy to use and the DojoIBL functionalities very well integrated.

More interpretations can be extracted from the semi-structured interviews. In general participants described the instant messaging as very convenient an intuitive resource to communicate and to ask for specific support. Thus this showed support for two of the components of any educational transaction defined in CoI [3, 10]: social and teaching presence. Regarding cognitive presence, participants found the possibility to discuss around inquiry activities very interesting. They argued that, while instant messaging provides a quick way to communicate an idea, the affordance to also comment on activities provide students time to reflect and to elaborate their contributions. Therefore, this way of communication might be preferable to instant messaging or even oral communication when the goal is to increase high-order cognitive learning [9].

Participants also reflected about the degree of awareness supported. It seemed that social and action awareness [14] were covered with the combination of using notifications and the timeline, as the participants found them convenient to track what others were doing. However, no evidences were reported about the support of activity awareness, which informs users about organizational or structural changes.

In summary, the overall impression from the participants was positive. Besides the feedback that will be addressed and included in the next round of development, participants were excited about the potential of DojoIBL. This was explicitly manifested when some participants showed their interest about future steps of DojoIBL in terms of interventions with students and the roadmap for future updates.

## 8 Future Work and Conclusion

This manuscript presented DojoIBL, a Learning Content Management System that aims at nurturing 'Community of Inquiry' (CoI), by helping students to co-create knowledge through social interactions. It combined essential elements to support inquiry-based learning (IBL) with social collaborative tools in order to facilitate better collaborative processes. In short, DojoIBL focused on adding value to teachers and students' IBL experiences by providing a simple, intuitive and flexible tool.

This formative study informed about how the users perceived DojoIBL, particularly the integration of collaborative tools into an IBL platform. The results showed a positive acceptance from participants, perceiving DojoIBL as a suitable tool to engage in collaborative inquiry processes. In addition, the results also showed that DojoIBL copes with the three essential components to any educational transaction described in CoI: cognitive, social and teaching presence.

In future developments of DojoIBL, the integration of role support [39] to enable testing the role taking strategy in IBL processes will be addressed. Roles, as a way to foster communities of inquiry by facilitating interactions between inquirers and fostering positive interdependence [40] will be further investigated. Additionally, although DojoIBL provides a 'liquid design' to be used in any device, a mobile app version is being develop for android, iOS and windows.

To conclude, this manuscript contributed DojoIBL, an open source platform that aims at fostering communities of inquiry for driving students' success facilitating the acquisition of the so called 21st century skills, e.g. communication and collaboration.



## References

1. Bruder, R., Prescott, A.: Research evidence on the benefits of IBL. ZDM Math. Educ. **45**, 811–822 (2013)
2. Vogel, B., Spikol, D., Kurti, A., Milrad, M.: Integrating mobile, web and sensory technologies to support inquiry-based science learning. In: 6th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education, pp. 65–72. IEEE (2010)
3. Peirce, C., Buchler, J.: Philosophical Writings of Peirce. Dover Publications, New York (1955). Selected and Edited, with and Introduction, by Justus Buchler
4. Mikroyannidis, A., Okada, A., Scott, P., Rusman, E., Specht, M., Stefanov, K., Boytchev, P.: weSPOT: a personal and social approach to inquiry-based learning. J. Univ. Comput. Sci. **19**(14), 2093–2111 (2013)
5. Bell, T., Urhahne, D.: Collaborative inquiry learning: models, tools, and challenges. Int. J. Sci. Educ. **32**(3), 349–377 (2010)
6. Dillenbourg, P.: What do you mean by collaborative learning. Collab. Learn. Cogn. Comput. Approaches **1**, 1–15 (1999)
7. Scardamalia, M., Bereiter, C.: Higher levels of agency for children in knowledge building: a challenge for the design of new knowledge media. J. Learn. Sci. **1**, 37–68 (1991)
8. Piaget, J.: The Language and Thought of the Child. Harcourt Brace & Company, New York (1926)

9. Garrison, D.R., Anderson, T., Archer, W.: Critical inquiry in a text-based environment: computer conferencing in higher education. Internet High. Educ. **2**, 87–105 (1999)
10. Pardales, M.J., Girod, M.: Community of inquiry: its past and present future. Educ. Philos. Theor. **38**, 299–309 (2006)
11. Garrison, D., Anderson, T., Archer, W.: Critical thinking, cognitive presence, and computer conferencing in distance education. Am. J. Distance Educ. **15**, 7–23 (2001)
12. Rourke, L., Anderson, T.: Assessing social presence in asynchronous text-based computer conferencing. Int. J. Distance Educ. **14**(3), 51–70 (2007)
13. Anderson, T., Rourke, L., Garrison, D., Archer, W.: Assessing teaching presence in a computer conferencing context. J. Asynchronous Learn. Netw. **5**(2), 1–17 (2001)
14. Carroll, J., Rosson, M., Convertino, G., Ganoe, C.: Awareness and teamwork in computer-supported collaborations. Interact. Comput. **18**, 21–46 (2006)
15. Carroll, J., Neale, D., Isenhour, P., Rosson, M., McCrickard, D.: Notification and awareness: synchronizing task-oriented collaborative activity. Int. J. Hum. Comput. Stud. **58**, 605–632 (2003)
16. Edelson, D.C., Gordin, D.N., Pea, R.D.: Addressing the challenges of inquiry-based learning through technology and curriculum design. J. Learn. Sci. **8**, 391–450 (1999)
17. Lehtinen, E.: Computer supported collaborative learning: a review. Reports on Education (1999)
18. van Joolingen, W.R., de Jong, T., Lazonder, A.W., Savelsbergh, E.R., Manlove, S.: Co-Lab: research and development of an online learning environment for collaborative scientific discovery learning. Comput. Hum. Behav. **21**, 671–688 (2005)
19. Mulholland, P., Anastopoulou, S., Collins, T., Feisst, M., Gaved, M., Kerawalla, L., Paxton, M., Scanlon, E., Sharples, M., Wright, M.: nQuire: technological support for personal inquiry learning. IEEE Trans. Learn. Technol. **5**, 157–169 (2012)
20. Gillet, D., de Jong, T., Sotirou, S., Salzmann, C.: Personalized learning spaces and federated online labs for STEM education at school. In: 2013 IEEE Global Engineering Education Conference (EDUCON), pp. 769–773. IEEE (2013)
21. Bodemer, D., Dehler, J.: Group awareness in CSCL environments. Comput. Hum. Behav. **27**(3), 1043–1045 (2011)
22. De Laat, M., Prinsen, F.: Social learning analytics: navigating the changing settings of higher education. Res. Pract. Assess. **9**, 51–60 (2014)
23. Barab, S., Squire, K.: Design-based research: putting a stake in the ground. J. Learn. Sci. **13**, 1–14 (2004)
24. Specht, M., Bedek, M., Duval, E., Held, P., Okada, A., Stevanov, K., Parodi, E., Kikis-Papadakis, K., Strahovnik, V.: WESPOT: Inquiry based learning meets learning analytics. In: 3rd International Conference on e-Learning, pp.15–20. Belgrade, Serbia (2012)
25. Pedaste, M., Mäeots, M., Siiman, L.A., de Jong, T., van Riesen, S.A.N., Kamp, E.T., Manoli, C.C., Zacharia, Z.C., Tsourlidaki, E.: Phases of inquiry-based learning: definitions and the inquiry cycle. Educ. Res. Rev. **14**, 47–61 (2015)
26. Donohoo, J.: Collaborative Inquiry for Educators: A Facilitator's Guide to School Improvement. HarperCollins, New York (2013)
27. Dillenbourg, P., Järvelä, S., Fischer, F.: The evolution of research on computer-supported collaborative learning. In: Balacheff, N., Ludvigsen, S., de Jong, T., Lazonder, A., Barnes, S. (eds.) Technology-Enhanced Learning, pp. 3–19. Springer, Netherlands (2009)
28. Hmelo-silver, C.E., Duncan, R.G., Chinn, C.A.: Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). Educ. Psychol. **42**, 99–107 (2007)

29. Ternier, S., Klemke, R., Kalz, M., Specht, M.: ARLearn: augmented reality meets augmented virtuality. J. Univ. Comput. Sci. Technol. Learn Across Phys. Virtual Spaces **18** (15), 2143–2164 (2012)
30. Tafoya, E., Sunal, D.W., Knecht, P.: Assessing inquiry potential: a tool for curriculum decision makers. Sch. Sci. Math. **80**, 43–48 (1980)
31. Ahern-Rindell, A.: Applying inquiry-based and cooperative group learning strategies to promote critical thinking. J. Coll. Sci. Teach. **28**(3), 203–207 (1998)
32. Stoddart, T., Abrams, R.: Concept maps as assessment in science inquiry learning - a report of methodology. J. Sci. Educ. **22**(12), 1221–1246 (2000)
33. Akinsanya, C., Williams, M.: Concept mapping for meaningful learning. Nurse Educ. Today **24**(1), 41–46 (2004)
34. Garrison, D.: Critical thinking and self-directed learning in adult education: an analysis of responsibility and control issues. Adult Educ. Q. **36**, 60–64 (1992)
35. Vavoula, G., Sharples, M.: KLeOS: a personal, mobile, knowledge and learning organisation system. In: IEEE International Workshop on Wireless and Mobile Technologies in Education (2002)
36. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 63–76. Springer, Heidelberg (2008)
37. Santoso, H., Barat, I.J., Schrepp, M.: Measuring user experience of the student-centered e-Learning environment. thejeo.com
38. Sauro, J.: Measuring usability with the system usability scale (SUS) (2011)
39. Strijbos, J.-W., De Laat, M.F.: Developing the role concept for computer-supported collaborative learning: an explorative synthesis. Comput. Hum. Behav. **26**, 495–505 (2010)
40. Johnson, D., Johnson, R., Smith, K.: Active Learning: Cooperation in the College Classroom. Interaction Book Company, Edina (1991)

# Inferring Student Attention with ASQ

Vasileios Triglianos[1(✉)], Cesare Pautasso[1], Alessandro Bozzon[2], and Claudia Hauff[2]

[1] Faculty of Informatics, University of Lugano, Lugano, Switzerland
{vasileios.triglianos,cesare.pautasso}@usi.ch
[2] Web Information Systems, Delft University of Technology, Delft, The Netherlands
{a.bozzon,c.hauff}@tudelft.nl

**Abstract.** ASQ is a Web application for broadcasting and tracking interactive presentations, which can be used to support active learning pedagogies during lectures, labs and exercise sessions. Students connect their smartphones, tablets or laptops to receive the current slide as it is being explained by the teacher. Slides can include interactive teaching elements (usually questions of different forms). In contrast to other existing platforms, ASQ does not only collect, aggregate and visualize the answers in real-time, it also supports the *data analytics in the classroom* paradigm by providing the teacher with a real-time analysis of student behaviour during the entire session. One vital aspect of student behaviour is *(in)attention* and in this paper we discuss how we infer — in real-time — student attention based on log traces ASQ collects.

## 1 Introduction

In the traditional post-secondary classroom-based learning, forty-five or ninety minute units of teaching are the norm. Students' attention during such teaching sessions varies significantly, as shown in a wide range of empirical studies that have either probed students directly for self-reports of attention (or day-dreaming and mind wandering) levels [2,5,8,10] or aimed to *infer* [in]attention based on (i) students' behaviour (e.g. their patterns of note-taking [9] or physical signs of inattention such as gazing [4]), (ii) physiological measures such as skin temperature [1], or, (iii) students' levels of knowledge retention [8,14].

Many of these techniques can only be employed at reasonable cost for a small subset of classes and/or a small subset of students due to their obtrusive nature (examples include physiological markers or minute-by-minute self-reports), issues of scale (e.g., the presence of external observers and the analyses of taken notes), and, the additional cognitive & timely burden placed on students (e.g., through retention tests). Moreover, with few exceptions, e.g. [12], these techniques do not enable lecturers to adapt their teaching on-the-fly, as they are not able to *continuously* determine students' attention *in real-time*; instead students are probed at specific intervals during the lecture or post-lecture data collection and data analyses steps are required.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 306–320, 2016.
DOI: 10.1007/978-3-319-45153-4_23

In this work, we investigate to what extent modern Web technologies can facilitate and enable the *continuous*, *scalable* and *unobtrusive* inference of student attention *in real-time*. We target the traditional classroom setting – so as to enable lecturers to *react* in a timely manner to the attention needs of their students – and we focus on Web-mediated teaching and formative assessment activities. We seek answer to the following Research Questions:

RQ1. *To what extent can students' attention be inferred from their interactions with a Web-based platform?*

RQ2. *Which type of interactions are most correlated with (in)attention?*

As common in previous works, we infer attention from students' retention levels. To this end, we have extended `ASQ` [16], a Web platform aimed at providing active classroom-based learning pedagogics such as enquiry based learning, problem based learning and collaborative learning. `ASQ` provides extensive logging capabilities, thus enabling the tracking and recording of real-time students' interactions during lectures. We deployed `ASQ` in the context of three ninety minute university-level lectures given by two different instructors, with varying interactivity levels and up to 187 students. Our results show that `ASQ` can provide fine-grained insights on students' attention states that relate to previous findings on the subject, thus demonstrating `ASQ`'s ability to obtain an accurate view of students' attention in a classroom setting.

## 2 Related Work

Measuring and influencing peoples' state of attention in their workplaces, daily lives and educational settings has been investigated for a number of decades in psychology and pedagogy; in more recent years technological advances have also led to contributions by the human computer interaction and the learning analytics communities [6,7].

Our research focus is in the measuring of students' attention in the post-secondary classroom, and thus in this section we narrow our overview to works that have investigated attention in the educational context only. Two important meta-studies [15,17], published in 2007 and 2013 respectively, not only summarize the current state of knowledge about student attentiveness, but also critically highlight the often contradictory findings — in [17] specifically, the assertion of the 10–15 min attention span of students is tackled in great detail. The contradictions are generally attributed to the nature of the individual experiments, which are typically conducted on a small number of students taking a class of less than one hour, which may have been specifically designed for the experiment. Factors which can explain the observed differences include the inherent variability of students' academic interests, instructor styles and means of measuring attention, which are usually not controlled for across experiments [15]. Of the many findings, we list here those which have been observed in several experiments[1]. *F1*: Students' attention drops over the class period [5]; as a consequence,

[1] Once more, it should be noted that for these findings some contradictory evidence exists as well.

in retention tests students tend to perform better on material presented early on in the class [8]. *F2*: attention breaks occur regularly and increase in frequency as the class progresses [4]. *F3*: As the class progresses, students tend to take less notes [9]. *F4*: the percentage of students attentive to the class varies significantly (depending on class topic, the instructor and the pedagogical tool employed). Between 40 % and 70 % of students are attentive at any moment during frontal teaching. Attention rises when interactive elements are introduced (discussions and problem solving) [2]. *F5*: immediately after interactive teaching elements, the level of distraction is lower than before the start of the interaction [2,3].

One common denominator of the aforementioned studies is their lack of technologies to determine students' attention directly or indirectly. Existing technology-based solutions, while enabling real-time insights, are also limited, due to the invasive technologies employed. In [12,13] EEG signals are recorded to infer students' attention — while accurate, those studies are restricted to either small classroom or lab settings. Sun et al. [11] rely, among others, on facial expressions to detect attention, which, while technologically feasible raises privacy concerns. Bixler et al. [1] find eye gaze and skin conductance and temperature recordings to be indicative of attention.

In contrast, in our work we explore the use of a *non-invasive* and *scalable* technological solution.

## 3 ASQ: From Low-Level Events to Attention States

ASQ is a Web-based tool for delivering interactive lectures. It builds upon the modern Web technology stack and allows teachers to broadcast HTML slides to students' devices and on-the-fly to receive and process their reactions and responses. The slides may contain exercises with interactive questions such as "choose one out of five", "highlight the text", "classify elements", "program a JavaScript function", or "write a SQL query" (Fig. 2) — these question types can be extended for different needs and new question types can easily be added to ASQ due to its modular nature. The answers students submit are available to the instructor for review and discussion *in real-time*. Moreover, most question types support the automatic aggregation and clustering of the answers, thus reducing the cognitive load of the instructor which in turn enables a quicker (and more accurate) feedback cycle. To reiterate, the main design driver of ASQ was to enable teachers to gather feedback live in the classroom and immediately assess the level of understanding of the entire classroom, by turning student devices from potential distractions into a novel communication channel — Fig. 1 shows an example session of ASQ in the classroom.

**Low-Level Event Capturing.** In order to capture the interactions with the taught material, and to understand how they contribute to the learning process and student attention, ASQ tracks various events (e.g. a user connects to the ASQ presentation, submits an answer or is idle for a number of seconds) generated by each learner's browser during a live presentation session. Note, that we do not

**Fig. 1.** `ASQ` in the classroom: most students' laptops are connected and focused on the slide material being explained.


**Fig. 2.** SQLite question from Lecture 1, *Advanced SQL*. It comprises a text editor (left) to write and execute SQL queries on an in-browser database instance, and a results pane (right) to visualize the query results. *(Best viewed in the electronic version.)*

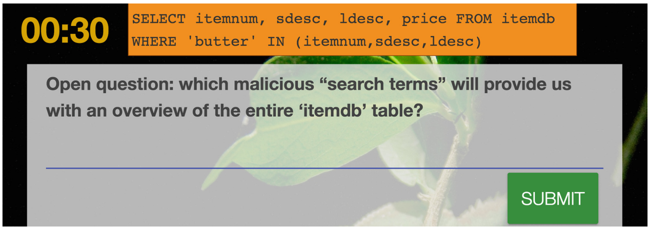



**Fig. 3.** A text input question from Lecture 3, *Web Security*.

require students to login to `ASQ`, as long as a student's browser is connected to the `ASQ` presentation relevant events will be captured; closing the browser tab that contains the `ASQ` presentation will disconnect the student. Specifically, in this work we consider the browser events listed in Table 1; events are generated not only when students interact with an `ASQ` question type, but also when they interact more generally with the browser window containing the `ASQ` tab.

Recall that our overarching goal is to infer student attention. To this end, based on the introduced low-level events, we define higher-level activity indicators, which denote the activity (or lack thereof) currently performed by a student in a lecture. Subsequently, we use these indicators to infer a basic model of *student attention states*.

**Activity Indicators.** Each low-level browser event occurs at a specific point in time; we map sequences of browser events generated by a student to one of six binary activity indicators, which we consider to be natural components of a student's attention state. These indicators are non-exclusive (i.e. several indicators can be true at the same time) and listed in Table 2: `exercise`, `connected`, `focus`, `idle`, `input` and `submitted`.

**Table 1.** Overview of web browser events monitored by the ASQ application

| Event name | Description |
|---|---|
| tabhidden | The browser tab that displays the ASQ web app becomes hidden. |
| tabvisible | The browser tab that displays the ASQ web app becomes visible. |
| windowfocus | The browser window that displays the ASQ web app receives focus. |
| windowblur | The browser window that displays the ASQ web app loses focus (blurs in HTML terminology). |
| exercisefocus | An ASQ exercise HTML element receives focus. |
| exerciseblur | An ASQ exercise HTML element blurs. |
| input | There is student input in the browser window that displays ASQ. |
| questioninput | Some ASQ question types emit this event when there is student input. |
| exercisesubmit | A student submits the solution to an ASQ exercise. |
| answersubmit | A student submits an answer for an ASQ question (an exercise can have multiple questions). |
| idle | Emitted by the browser window that displays the ASQ web app when none of the above events has occurred for 10 seconds. |
| connected | A student connects to the ASQ server. |
| disconnected | A student disconnects from the ASQ server. |

**Table 2.** Overview of activity indicators based on browser events

| Name | Description |
|---|---|
| exercise | True when the current slide has an exercise. |
| connected | True when the student browser is connected. |
| focus | True when the browser has focus on the tab or exercise related to the lecture. |
| idle | True from the time of an idle event until one of tabhidden, tabvisible, windowfocus, windowblur, focusin, focusout, exercisefocus, exerciseblur, input, questioninput, exercisesubmit and answersubmit occurs. |
| input | True when an input or questioninput event occurs. This state is valid only on slides that contain exercises. |
| submitted | True when the student has submitted at least once this exercise (as indicated by an exercisesubmit event). This state is valid only on slides that contain exercises. |

**Student Attention States.** We take a data-driven approach to the exploration of the activity indicators and in Table 3 list all the 17 combinations of indicators that we observed in our data traces (described in detail in Sect. 4). We manually assign ten different semantic labels to each combination. For instance, a student who has `submitted` an answer to an `exercise` and is now `idle` with `ASQ` in `focus` is considered to be *Waiting* (e.g. for the instructor to provide feedback), while a student who also `submitted` an answer and is neither `idle` nor having `ASQ` in `focus` is considered to be *Bored* (and occupying himself with other activities on the device). Thus, at each point in time a student is in exactly one of the ten listed attention states. Figure 4 showcases the progression of two students' attention states across an entire lecture; while *Student 1* starts off the lecture at a high level of attention (indicating by the continuous *Following* state) and later on toggles between the *Following* and *Distracted* states, *Student 2* starts off the lecture in a *Distracted* state and only exhibits short bursts of attention shortly before or after some of the interactive exercises.

Although we are using psychological terms such as *Bored*, *Distracted*, *Thinking*, and the like, these should be not be interpreted beyond the strict definition of Table 3 as our goal is to give a readable representation of the aggregated activity indicators that can be amenable of further analysis and experimentation. In the remainder of this paper we analyze to what extent our definition of inferred attention states is suitable to reproduce findings from the literature.

**Table 3.** Modeling student attention based on activity indicators. Activity indicators are binary, ✓ represents `True`, and - represents `False`.

| exercise | connected | focus | idle | input | submitted | Inferred attention state |
|---|---|---|---|---|---|---|
| - | - | - | - | - | - | Disconnected |
| ✓ | - | - | - | - | - | Disconnected |
| ✓ | - | - | - | - | ✓ | Disconnected |
| - | ✓ | - | - | - | - | Distracted |
| - | ✓ | - | ✓ | - | - | Distracted |
| ✓ | ✓ | - | - | - | - | Searching for a solution |
| ✓ | ✓ | - | ✓ | - | - | Searching for a solution |
| - | ✓ | ✓ | - | - | - | Interacting with non-question slide |
| - | ✓ | ✓ | ✓ | - | - | Following |
| ✓ | ✓ | ✓ | - | - | - | Thinking |
| ✓ | ✓ | ✓ | ✓ | - | - | Thinking |
| ✓ | ✓ | - | - | - | ✓ | Bored |
| ✓ | ✓ | - | ✓ | - | ✓ | Bored |
| ✓ | ✓ | ✓ | - | - | ✓ | Waiting |
| ✓ | ✓ | ✓ | ✓ | - | ✓ | Waiting |
| ✓ | ✓ | ✓ | - | ✓ | - | Working on an answer |
| ✓ | ✓ | ✓ | - | ✓ | ✓ | Reworking answer |

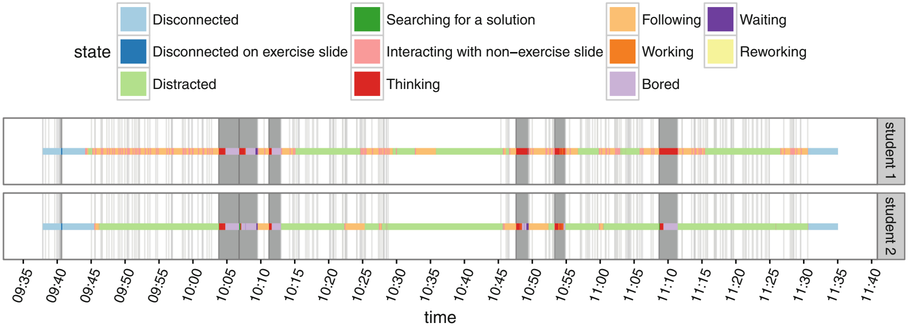



**Fig. 4.** Two example progressions of inferred attention states during the course of a single 90-min lecture (specifically: Web Security). The dark-grey areas represent slides with interactive exercises (6 in total), while the light-grey vertical bars indicate slide transitions. While *Student 1* starts off highly attentive, *Student 2* is inattentive from the start of the lecture.

## 4 ASQ Deployment and Data Collection

We deployed ASQ in the 2015/16 edition of *Web and Database Technology*, a compulsory course for $1^{st}$ year BSc Computer Science and an elective for $3^{rd}$ year BSc minor students, at the Delft University of Technology. The course was followed by 310 students in total: 260 $1^{st}$ year and 50 minor students. Across the eight course weeks, fifteen 90-minute lectures were given. We utilized ASQ in three of those sessions, identified as suitable for experimentation: at regular intervals, the lecture material was interspersed with interactive elements consisting of live programming exercises, multiple choice questions, and visual question types.

At the beginning of each ASQ session, students in the lecture hall were instructed (though not compelled) to open the lecture presentation in the browser. Students connected anonymously; a random identifier was assigned to each connection, enabling us to group all interactions made by the same student within one lecture together (identifying markers *across* lectures were not retained for privacy reasons). The lecture slides were not only visible in the students' browser but also on the lecture hall screen and thus students who decided not to use ASQ treated the sessions as standard lectures.

We posed questions of three question types that depended on the lecture material and assessment goals of each class: (A) multiple-choice, (B) SQLite programming (Fig. 2), and (C) text-input (Fig. 3). Table 4 summarizes the main characteristics of the three lectures, including the lectures' topic, the number of students participating through ASQ and the number of questions posed per type. Note that *Lecture 1. Advanced SQL Topics* has generated almost seven times more browser events than the other two lectures due to its usage of SQLite programming quizzes: not only the large amount of typing contributed to the events generation, but also the question setup which required the students to consult

**Table 4.** Overview of the three ASQ lecture sessions each given by one of two instructors (identified as I1 and I2). For each session, the number of students participating, the number of exercises (per type) and the number of ASQ low-level browser events logged are listed.

| | Instr. | Topic | #Students using ASQ | #ASQ events | #Question types | | |
|---|---|---|---|---|---|---|---|
| | | | | | A | B | C |
| 1 | I1 | *Advanced SQL Topics* | 143 | 121,062 | 0 | 7 | 0 |
| 2 | I1 | *ER Conceptual Design* | 111 | 17,460 | 8 | 0 | 0 |
| 3 | I2 | *Web Security* | 187 | 17,562 | 4 | 0 | 2 |

a database schema diagram resulting in a considerable amount of blur/focus events between ASQ and the diagram.

## 5 Analysis

In our exploration of the collected logs, we are guided by our research questions and the five main findings of prior works (identified in Sect. 2) exploring students' attentiveness in the classroom.

*F1: Students' attention drops over the class period.* For all lecture logs, we translated low-level browser events into activity indicators (Fig. 5) and subsequently inferred attention states (Fig. 6). We consider the two activity indicators `connected` and `focused` and the union of the states *Following/Thinking/Working* as well as *Distracted/Bored* as most suitable representatives of student attention and inattention respectively. To explore how attention changes over time, we correlate the lecture time (in units of 1 second) with the number of students in the specific state(s) or activity setting. If, as expected student attention drops over time, we will observe a decrease in `focus` over time and an increase in *Distracted/Bored* students. The results in Table 5 show that this is indeed the case: inattention-oriented activities/states are positively correlated with time while attention-oriented activities/states are negatively correlated with time. Moreover, the high-level inferred attention states achieve higher absolute correlations, indicating that they are more suitable to infer (in)attention than our low-level activity indicators.

We thus posit that based on the events logged in ASQ, we are able to infer in real-time (and live in the classroom) when and to what extent attention drops over time, relying on either the `focus` activity indicator as a basic measure or a combination of the more high-level attention states *Following/Thinking/Working* (and their counterparts).

*F2: Attention breaks occur regularly and increase in frequency as the class progresses.* For each second of the lecture we track the number of *attention breaks*, that is, the number of students that switch their device from `focused` on ASQ to some other activity. We also track *attention recovery* which we define as the

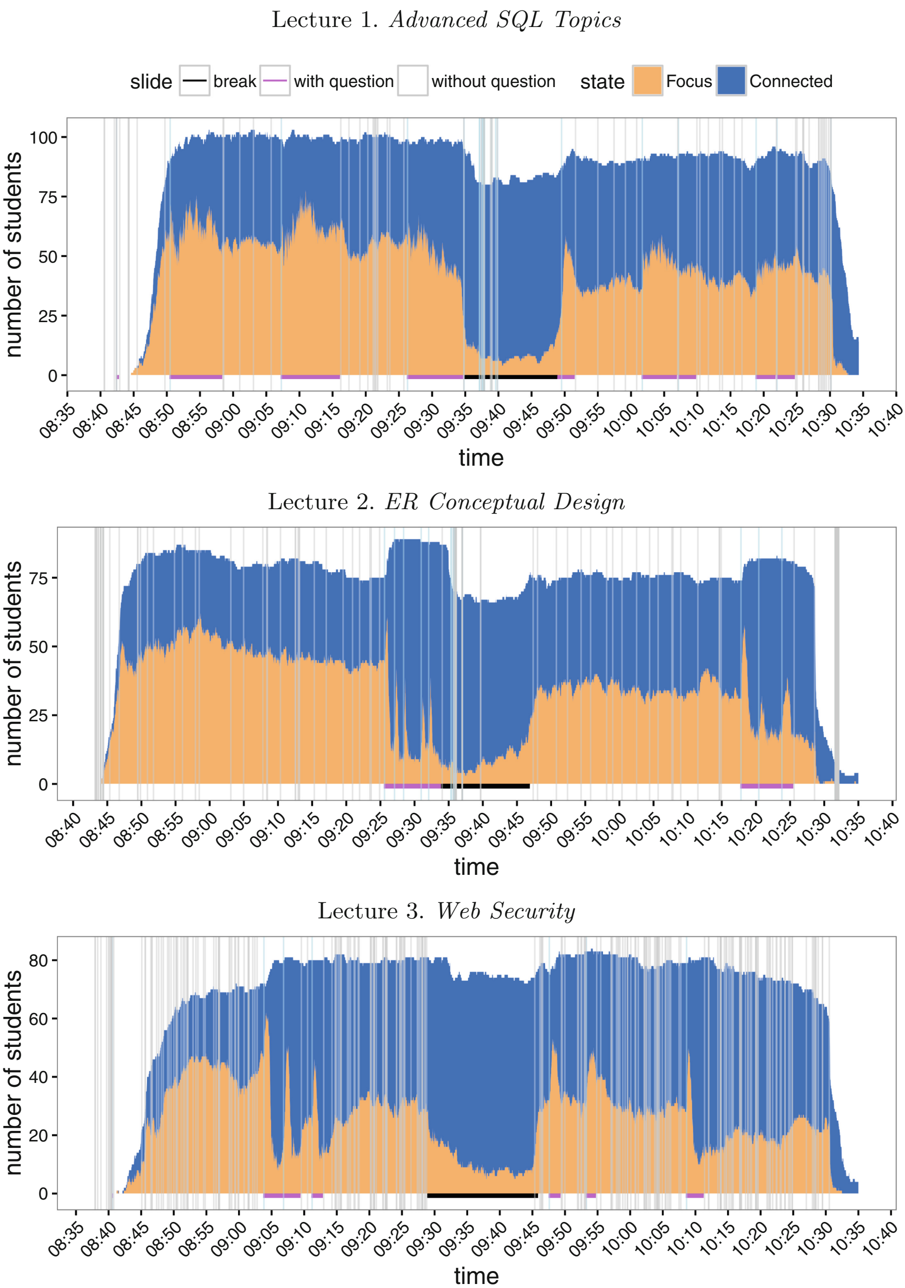



**Fig. 5.** Connected and focused activity indicators for all the sessions

**Table 5.** Linear correlation coefficient (significant correlations at the $p < 0.05$ level are marked †) between time and number of students exhibiting a particular activity indicator (top part) or one of a set of inferred attention states (bottom part).

| +++ Activity indicators +++ | | | | | |
|---|---|---|---|---|---|
| | Lecture | All slides | | Slides w/o exercises | |
| | | Connected | Focus | Connected | Focus |
| 1 | Advanced SQL Topics | 0.176† | −0.182† | 0.281† | −0.059 |
| 2 | ER Conceptual Design | −0.224† | −0.569† | −0.284† | −0.637† |
| 3 | Web Security | 0.263† | −0.177† | 0.284† | −0.228† |
| +++ Attention states +++ | | | | | |
| | Lecture | All slides | | Slides w/o exercises | |
| | | Distracted/Bored | Following/Thinking/Working | Distracted | Following/Thinking |
| 1 | Advanced SQL Topics | 0.324† | −0.274† | 0.450† | −0.257† |
| 2 | ER Conceptual Design | 0.039 | −0.549† | 0.230† | −0.657† |
| 3 | Web Security | 0.391† | −0.262† | 0.458† | −0.390† |

number of students whose device switches back to `focus` on `ASQ`. The *attention focus variation* is the net sum of attention recoveries minus the attention breaks observed during the same period (a window of 30 s). For each of the three lectures we present their *attention focus variations* in Fig. 7. We observe that attention breaks occur regularly but there is no noticeable increase in frequency as the class progresses. We note that although this is in contrast to $F2$, not all empirical studies in the past observed this increase in attention breaks [15].

*F3: Attention rises when interactive elements are introduced.* Drawing on our analysis of attention focus variation, we observe that whenever there are interactive elements in the slide, in the form of questions, we observe spikes of attention recovery (Fig. 7) and an increase of connected students (Fig. 5). While introducing interactive elements thus captures the attention of the students (positive attention focus variation), shortly thereafter we observe the subsequent loss of focus due to students waiting on each other to answer. Likewise, students might be searching for solutions using their devices, something `ASQ` cannot distinguish from students simply leaving the application to do something else. As we can observe in the charts of Fig. 6 for all the lectures, whenever there is a slide with a question, the number of students that have their `ASQ` page out of focus (`Searching` state) is always lower than in slides without a question (`Distracted` state). Similarly, the magnitude of attention focus variation is smaller for slides without questions than for slides with questions, which literally appear to send jolts through the collective attention span of the students in the classroom (Fig. 7). Our results thus confirm previous findings of rising attention at interactive elements.

*F4: Immediately after interactive teaching elements, the level of distraction is lower than before the start of the interaction.* While there is a peak of interest as soon as questions are asked, after students submit their answers, their focus switches to other activities. Hence, as shown in Fig. 6 towards the end of the question, the number of students we infer to be in a *Distracted* state rises

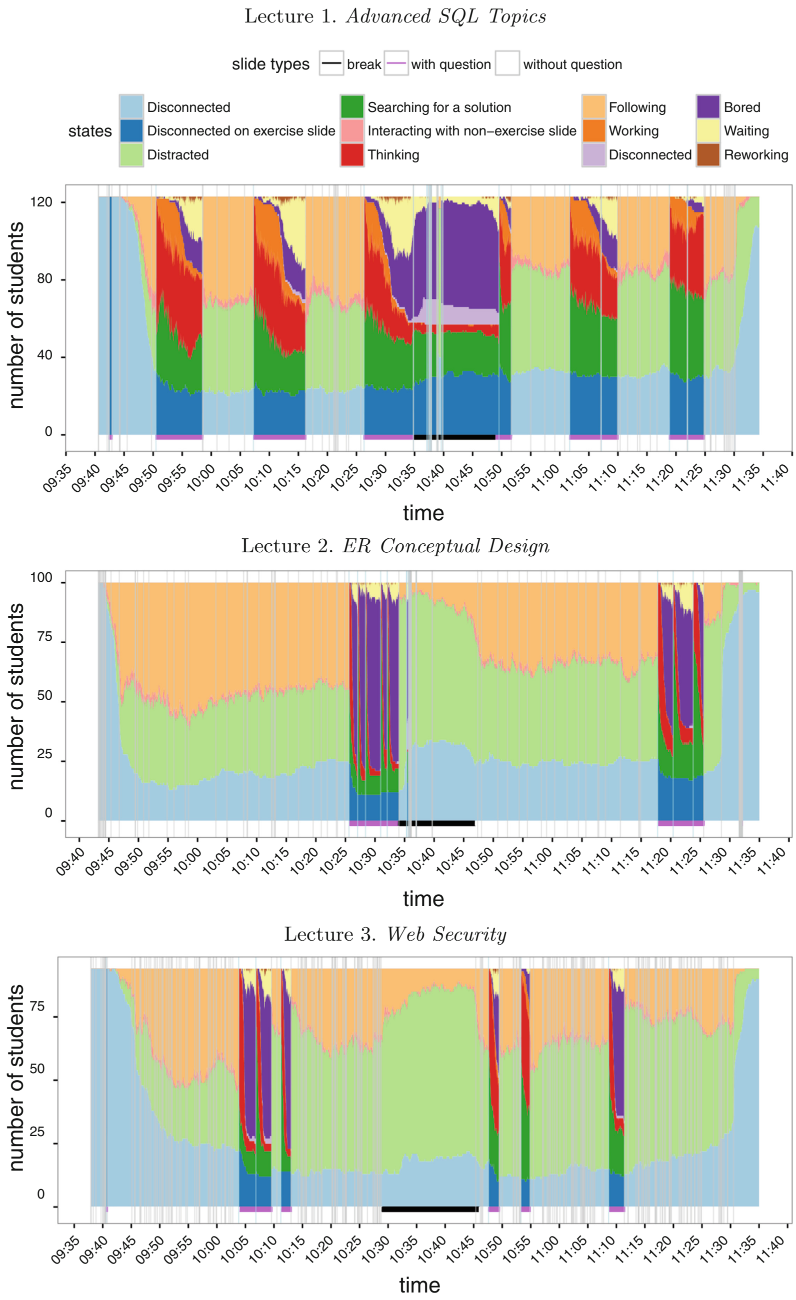



**Fig. 6.** Inferred student attention state for all the sessions

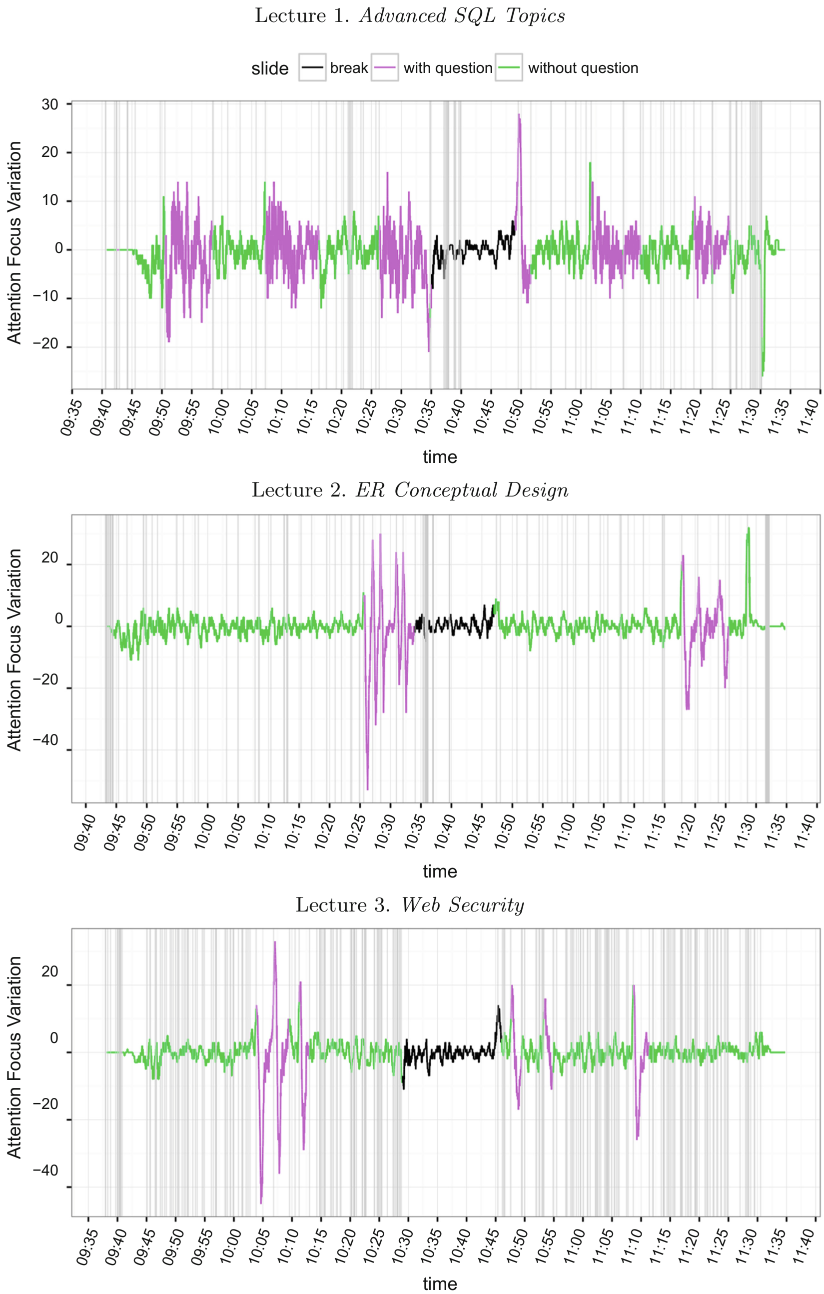



**Fig. 7.** Attention Focus Variation: how many students have changed the focus of attention during lecture (moving sum of attention breaks and recoveries using a window of 30 s).

considerably and is almost always higher than right before the interactive teaching element. The effect depends on the length of time students have to wait for other students to complete the exercise (before the instructor moves on in the lecture) and on the type of feedback given either individually or globally on the submitted answer. This result is a clear deviation from prior works and suggests that our attention model, in particular the *Distracted* state captures more than just students' distraction.

*F5: In retention tests, students tend to perform better on material presented early on in the class.* Instead of dedicated retention tests, we rely on the multiple choice (MC) questions as a retention proxy (we restrict ourselves to MC questions as the open question types require manual grading to achieve highly accurate results).

Table 6 lists the accuracy of the student answers for the eight MC questions of *Lecture 2. ER Conceptual Design* as well as the specific time they were posed in the lecture. Note that shortly after 10:30 am the official 15 min break commenced. We observe that students tend to perform better in the first half of the class than the second. Although a subset of questions from a single lecture do not provide enough evidence to support or reject this finding in the context of `ASQ` it shows once more `ASQ`'s capabilities to provide fine-grained real-time logging and analyses to the instructor.

**Table 6.** Correct vs incorrect answers ordered by time of question for Lecture 2

| Question | Start time | Correct% | Correct | Incorrect | Total |
|---|---|---|---|---|---|
| 1 | 10:25:36 | 92.31 | 72 | 6 | 78 |
| 2 | 10:27:04 | 77.22 | 61 | 18 | 79 |
| 3 | 10:28:24 | 82.50 | 66 | 14 | 80 |
| 4 | 10:30:58 | 1.43 | 1 | 69 | 70 |
| 5 | 10:32:04 | 90.79 | 69 | 7 | 76 |
| 6 | 11:17:48 | 70.77 | 46 | 19 | 65 |
| 7 | 11:20:25 | 75.41 | 46 | 15 | 61 |
| 8 | 11:23:48 | 65.00 | 39 | 21 | 60 |

## 6 Conclusions and Future Work

`ASQ` is an interactive Web-based teaching platform that allows capturing browser events to observe and categorise the behavior of its users. `ASQ` is able to provide *real-time data analytics in the classroom*, thus providing a lecturer with the capability to observe her students in a data-driven manner. In this paper we have shown how `ASQ` can be employed to infer student attention, based on either activity indicators and states we aggregate based on low-level browser events. The visualizations presented here enable instructors to observe at a very fine-grained level the behavior of an entire class with hundreds (or potentially thousands) of students. Our analysis confirms existing research findings, whereby: (1) student attention drops during the class period; (2) attention breaks occur regularly as the class progresses; and (3) attention rises when interactive elements are introduced. Additionally, we could observe a drop in attention as soon as the interactive activity is completed by individual students, which should be taken into account when planning to introduce questions and interactive exercises within a lecture.

`ASQ` can also be used to support adaptive teaching. As future work, we will further exploit `ASQ`'s attention level monitoring capabilities to recommend teachers subject-related questions that could used to restore focus, if an attention drop is detected during the presentation of slides. The current version of `ASQ` is only a first step to a highly sensor- and data-driven classroom. In our future work we plan to complement `ASQ`'s data collection and aggregation abilities with additional sensors and technologies (eye-tracking and activity sensors) in order to acquire a more complete picture of the students in the classroom.

**Acknowledgements.** This research was partially supported by the Extension School of the Delft University of Technology.

## References

1. Bixler, R., Blanchard, N., Garrison, L., D'Mello, S.: Automatic detection of mind wandering during reading using gaze and physiology. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI 2015, pp. 299–306 (2015)
2. Bunce, D.M., Flens, E.A., Neiles, K.Y.: How long can students pay attention in class? a study of student attention decline using clickers. J. Chem. Educ. **87**(12), 1438–1443 (2010)
3. Burke, L.A., Ray, R.: Re-setting the concentration levels of students in higher education: an exploratory study. Teach. High. Educ. **13**(5), 571–582 (2008)
4. Johnstone, A.H., Percival, F.: Attention breaks in lectures. Educ. Chem. **13**(2), 49–50 (1976)
5. Lindquist, S.I., McLean, J.P.: Daydreaming and its correlates in an educational environment. Learn. Individ. Differ. **21**(2), 158–167 (2011)
6. Persico, D., Pozzi, F.: Informing learning design with learning analytics to improve teacher inquiry. Br. J. Educ. Technol. **46**(2), 230–248 (2015)
7. Santos, L.P.P., et al.: Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors. In: International Learning Analytics and Knowledge, No. EPFL-CONF-216918 (2016)
8. Risko, E.F., Anderson, N., Sarwal, A., Engelhardt, M., Kingstone, A.: Everyday attention: variation in mind wandering and memory in a lecture. Appl. Cogn. Psychol. **26**(2), 234–242 (2012)
9. Scerbo, M.W., Warm, J.S., Dember, W.N., Grasha, A.F.: The role of time and cuing in a college lecture. Contemp. Educ. Psychol. **17**(4), 312–328 (1992)

10. Stuart, J., Rutherford, R.: Medical student concentration during lectures. The Lancet **312**(8088), 514–516 (1978)
11. Sun, H.J., Huang, M.X., Ngai, G., Chan, S.C.F.: Nonintrusive multimodal attention detection. In: The Seventh International Conference on Advances in Computer-Human Interactions, ACHI 2014, pp. 192–199. Citeseer (2014)
12. Szafir, D., Mutlu, B.: Pay attention!: designing adaptive agents that monitor and improve user engagement. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2012, pp. 11–20 (2012)
13. Szafir, D., Mutlu, B.: Artful: adaptive review technology for flipped learning. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2013)
14. Szpunar, K.K., Khan, N.Y., Schacter, D.L.: Interpolated memory tests reduce mind wandering and improve learning of online lectures. Proc. Nat. Acad. Sci. **110**(16), 6313–6317 (2013)
15. Szpunar, K.K., Moulton, S.T., Schacter, D.L.: Mind wandering and education: from the classroom to online learning. Front. Psychol. **4**, 495 (2013)
16. Triglianos, V., Pautasso, C.: ASQ: Interactive web presentations for hybrid MOOCs. In: Proceedings of WWW, pp. 209–210 (2013)
17. Wilson, K., Korn, J.H.: Attention during lectures: beyond ten minutes. Teach. Psychol. **34**(2), 85–89 (2007)

# Chronicle of a Scenario Graph: From Expected to Observed Learning Path

Mathieu Vermeulen[1(✉)], Nadine Mandran[2], and Jean-Marc Labat[1]

[1] LIP6, UPMC, Sorbonne Universités, Paris, France
{mathieu.vermeulen,jean-marc.labat}@lip6.fr
[2] LIG, Université Grenoble ALPES, Grenoble, France
nadine.mandran@imag.fr

**Abstract.** This paper proposes an analysis of student paths into the scenario graph for a learning game that uses a formal model of serious games understandable and usable by teachers. Screenwriting, implemented with a mental map, includes an expected path: the one that includes the most interesting nodes of the scenario graph from the point of view of the teacher, and achieves the training objectives. Through the analysis of the paths taken by the students, we will show the advantages and the benefits of this screenwriting. For that we indicate the different paths, the exit points (nodes presenting the case of abandonment of the student) and the various categories of paths (with achievement or non-achievement of training objectives). Finally, we propose solutions (tools and methods) to improve the reengineering process and the design of the scenario by the teachers.

**Keywords:** Learning games · Screenwriting · Design · Teachers · Reengineering

## 1 Introduction

The "serious games" term (or learning game for our case) has several definitions depending on the context and authors such as Abt [1] or Michael and Chen [16]. Alvarez and Djaouti provide a definition clarifying the ambiguity of the concept: a serious game is a computer application for which the original intention is to combine with consistency, serious aspects, in this case learning, with playful elements taken from the video game [2]. Many achievements have shown their interest in the transfer of skills and knowledge by developing the attractiveness and promoting the motivation to learners. Nonetheless, that interest is tempered by the lack of tools and methodologies for the design and production [11]. Meanwhile, the world of higher education is impacted, but with less enthusiasm. Teachers in higher education, even those that are convinced of the potential of digital education, have difficulties to create and to adapt learning games to their pedagogy [5]. Particularly, their implication in the design of the scenario is a crucial point of the learning game development [13]. In addition, we know that


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 321–330, 2016.
DOI: 10.1007/978-3-319-45153-4_24

reengineering of learning games and of their scenario is a criterion of adoption by the teachers [14].

The research pointed out that the reengineering process needs feedbacks on the usage of the students of the TEL Systems [3]. In our case, these feedbacks are logs files and data that describe the path of the student through the scenario graph. We interpret these data in order to compare the teacher's expected path with the student's path. We could use the term of "expected scenario" and "observed scenario" [8].

In this paper, through the use of a learning game presented in the next section, we analyse the data of 155 learners (we use this term rather than students in this paper due to the profile type of this population as describe in Sect. 3). They have used this game from 2016 March, 7th to 2016 March, 18th. This analysis would allow to make a reengineering process of these learning game with the teachers. With this, we search to make the game more efficient and to allow learners to acquire competencies, and for that to pass through the important nodes of the scenario graph. Another goal is to affine our methodology for designing learning games.

## 2 Development of the Learning Game

We have developed a learning game called "Les ECSPER" which allows to evaluate knowledge in Statistic and to acquire competencies about methodology in statistic problems in enterprise (Fig. 1).

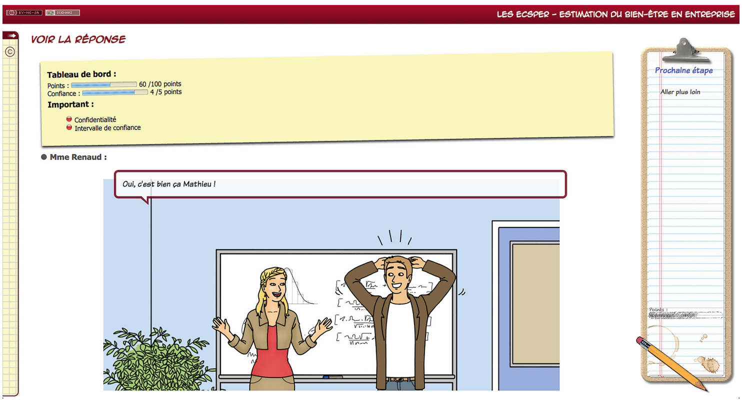



**Fig. 1.** A screen of the learning game Les ECSPER

The learner makes hypotheses and choices while playing as a young engineer employed as statistician in a big company. His/her mission is to estimate the wellness of the employees. We have two manners to evaluate the learner: a score for progression (points) and lives (3 lives in the beginning).

As we have already indicated, numerous works pointed out the difficulties for teachers to design and to create a learning game [4,10,12]. Particularly, if we focus on screenwriting and scenario, researches have been realized and some generic models and tools deal with this aspect of the design process. Marne et al. [12,14] have extracted from the study of different authoring tools three fundamental features about the scenario of learning games:

– The scenario is divided into components (and partially independent from each other),
– These components (e.g. levels or stages) are organized and connected by the hierarchical structure of their goals.
– The scenario components can dynamically adapt to the choices and performances of the learner (or "player").

Furthermore, we have designed the game following a conceptual framework named "the six facets" [15]. The facet "Problems and Progression" concerns "which problems to give the players to solve and in which order". This facet was a real challenge for the authors because it implies both the teachers and the game experts who must communicate with each other.

The scenario of Les ECSPER follows these features. Thus, it was inspired by gamebooks and divided into components; each of them is a case study with an educational goal. The teachers designed the scenario step by step with an iterative process and implemented it with a mental map and the tool XMind (Fig. 2). This tool is easy to use and can be used by the teachers who are not computer specialist, and of course by the game experts.

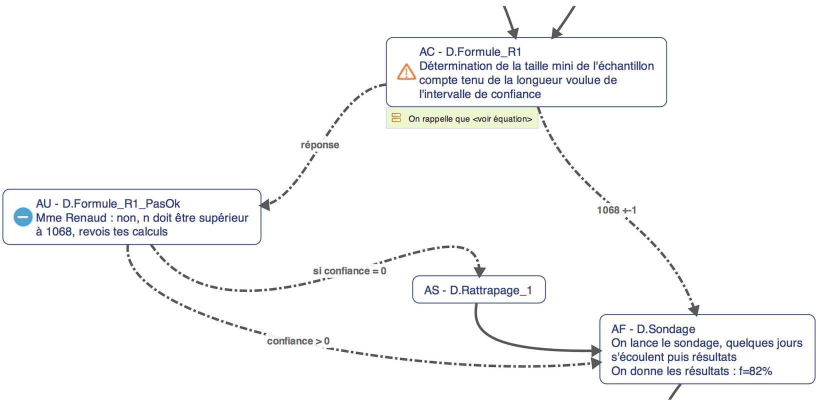



**Fig. 2.** An extract of the mind map describing the scenario graph

We obtained a scenario graph that contains the expected path e.g. the one which includes the most interesting nodes of the scenario graph from the point of view of the teachers, and achieves the training objectives. It is composed by different types of nodes, which were defined by their objective. We count a very large number of paths, more than 16000, and 65 nodes; each of them is a case study, an expositive step or an end step. Some nodes have a large degree that represents numerous hypotheses.

## 3 Context and Background of the Learners

The learning game Les ECSPER have been designed to be an activity of a MOOC (Massive Open Online Course) named “Statistique pour l’ingénieur” (in French and means statistic for engineers). This MOOC is deployed on FUN (e.g. France Université Numérique), based on the LMS Open edX (Learning Management System). The teachers want to multiply activities and to include learning games in this MOOC [7]. The game is deployed on the LMS Moodle. We have implemented an API on this LMS to record different data such as steps (nodes) ordered chronologically, time spent by step and scores. With that data, we can induce the paths taken by each learners. The learners accede anonymously to Les ECSPER through the IMS LTI protocol [6] with a unique number (id). If another session is made for an id, it count as a second (or third) try associated with the same id. Thus, each record is associated to one unique user and available in Moodle in a .csv file. In sum, FUN provides information about the learners profile and Moodle provides data about the use of the game. However, we can’t link these data because the policy of FUN imposes anonymity.

The LMS FUN provides the profile of the 6958 learners registered (as noticed at 2016 March, 24). We can state that a large part of them have at least a master’s or professional degree (64 %); they have or search a job (68 %) and have more than 25 years old (74 %).

## 4 Analysis

As we have already indicated, the data were recorded from 2016 March, 7 to 2016 March, 18. Learners were on the third week of the MOOC and the prerequisites were viewed since 2 weeks through different classical activities (lectures, videos, documents and quiz). These data show a lot of different paths: 139 paths for 155 unique learners. We have categorized these paths and compared them to the expected path design by the teachers. The first step of our analyse is to count, for each learner, the number of viewed nodes (Fig. 3). We can notice that the learners who finished the game have viewed more than 38 nodes. This reflects the effort that was needed to achieve the game. We have choosen to describe 4 categories in this paper:

1. The dropouts, learners who have viewed 8 nodes or less.
2. The learners who have achieved the game with success.
3. The learners who have achieved the game with a “game over”.
4. The abandonments, learners who haven’t achieved the game (they have enventually done another game session) and made more than 35 % of the game.

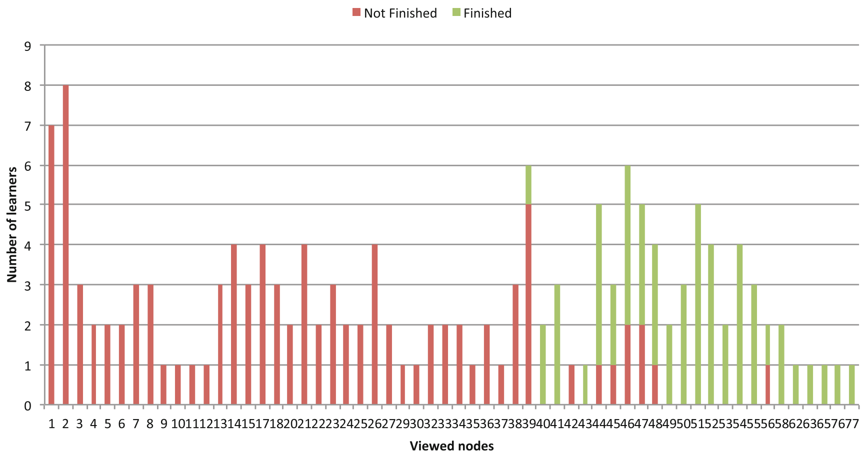



**Fig. 3.** Number of learners per number of viewed nodes

This choice allows us to show the interest of this analysis for the teachers in order to make a reengineering of the learning game. For that we pointed out a major mistake of the learners made late in the game.

### 4.1 First Category: The Dropouts

30 learners have viewed less than 8 steps (which represent 10 % of the game Les ECSPER). Among these: 5 have followed the 7 first nodes of the expected path and have attained the key node of the game (the working place of the main character); 3 have chosen the wrong answer at the first quiz. The exit node directly follows this step, when they have known they have made a mistake; 15 have only saw the first and or the second step without getting into the game. It will be interesting at the end of the MOOC if they make another try and what path they will take. If they will not do this, the design of the first and second steps would be revised.

### 4.2 Second Category: The Good Paths

44 learners have finished the game with one of the three good endings. It means that they have achieved the training objectives but they could have made some mistakes (only one learner have achieved the game with the highest score: "Excellent!").

Thus, among these learners paths, we have isolated an interesting fact. 25 learners of this category have made a major mistake at the step AB, a quiz step. We have extracted a sequence of nodes made by these 25 learners: [AB, AC, AU, AF] (Fig. 4). On Fig. 4, the expected path is the edge in light grey between the node AB and AF: [AB, AF]. Another sequence, [AB, AC, AU, AS, AF] includes the node AS which gives a second chance to the learner if he/she has one life left. We coloured the node AB and AC with the same colour (red) because this is the same screen for

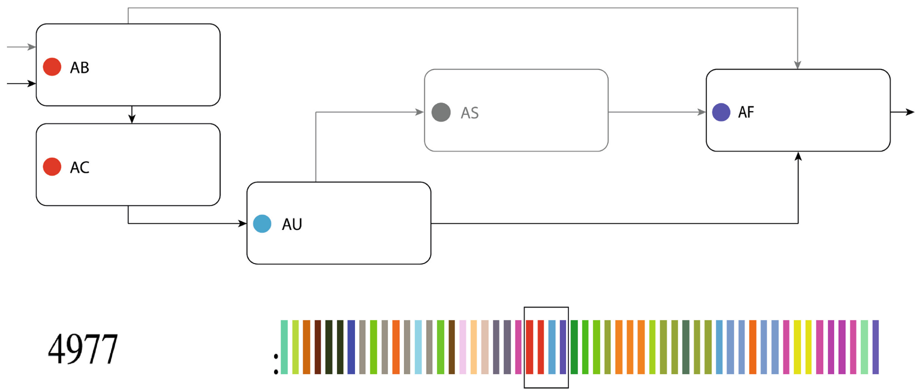



**Fig. 4.** A major mistake in the path of one learner (represented with Undertracks) and the part of the Mind Map associated (Color figure online)

the learner but with a feedback (this is a second chance to find the true answer). We focus about this sequence in this analysis. Data visualization are built with Undertracks platform [9] (http://undertracks.imag.fr), a tool to capitalize and analyse data; it is maintained by LIG (Laboratoire Informatique de Grenoble).

### 4.3 Third Category: The Bad Paths

9 learners have finished the game with a "game over". In fact, they have made 4 mistakes that the teachers qualified as major mistakes (thus learners have lost one life for each of these). 5 of them have spent short time on the game: less than 20 min. The authors have designed the game for at least one hour of play and work. We think that these learners just wanted to make a try and to see this new activity. We'll analyse their future try, if they'll make it, in a future work. On the Fig. 5, we note that the sequence [AB, AC, AU, AF] (we pointed out this with a black box) was always in the path.

### 4.4 Fourth Category: The Abandonments

About the other learners, we have analysed the paths of the 40 learners who have made more than 35 % of the game. We have pointed out 2 situations of abandonment. A reason seems to be the type of the exit node: 29 have left the game on a quiz step noted "Quiz" and in blue in the Fig. 6. These nodes depict the important steps of the scenario. They are the key nodes of the predicted scenario defined by the teachers. Regarding the sequence [AB, AC, AU, AF], we have observed that 25 learners sequences of this category contained this sequence of nodes. The node AB is a quiz step, this is a typical example of key node. 6 learners of this category exited Les ECSPER at this step (Fig. 6). 13 learners exited the game at the step just following the node AB and all of them lost a point of life with the sequence [AB, AC, AU, AF]. Another reason is the case where they were in a situation of

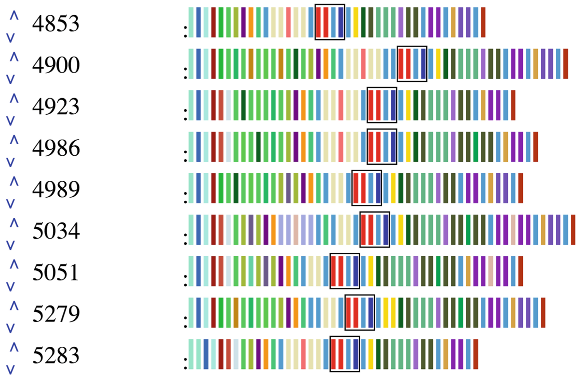



**Fig. 5.** Data visualization with Undertracks of the 9 learners who finished the game with a Game Over with sequence [AB, AC, AU, AF]

failure. Thus, 70 % of learners of this category (Table 1) have lost 2 or more lives (the lives represent the level of confidence of the player). And of course, they know that if they lost 4 lives, the game is over. We don't know if they have left the game definitely or if they have made another try. The data exists and we'll analyse them in another paper.

Finally we could say that neither the time spent to play nor the score that calculates progression determine a situation of abandonment.

### 4.5 To Sum up

If we focus on the sequence [AB, AC, AU, AF], we observe that the learners lost often a point of life at the node AB (Table 2) even between those who end the game successfully (57 % have made a mistake at node AB). This fact was noticed to the teachers, and this is an important element for a future reengineering of this

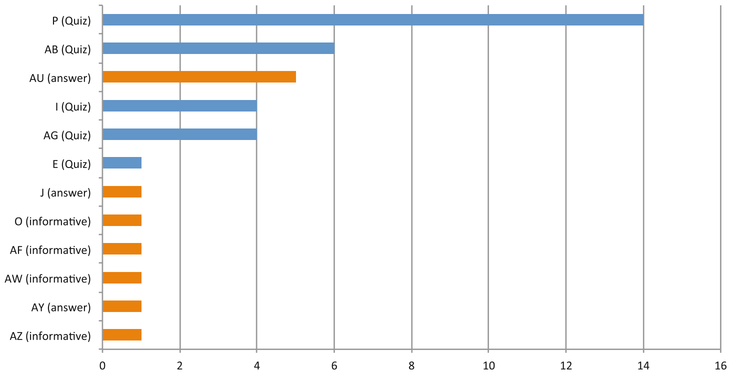



**Fig. 6.** Exit nodes for the learners who have made more than 35 % of the game

**Table 1.** Lives left when the learners exit the game

| No life left | 1 life left | 2 lives left | 3 lives left | 4 life left |
|---|---|---|---|---|
| 5 % | 20 % | 45 % | 18 % | 13 % |

**Table 2.** Learners that lose life at the node AB

| Good Paths | Bad Paths | Abandonments |
|---|---|---|
| 57 % | 100 % | 63 % |

learning game. Thus, for this sequence we have worked with one of the tesachers, and he has proposed to add a step before the node AB with reminders. This new node will be integrated to the new version of the learning game.

## 5 Perspectives

We pointed out elements about the scenario that we could avoid. Three points need our attention:

– We have seen that the first and second steps will have to be revising to avoid an immediate exit of the game.
– The teachers could prepare the learners when the nodes include a quiz (or evaluation). To do that, they could add another fun elements and include them in the screenwriting. Also, they could rewrite the text of the quiz if it is needed.
– Finally, they would improve the feedback, especially when the learner make a mistake.

To improve all these points and to ease the reengineering, we would make a qualitative analyse of these data with the teachers. When they design this type of learning game, they could improve the game after the first real use. Thus, we must allow them to rework the screenwriting and so, we could make a reengineering of the learning game. Furthermore, we'll use a model and a tool created by Bertrand Marne, named respectively MoPPLiq and APPLiq [13,14]. MoPPLiq is a generic model able to describe the playful and educational aspects of the scenario of learning game and makes the scenario understandable and capable of being manipulated by teachers. This model comes with a tool called APPLiq enabling manipulation of the scenario to fit it into the educational background of teachers. This tool would allow us to ease the process of screenwriting, as the iterative step (as seen above). Indeed, in this step of design, we need a tool that enables the manipulation of the scenario. We define actually a formal model and predefined templates to ease the design phase, particularly the screenwriting. This model will guide teachers more closely in designing serious games.

The MOOC "Statistique pour l'ingénieur" is always open at the time of writing this paper and there will be another session in September 2016 which will increment the quantity of data. We have already a second (and sometimes a third) path recorded for learners who have not achieved the game. We prepare another paper including analyse of these new data and the qualitative analyse made with the teachers.

**Acknowledgments.** This work was supported in part by the Institut Mines Telecom, UNISCIEL and Mines Douai. We would like to thank them for their support in the development of this project. The authors would like to thank Gaëlle Guigon, Carole Portillo and Rémy Pinot for their very helpful support to collect all this data; and of course the authors would like to thank the teachers who designed the learning game Les ECSPER, Michel Lecomte et Frédéric Delacroix based on an idea from Jean-Loup Cordonnier.

## References

1. Abt, C.C.: Serious Games. University Press of America, Lanham (1987)
2. Alvarez, J., Djaouti, D.: An introduction to serious game definitions and concepts. In: Serious Games & Simulation for Risks Management, p. 11 (2011)
3. Choquet, C., Iksal, S.: Modeling tracks for the model driven re-engineering of a tel system. J. Interact. Learn. Res. **18**(2), 161 (2007)
4. Djaouti, D.: Serious game design: considérations théoriques et techniques sur la création de jeux vidéo à vocation utilitaire. Ph.D. thesis (2011)
5. Egenfeldt-Nielsen, S.: Practical barriers in using educational computer games. Horizon **12**(1), 18–21 (2004)
6. Forment, M.A., Guerrero, M.J.C., González, M.Á.C., Peñalvo, F.J.G., Severance, C.: Interoperability for LMS: the missing piece to become the common place for elearning innovation. In: Lytras, M.D., et al. (eds.) WSKS 2009. LNCS, vol. 5736, pp. 286–295. Springer, Heidelberg (2009)
7. Freire, M., Blanco, A.D., Fernandez-Manjon, B.: Serious games as edX MOOC activities. In: 2014 IEEE Global Engineering Education Conference (EDUCON), pp. 867–871, April 2014
8. Lejeune, A., Pernin, J.P.: A taxonomy for scenario-based engineering. In: CELDA, pp. 249–256 (2004)
9. Mandran, N., Ortega, M., Luengo, V., Bouhineau, D.: DOP8: merging both data and analysis operators life cycles for technology enhanced learning. In: Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, pp. 213–217. ACM (2015)
10. Marfisi-Schottman, I., Labat, J.M., Carron, T.: Building on the case teaching method to generate learning games relevant to numerous educational fields. In: 2013 IEEE 13th International Conference on Advanced Learning Technologies (ICALT), pp. 156–160. IEEE (2013)
11. Mariais, C., Michau, F., Pernin, J.P., Mandran, N.: Learning role-playing games: méthodologie et formalisme de description pour l'assistance à la conception-Premiers résultats d'expérimentation. In: Environnements Informatiques pour l'Apprentissage Humain, Conférence EIAH 2011. pp. 95–107. Editions de l'UMONS (2011)
12. Marne, B.: Modèles et outils pour la conception de jeux sérieux: une approche meta-design. Ph.D. thesis (2014)

13. Marne, B., Carron, T., Labat, J.M., Marfisi-Schottman, I.: MoPPLiq: a model for pedagogical adaptation of serious game scenarios. In: 2013 IEEE 13th International Conference on Advanced Learning Technologies (ICALT), pp. 291–293. IEEE (2013)
14. Marne, B., Labat, J.M.: Model and authoring tool to help teachers adapt serious games to their educational contexts. Int. J. Learn. Technol. **9**(2), 161–180 (2014)
15. Marne, B., Wisdom, J., Huynh-Kim-Bang, B., Labat, J.M.: The six facets of serious game design: a methodology enhanced by our design pattern library. In: 21st Century Learning for 21st Century Skills, pp. 208–221 (2012)
16. Michael, D.R., Chen, S.L.: Serious Games: Games that Educate, Train, and Inform. Muska & Lipman/Premier-Trade, New York (2005)

# Adaptive Testing Using a General Diagnostic Model

Jill-Jênn Vie[1](✉), Fabrice Popineau[1], Yolaine Bourda[1], and Éric Bruillard[2]

[1] LRI – Bât. 650 Ada Lovelace, Université Paris-Sud, 91405 Orsay, France
{jjv,popineau,bourda}@lri.fr

[2] ENS Cachan – Bât. Cournot, 61 Avenue du Président Wilson, 94235 Cachan, France
eric.bruillard@ens-cachan.fr

**Abstract.** In online learning platforms such as MOOCs, computerized assessment needs to be optimized in order to prevent boredom and dropout of learners. Indeed, they should spend as little time as possible in tests and still receive valuable feedback. It is actually possible to reduce the number of questions for the same accuracy with computerized adaptive testing (CAT): asking the next question according to the past performance of the examinee. CAT algorithms are divided in two categories: summative CATs, that measure the level of examinees, and formative CATs, that provide feedback to the examinees at the end of the test by specifying which knowledge components need further work. In this paper, we formalize the problem of test-size reduction by predicting student performance, and propose a new hybrid CAT algorithm `GenMA` based on the general diagnostic model, that is both summative and formative. Using real datasets, we compare our model to popular CAT models and show that `GenMA` achieves better accuracy while using fewer questions than the existing models.

## 1 Introduction

Computerized assessments are becoming increasingly popular. Meanwhile, the Obama administration has urged schools to make exams less onerous and more purposeful [1]. To reduce over-testing, we need to optimize the time spent on tests, asking only informative questions about the learners' ability or knowledge. This is the idea behind Computerized Adaptive Testing (CAT): selecting the next question to ask according to the previous answers of the examinee. As an example, 238,536 such adaptive tests have been administered by the Graduate Management Admission Council in 2012–2013 [2] and adaptive assessment is getting more and more necessary in the current age of massive online open courses (MOOC), in order to minimize dropout.

Primarily, CATs have been relying on item response theory, that provides a framework to measure effectively scores called *latent traits* in order to rank students on a scale. The idea is to calibrate the difficulty of questions using a history of people having already taken the test. In 2001, The No Child Left Behind Act


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 331–339, 2016.
DOI: 10.1007/978-3-319-45153-4_25

has called for more formative assessments, providing feedback to learners and teachers at the end of the test. Such formative assessments may detect students with cognitive disabilities or simply build a profile that specifies which knowledge components seem to be mastered and which ones do not. A straightforward application would be a personal assistant that asks a few questions, then highlights the points that need further work, and possibly suggests useful material for remediation. In 2003, to address this need, new CATs have been developed relying on cognitive diagnosis models, the most popular being the DINA model [3] based on a q-matrix: a matrix that maps items (aka questions) to knowledge components involved in their resolution. Other cognitive models, less known, tend to unify scoring and formative assessments, but to date, they have not be used for adaptive testing [4].

In this paper, we formalize the problem of *test-size reduction by predicting student performance* (TeSR-PSP), inspired by [5] and present a new algorithm for CAT called `GenMA`, based on the general diagnostic model [6] that encompasses both the recovery of the latent knowledge components (KC) and for each KC, a degree of proficiency represented by a difficulty parameter.

To compare our algorithm for adaptive testing to the other ones mentioned above, we present an experimental protocol and execute it on real data. We show that `GenMA` outperforms existing models.

Our paper is organized as follows. In Sect. 1, we present the related work in CAT models. In Sect. 2, we formalize the problem of test-size reduction by predicting student performance and our new model, `GenMA`. In Sect. 3, we present an experimental protocol devised to compare these models, the real dataset used for evaluation, and our results. Finally, we discuss further work.

## 2 Related Work

### 2.1 Computerized Adaptive Testing and the Problem of Test-Size Reduction

In a non-adaptive test, every examinee receives the same set of questions. In an adaptive test, the next item asked by the system is chosen according to a certain criterion (the *selection item rule*), until the termination criterion holds, for example until a threshold over the parameters of the chosen model is guaranteed. Therefore, asking questions in an adaptive way means that the next question can be chosen as a function of the previous responses of the examinee.

In the problem of *test-size reduction* [5], one wants to reduce the number of questions asked as much as possible. Given a student model, we thus need to carefully choose the next question in order to still recover the model parameters. Formally, we want to decrease as much as possible the distance between the estimated and true user parameters after each question.

But in real data analysis, the true user parameters are unknown. For their evaluation, [5] replace the true user parameters with the estimated parameters they obtain after all questions have been asked, even if those estimated parameters do not fit the data at all.

In what follows, we will assume $n$ learners take a test of a total of $m$ questions. We assume the student data is dichotomous, which means every student either fails on succeeds over an item.

### 2.2 Item Response Theory (IRT)

The most simple model in item response theory for adaptive testing is the Rasch model, also known as 1-parameter logistic model. It models the behavior of a learner $i \in \{1, \ldots, n\}$ with a single parameter $\theta_i \in \mathbf{R}$ called ability, and models the item $j \in \{1, \ldots, m\}$ with a single parameter $d_j \in \mathbf{R}$ called difficulty. The tendency for a learner to solve an item only depends on the difference between the difficulty and the ability:

$$Pr(\text{“learner } i \text{ answers item } j\text{”}) = \Phi(\theta_i - d_j)$$

where $\Phi : x \mapsto 1/(1 + e^{-x})$ is the logistic function.

Being a unidimensional model, the Rasch model alone is not suitable for fine-grained feedback: it can only provide the level of the examinee at the end of the test. Still, it is really popular because of its simplicity, its stability and its sound mathematical framework [7,8]. Also, [9] has showed that if the items are splitted into categories, the Rasch model is enough to provide to the examinee a useful deviation profile, specifying which category subscores were lower or higher than expected.

It is natural to extend the Rasch model to multidimensional abilities. In Multidimensional Item Response Theory (MIRT) [10], both learners and items are modelled by vectors of a certain dimension $d$, and the tendency for a learner to solve an item depends only on the dot product of those vectors. Thus, if learner $i \in \{1, \ldots, n\}$ is modelled by vector $\boldsymbol{\theta_i}$ and item $j \in \{1, \ldots, m\}$ is modelled by vector $\boldsymbol{d_j}$:

$$Pr(\text{“learner } i \text{ answers item } j\text{”}) = \Phi(\boldsymbol{\theta_i} \cdot \boldsymbol{d_j}).$$

Thus, a learner has greater chance to solve items correlated with its ability. Nevertheless, those richer models involve many more parameters, and have proven to be much harder to converge [7].

### 2.3 Cognitive Diagnosis

[11] have used adaptive testing strategies applied to cognitive diagnosis (CD) models, notably the DINA model. These cognitive models rely on a specification of the knowledge components (KCs) involved in the resolution of the items proposed in the test, in the form of a q-matrix, which simply maps items to KCs: $q_{ik}$ is 1 if item $i$ involves the KC $k$, 0 otherwise. Several algorithms have been proposed and compared for CD-CATs, using for example Kullback-Leibler divergence [3,12].

If there are $K$ KCs involved in a test, the learner can be modelled by a vector of $K$ bits called *state*, specifying which KCs are mastered and which ones do not.

Knowing the state of a learner, we can infer his performance over the different questions of the test. Slip and guess parameters capture careless errors. Throughout the assessment, a probability distribution over the $2^K$ states is maintained, and updated after each answer in order to fit the learner's behavior. In the particular case of the DINA model, the KCs involved in the resolution of an item are required to solve it. If the learner masters all KCs required, it still has a probability to slip over the question; if it lacks a KC, it still has a probability to guess correctly the answer.

# 3 Our Contribution

## 3.1 TeSR-PSP: Test-Size Reduction by Predicting Student Performance

In this paper, we propose a new problem called *test-size reduction by predicting student performance*: if we can ask only $k$ questions in an adaptive way, which ones should we pick so as to predict the examinee's performance over the remaining questions of the test?

Usually, adaptive tests keep going until a suitable confidence interval over the learner parameters is obtained. In our case, we want to specify in advance the maximal number of questions that will be asked to every student, in order to prevent boredom from the learner.

## 3.2 `GenMA`: Using the General Diagnostic Model for Adaptive Testing

[6] has proposed a unified model that takes many existing IRT models and cognitive models as special cases: the general diagnostic model for partial credit data:

$$Pr(\text{“learner } i \text{ answers item } j\text{”}) = \Phi\left(\beta_i + \sum_{k=1}^{K} \theta_{ik} q_{jk} d_{jk}\right)$$

where $K$ is the number of KCs involved in the test, $\beta_i$ is the main ability of learner $i$, $\theta_{ik}$ its ability for KC $k$, $q_{jk}$ is the $(j,k)$ entry of the q-matrix which is 1 if KC $k$ is involved in the resolution of item $j$, 0 otherwise, $d_{jk}$ the difficulty of item $j$ over KC $k$. Please note that this model is similar to the MIRT model specified above, but only parameters that correspond to a nonzero entry in the q-matrix are taken into account.

To the best of our knowledge, this model has not been used in adaptive testing [4]. This is what we present in this paper: `GenMA` relies on a general diagnostic model, thus requires the specification of a q-matrix by an expert. The parameters $d_{jk}$ for every item $j$ and KC $k$ are calibrated using the history of answers from a test and the Metropolis-Hastings Robbins-Monro algorithm [13,14]. For the selection item rule of `GenMA`, we choose to maximize the Fisher information at each step, details of the implementation can be found in [13].

The problem TeSR-PSP becomes: after $k$ questions asked to a certain learner $i$, how to estimate its main ability $\beta_i$ and ability for each KC $\theta_{ik}$ that can explain its behavior throughout the test?

In real tests, items usually rely on only few KCs, hence there are fewer parameters to estimate than in a general MIRT model, which explains why the convergence is easy to obtain for `GenMA`. We can thus use the general diagnostic model to create an adaptive test that makes best of possible worlds: providing feedback under the form of degrees of proficiency over several KCs at the end of test, represented by the vector $\boldsymbol{\theta_i} = (\theta_{i1}, \ldots, \theta_{iK})$, and being easy to converge. `GenMA` is both summative and formative, thus a hybrid model. Such feedback can be aggregated at various levels (e.g., from student, to class, to school, to city, to country) in order to enable decision-making [9,15].

## 4 Evaluation

In this section, we detail the experimental protocol used to compare the following models for TeSR-PSP: the Rasch model, the DINA model and `GenMA`. For the sake of equality, we decide to define the same selection item rule for all models: all of them pick the question that maximizes Fisher information, which means the question of estimated probability closest to 1/2.

### 4.1 Experimental Protocol

Our experimental protocol is based on double cross-validation. For each experiment, we need:

- a *train* student set, in order to calibrate the parameters of our model (for example, the difficulty of questions in the case of the Rasch model);
- a *test* student set, which will take our adaptive test;
- a *validation* question set $V_Q$, which is used for training, but kept out of the adaptive tests, used only to evaluate the prediction of performance of the students from the test set.

To evaluate the score of a model for our problem, we first train it using the train student set. Then, for each student from the test set, we let the model pick a question, we reveal the student answer for this question, the model updates its parameters accordingly and outputs a probability of correctly answering the questions from the validation set, that we can evaluate using negative log-likelihood, hereby denoted as *error*:

$$error(pred, truth) = \sum_{q \in V_Q} truth_q \log pred_q + (1 - truth_q) \log(1 - pred_q).$$

Then the model picks the second question, and so on. Thus, after $k$ questions we can compute a prediction error over the validation question set for every model and every test student.

### 4.2 Real Dataset: Fraction subtraction

Tatsuoka's fraction subtraction dataset contains the dichotomous responses of 536 middle school students over 20 fraction subtraction test items. The corresponding q-matrix maps the 20 items to the following 8 knowledge components (KCs):

- convert a whole number to a fraction,
- separate a whole number from a fraction,
- simplify before subtracting,
- find a common denominator,
- borrow from whole number part,
- column borrow to subtract the second numerator from the first,
- subtract numerators,
- reduce answers to simplest form.

The cross-validation was performed using a random split into training and test sets: the split was 5-fold over the students and 4-fold over the questions: each student set had a size of 20 % while each validation question set had a size of 25 % so 5 questions, therefore there were 20 experiments in total, of which the mean error was computed.

### 4.3 Implementation Details

Our Rasch model implementation comes from the `ltm` R package. We made our custom implementation of the DINA model but the slip and guess calibration is held by the `CDM` package. `GenMA` is built upon the `mirt` package [13].

### 4.4 Results

For each number of questions asked from 1 to 15, we plotted the mean error of each model (Rasch model, DINA model and `GenMA`) over the test student set in Fig. 1, and as an insight easier to comprehend, the mean number of questions incorrectly guessed in Fig. 2.

Figure 1 shows that 8 questions over 15 are enough for the Rasch model to converge on the fraction subtraction dataset. Figure 2 shows that no matter how many questions are asked, the Rasch and DINA models can't predict correctly more than 4 questions in average over 5 in the validation question set, while `GenMA` can achieve this accuracy with only 4 questions, then keeps on improving its predictions. The DINA model takes a long time to converge because the first questions require a single KC, therefore they do not bring a lot of information about the user state. But still, the simplest, unidimensional Rasch model performs surprisingly well compared to `GenMA` which is over 8 dimensions, one per KC of the q-matrix.

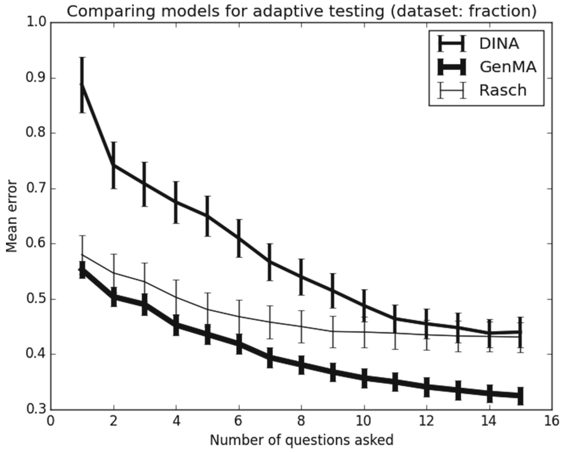



**Fig. 1.** Comparing adaptive testing models. Evolution of error (negative log-likelihood) over the validation question set, after a certain number of questions have been asked.

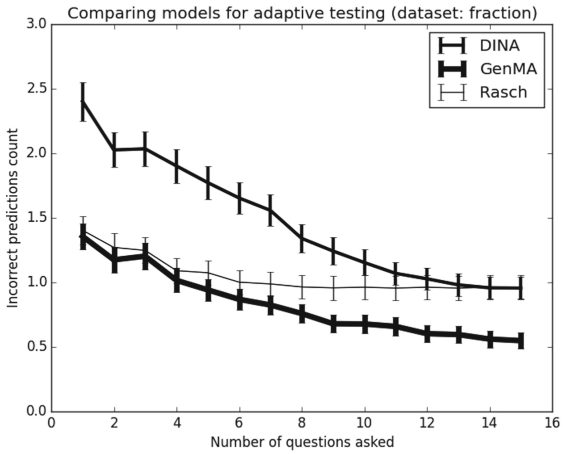



**Fig. 2.** Comparing adaptive testing models. Evolution of the number of questions in the validation question set incorrectly predicted, after a certain number of questions have been asked.

## 5 Conclusion and Future Work

In this paper, we formulated the problem of test-size reduction by predicting student performance, and presented our new adaptive testing algorithm `GenMA` to tackle it, based on the general diagnostic model. As this model is richer than other models as Rasch or DINA, it could be prone to overfitting: having more parameters, it may have a good score over train data but poor generalization over the test data. But we showed it actually achieves a better accuracy at predicting student performance, using fewer questions than the existing models on a real dataset.

The idea of a hybrid model combining several KCs and weights for each of them is not new: MIRT models can be seen this way, but there are many parameters to estimate, leading to convergence issues. [5] presented sparse factor analysis (SPARFA), a model that combines q-matrices and weights but their KCs are specified automatically, not by experts, thus it is not possible to provide a feedback at the end of the test.

In order to overcome the complexity of $O(2^K)$ of the DINA model, some knowledge representations such as Attribute Hierarchy Model [16,17] or Knowledge Space Theory [18,19] have been devised, relying on dependencies over KCs in the form of a directed acyclic graph. We would like to compare these approaches to `GenMA`. Also, richer models using ontologies [20,21] are a promising direction of research.

**Acknowledgements.** This work is supported by the Paris-Saclay Institut de la Société Numérique funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02.

## References

1. Zernike, K.: Obama administration calls for limits on testing in schools (2015)
2. Council, G.M.A.: Profile of GMAT Candidates - Executive Summary (2013)
3. Huebner, A.: An overview of recent developments in cognitive diagnostic computer adaptive assessments. Pract. Assess. Res. Eval. **15**(3), n3 (2010)
4. Yan, D., von Davier, A.A., Lewis, C.: Computerized multistage testing (2014)
5. Lan, A.S., Waters, A.E., Studer, C., Baraniuk, R.G.: Sparse factor analysis for learning and content analytics. J. Mach. Learn. Res. **15**, 1959–2008 (2014)
6. Davier, M.: A general diagnostic model applied to language testing data. ETS Research Report Series, p. i-35 (2005)
7. Desmarais, M.C., Baker, R.S.: A review of recent advances in learner and skill modeling in intelligent learning environments. User Model. User Adap. Inter. **22**, 9–38 (2012)
8. Bergner, Y., Droschler, S., Kortemeyer, G., Rayyan, S., Seaton, D., Pritchard, D.E.: Model-based collaborative filtering analysis of student response data: machine-learning item response theory. In: International Educational Data Mining Society (2012)
9. Verhelst, N.D.: Profile analysis: a closer look at the PISA 2000 reading data. Scand. J. Educ. Res. **56**, 315–332 (2012)
10. Reckase, M.: Multidimensional Item Response Theory. Statistics for Social and Behavioral Sciences. Springer, New York (2009)
11. Xu, X., Chang, H., Douglas, J.: A simulation study to compare CAT strategies for cognitive diagnosis. In: Annual Meeting of the American Educational Research Association, Chicago (2003)
12. Cheng, Y.: When cognitive diagnosis meets computerized adaptive testing: CD-CAT. Psychometrika **74**, 619–632 (2009)
13. Chalmers, R.P.: mirt: A multidimensional item response theory package for the R environment. J. Stat. Softw. **48**, 1–29 (2012)

14. Cai, L.: Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. J. Educ. Behavi. Stat. **35**, 307–335 (2010)
15. Shute, V., Leighton, J.P., Jang, E.E., Chu, M.-W.: Advances in the science of assessment. Educ. Assess. **21**(1), 34–59 (2015)
16. Leighton, J.P., Gierl, M.J., Hunka, S.M.: The attribute hierarchy method for cognitive assessment: a variation on Tatsuoka's rule-space approach. J. Educ. Measur. **41**, 205–237 (2004)
17. Rupp, A., Levy, R., Dicerbo, K.E., Sweet, S.J., Crawford, A.V., Calico, T., Benson, M., Fay, D., Kunze, K.L., Mislevy, R.J.: Putting ECD into practice: the interplay of theory and data in evidence models within a digital learning environment. JEDM-J. Educ. Data Min. **4**, 49–110 (2012)
18. Doignon, J.-P., Falmagne, J.-C.: Knowledge Spaces. Springer Science & Business Media, New York (2012)
19. Lynch, D., Howlin, C.P.: Real world usage of an adaptive testing algorithm to uncover latent knowledge (2014)
20. Mandin, S., Guin, N.: Basing learner modelling on an ontology of knowledge and skills. In: 2014 IEEE 14th International Conference on Advanced learning technologies (iCALT), pp. 321–323. IEEE (2014)
21. Kickmeier-Rust, M.D., Albert, D.: Competence-based knowledge space theory. In: Measuring and Visualizing Learning in the Information-Rich Classroom, p. 109 (2015)

# How Teachers Use Data to Help Students Learn: Contextual Inquiry for the Design of a Dashboard

Françeska Xhakaj(✉), Vincent Aleven, and Bruce M. McLaren

Human Computer Interaction Institute,
Carnegie Mellon University, Pittsburgh, PA, USA
{francesx, aleven, bmclaren}@cs.cmu.edu

**Abstract.** Although learning with Intelligent Tutoring Systems (ITS) has been well studied, little research has investigated what role teachers can play, if empowered with data. Many ITSs provide student performance reports, but they may not be designed to serve teachers' needs well, which is important for a well-designed dashboard. We investigated what student data is most helpful to teachers and how they use data to adjust and individualize instruction. Specifically, we conducted Contextual Inquiry interviews with teachers and used Interpretation Sessions and Affinity Diagramming to analyze the data. We found that teachers generate data on students' concept mastery, misconceptions and errors, and utilize data provided by ITSs and other software. Teachers use this data to drive instruction and remediate issues on an individual and class level. Our study uncovers how data can support teachers in helping students learn and provides a solid foundation and recommendations for designing a teacher's dashboard.

**Keywords:** Intelligent Tutoring Systems · Dashboard · Contextual Inquiry

## 1 Introduction

Much recent research focuses on designing and evaluating instructor dashboards [1, 4, 13, 20–22, 25]. It is reasonable to assume that the large amount of student interaction data that is routinely collected by educational technologies can be helpful to teachers and instructors, when presented on a dashboard in concise and actionable format. It might inform key decisions that teachers make, such as deciding the focus of discussion for a class lecture or identifying students who need one-on-one attention, with potentially a positive effect on student learning. Dashboards have been designed for a large variety of educational technologies such as multi-tabletop learning [20], collaborative learning in digital learning environments [22, 25], web-based distance courses [21], online courses [18], Intelligent Tutoring Systems [12], etc. The use of student data for instructional decision-making is not restricted to educational technology. For example, mastery learning, a highly effective data-driven instructional method, can be implemented without technology [15]. Also, in 2009 the Institute for Education Sciences (IES, part of the U.S. Department of Education) published a practice guide with recommendations for teachers on how to use data to inform instruction [11]. The IES


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 340–354, 2016.
DOI: 10.1007/978-3-319-45153-4_26

Practice Guide also points out, however, that there is limited scientific evidence that data-driven classroom practices actually improve educational outcomes, indicating a need for more research.

A very small number of studies suggest that a teacher dashboard can lead to improvements in students' learning outcomes. In one such study, the data-driven redesign of a statistics course yielded improved student learning in half the time [18]. A dashboard was one novel component of the redesigned course, but there were other changes as well, so the improvement cannot be attributed solely to the dashboard. Kelly et al. (2013) demonstrated benefits of teacher reports in a web-based tutoring system for middle school mathematics [14]. Relatedly, research with Course Signals system illustrates that using learning analytics to identify students who are falling behind, can have a positive effect on student retention [6]. In contrast to the current research, this project focused on university students and on feedback directly to students rather than teachers.

We are working towards creating a dashboard for middle and high school teachers who use an Intelligent Tutoring System (ITS) in their classrooms. ITSs are an advanced learning technology that provides detailed guidance to students during complex problem-solving practice, while being adaptive to student differences [5, 26, 29]. A number of meta-reviews indicate that ITS can enhance student learning in actual classrooms, compared to other forms of instruction [16, 19, 23, 24, 27]. ITS have also proven to be commercially viable [10]. Although ITSs typically produce a wealth of data about student learning, relatively little effort has been expended to investigate how this data can best be leveraged to help teachers help their students. Much more research has focused on how this information can be presented to students (e.g., in the form of an open learner model [9]).

A central assumption in our work is that in order to design an effective dashboard, it helps to understand how teachers use data about students' performance and learning in their day-to-day pedagogical decision-making. Therefore, we started off studying teachers' use of data using Contextual Inquiry, a method often used in user-centered design [8]. Although the use of user-centered design methods for dashboard design is quite common, we are unaware of prior studies that investigate teacher data needs through Contextual Inquiry, as we do in the current work. Some studies involved teachers as part of a user-driven design process that included interviews, prototypes and empirical evaluations of dashboard designs [20], surveys conducted to determine the information instructors may need [21], questionnaires used to evaluate and iterate on the features of a learning analytics tool for a web-based learning environment [3], or semi-structured interviews as part of the developing process of a web-based learning analytics tool with a dashboard component [7]. Another study applied participatory design and other design methods to create a dashboard for an educational game app [1]. Other studies do not mention teachers as part of the dashboard design, do not report on the methods used to interpret and select the data, or use theoretical work and previous literature to determine the appropriate design [4, 13, 25].

In this paper, we describe how we used Contextual Inquiry to better understand (1) what student data teachers need to be effective and (2) how teachers use data to inform and adjust their instruction. This work will inform the design of a teacher's dashboard in an ITS environment. We focus our design on Lynnette [17, 28], a tutor for middle school mathematics (grades 6–8) with a proven track record in helping students learn to solve linear equations.

## 2 Methodology

### 2.1 Gathering Data on Teacher Practices

We conducted Contextual Inquiry interviews to study teacher practices in using student data to adjust or individualize instruction. Contextual Inquiry is a user-centered design process, part of the Contextual Design method [8]. Contextual Inquiry is widely used to gather field data from users with the aim of understanding who the users are and how they work in their day-to-day basis. During a Contextual Inquiry interview, the researcher meets one-on-one with the participant and observes the participant conduct one of their daily activities in the participant's workplace. In this process, the researcher is considered to take up the role of an "apprentice" and the participant takes on the role of the "master." The researcher does not actively interview the participant with a set of pre-determined questions; rather, she or he observes the participant conduct one of the daily activities or normal tasks. The researcher asks questions occasionally to clarify and understand what and why the participant is doing something. Contextual Inquiry allows gathering of detailed and highly reliable information. It can reveal knowledge and information about the user's work that they themselves are unaware of.

We recruited teachers from various schools in our area that had previously participated in studies with our institution. We also requested assistance from Carnegie Learning to recruit teachers who currently use the Carnegie Learning (CL) tutor [10], a mathematics Cognitive Tutor – Cognitive Tutors are a type of ITS grounded in cognitive theory [5] – for grades 6–12 (Fig. 1). We ran Contextual Inquiry interviews with 6 teachers from 3 different schools in our area, namely, 4 middle-school teachers from a suburban, medium-achieving school (2 male and 2 female), 1 female high-school teacher from an urban, low-achieving school, and 1 female middle-school teacher from a suburban, medium-achieving school. Out of the teachers we interviewed, 2 teachers had used the CL tutor before in their classrooms and 1 teacher was using it currently. In addition, 2 other teachers had used in previous years other ITSs as part of various short-term studies from our institution. Lastly, all teachers used digital grade books or other technology in their classrooms. Thus, the teachers in our sample exhibit substantial variability regarding important variables such as whether they work in high- versus low-performing districts, whether they have experience with an ITS versus not, as well as the methods they devised themselves for using student data to guide their teaching, and their use of technology in their classrooms.

The focus of our Contextual Inquiry interviews was to observe the teacher in how and what data they generated on their students' performance (from materials such as exams, quizzes, assignments, etc.), and how they used this data to drive instruction and prepare for a class. After the Contextual Inquiry interview, we observed the teacher conduct the class they prepared for. During this process we silently observed in the classroom and followed up with an interview with the teacher with questions regarding the classroom observation. Due to constraints in the teachers' schedules, with some of

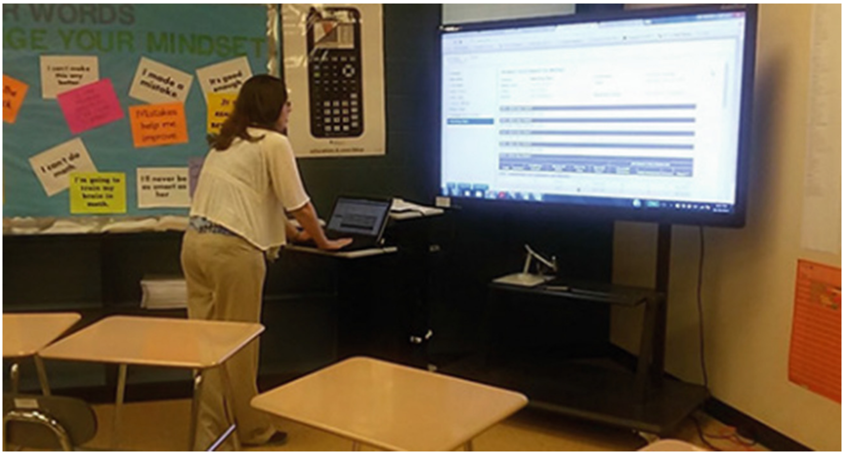


**Fig. 1.** Teacher during a Contextual Inquiry interview working on her laptop and smart screen with an ITS report.

the teachers we conducted the Contextual Inquiry interviews after doing a classroom observation, and then followed with an interview with the teacher with follow-up questions. With two of the teachers who participated in our study, we conducted Contextual Inquiry interviews on one teacher's previous use and another's current use of the reports generated by the CL tutor. These teachers reported that they used the CL tutor 2 days during the week, while the other 3 days they would have lectures in the classroom, outside the tutor environment. Lastly, we observed teachers' use of reports and other technology or software in the classroom. The Contextual Inquiry interviews were video recorded and resulted in a total of approximately 11.5 h of recording.

## 2.2 Interpretation Sessions and Affinity Diagramming

The video recordings of the Contextual Inquiry interviews were transcribed to text. A team composed of a PhD student (the first author of this paper) and a Master's student, both from our institution, worked through the transcriptions to analyze and synthesize the data from the transcribed interviews. Two standard techniques from Contextual Design were used: Interpretation Sessions and Affinity Diagramming. Interpretation Sessions are team-based tasks aimed to create a shared understanding of the collected data by recording on post-it notes, simple observations and key issues and insights from the interviews of each participant. Affinity Diagramming is a widely-used method that aims to discover patterns that define the whole population by grouping and organizing the post-it notes based on content similarity into a hierarchy that reveals common issues and themes [8]. The way of clustering the post-it notes into an Affinity Diagram has an element of subjectivity. However, the categories in this diagram emerge from clustering the data and are not pre-defined. The Affinity Diagram process does not require a coding schema or inter-rater reliability.

From 11.5 h of transcribed video interviews, we conducted several Interpretation Sessions, during which we walked through the transcribed video interviews for each participant and created post-it notes. We gathered approximately 2000 yellow notes, as illustrated in Fig. 2 (the two rows from the bottom). We initially followed the traditional Interpretation Session approach and recorded the observations in physical post-it notes. Given the large amount of interview data we had collected, we decided to instead store the notes electronically in a Google Spreadsheet. We also approached the Affinity Diagramming in a traditional way first, namely, by using printed copies of the digital notes and organizing them on large sheets of paper. However, given the large number of notes, we resorted to creating and keeping the Affinity Diagram in a Google Spreadsheet as well, as shown in Fig. 2.

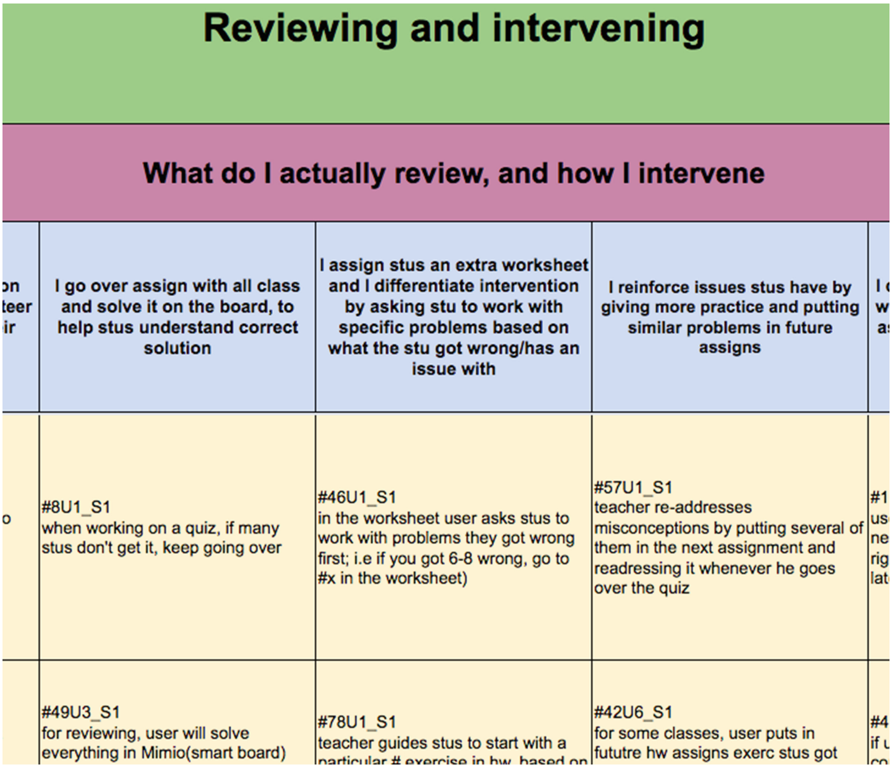



**Fig. 2.** Partial view of our final Affinity Diagram. (Color figure online)

We organized the yellow notes into categories based on patterns we identified and similarities in their content. Following the Affinity Diagramming technique, for each category, we recorded the synthesized content of all the yellow notes within the blue categories (third row from the top in Fig. 2). We then grouped together blue categories based on similarity of content and recorded the information they conveyed within the pink categories (second row from the top in Fig. 2). Lastly, we grouped pink categories and synthesized their content within the green categories (first row from the top in Fig. 2). Our final Affinity Diagram had 335 blue level categories (with 1–2 up to 12–14 yellow notes per category), 81 pink and 33 green level ones.

Based on the initial focus of our Contextual Inquiry interviews, namely how and what data teachers generate about their students' performance, and how they use this data to drive instruction and prepare for a class, we focused on the categories of the Affinity Diagram that contained the most important information relevant to this focus. We initially went through the final Affinity Diagram and selected the blue, pink and green categories that contained such information. We then recorded in two lists – what data teachers generate and how they use this data – a summary of the selected categories, in the form of short sentences and keywords. Each of the lists individually was then synthesized based on similarities in content, and our final results are presented in the following section.

## 3 Findings from the Contextual Inquiry Interviews

### 3.1 What Data Do Teachers Use to Help Students?

From the Contextual Inquiry interviews, we found that teachers continuously generate and use data on the progress and performance of their students. They also use data generated by technology such as the CL tutor or other software they use as part of their classroom instruction.

Teachers gather data when grading written student assignments, as well as by having one-on-one interactions with students during or outside of class. In particular, teachers pay attention to whether the overall class or individual students have mastered particular concepts. A concept can be an entire problem that exercises a skill (e.g., finding the greatest common denominator) or one of the steps that leads to the solution of the problem (e.g., graphing the direction of an inequality in the number line as part of graphing the inequality itself on the number line). In addition, teachers try to understand, on a class and individual student level, what causes students the most trouble, i.e., what are the most common misconceptions and errors.

Data provided by technology includes reports and analytics on student progress and performance in the CL tutor or in other software used by the teachers. For example, among the many reports that are offered by this tutor, the teachers we interviewed made the most use of the reports that give information on the overall class performance and on the individual student performance in the tutor. Teachers also pay attention to the number of skills students have mastered or not mastered and, less frequently, to time spent working in the tutor.

We also found that teachers use many different ways to record, keep track and organize student data. Some data gets initially recorded on paper and then is transferred to software. For example, some teachers recorded and kept grades in a paper grade book before transferring that information to a digital grade book. Other data on student performance is initially generated through software (such as CL tutor reports or other software reports), and the teacher prints and stores it offline. It is challenging for the teachers to keep track of and integrate both offline and online data.

Some (though not all) of the teachers we interviewed kept track of student errors and misconceptions at a surprising level of detail, as illustrated in Fig. 3. In the tally sheet on the left of Fig. 3, a teacher keeps track of the frequency of particular misconceptions (shown in columns) for each problem in an assignment (shown in rows). As the teacher describes, "*I will go through each problem and will start writing down where they made their errors. And I will just put tallies. And where I see different things I make sure I circle them so I can focus there whenever I am reviewing that*", referring to the misconceptions that most students had and thus should be discussed with that class. In addition, the teacher writes, at the top right of the tally sheet (covered), the name(s) of the student(s) who had the most trouble with a particular concept or concepts. To be consistent across periods, the teacher initially grades all tests or exams for each period and then creates the tally sheet template from the first period, copying it to the tally sheet for other periods. The teacher finishes tallying the sheet for one period before they move on to the next period. If the teacher notices a different or miscategorized misconception in another period, they go back and correct the tallies for that misconception in all the other periods.

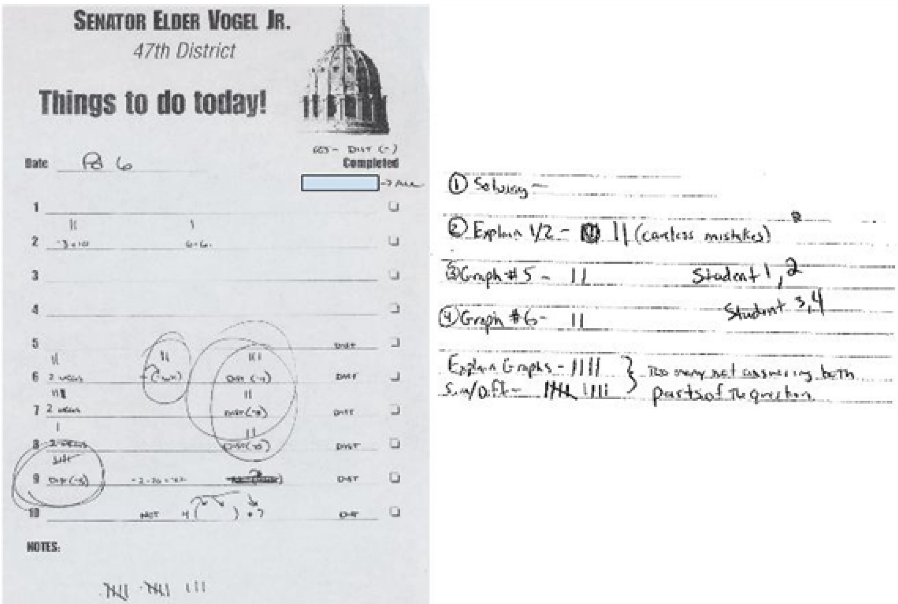



**Fig. 3.** Tally sheet from teacher 1 and teacher 2. Student identifiers have been removed.

Another teacher we interviewed uses the tally sheet on the right of Fig. 3 to tally students who got a problem (or parts of a problem) wrong in an assignment. Each problem in this particular assignment represented a high level concept (for example, exercise 1 was related to solving two inequalities, while exercise 2 asked students to explain the steps to those solutions). For some exercises, the teacher also notes in the tally sheet the reasons the students made the mistakes (for example, careless mistakes or not answering both parts of the question). Lastly, the teacher writes down the names of the students who they want to call on in class (represented by student 1, 2 and student 3, 4 in Fig. 3).

## 3.2 How Do Teachers Use Data to Drive Instruction and Help Students?

We found that teachers use data to drive and adjust their instruction in many ways. Most of the teachers differentiate how they use data and tune the level of detail to determine whether the best remedy is a classroom intervention or individual, one-on-one sessions with particular students.

### 3.2.1 Class-Level Decisions

**Decide to Move on to the Next Topic and Build on Current Concepts.** After generating data on the overall class performance in an assignment or test, the teacher analyzes it to assess the current status of the class and to decide whether to move on to the next topic. If, in the teacher's judgment, the majority of the class has mastered a concept or a set of concepts, the teacher decides to move on with the instruction and build on the current concept(s). As one teacher describes, "*there's times where I'm like 'Ok if they don't know this, I have to start here. But if they do know it, I can start here,' in a different position.*"

**Determine that the Class Needs Intervention.** The teacher notices when many students have not mastered certain concept(s), or when there are many different errors and issues in an assignment. The teacher decides to intervene and devote more time and attention in class to specific concepts, misconceptions or errors to help students remedy their issues.

**Identify the Focus of Intervention.** Based on the number of students who have not mastered the concept(s), or have misconceptions and errors, the teacher determines what is important to cover during a class lecture. The teacher can also create worksheets with exercises to allow students to practice the concepts they are missing or having the most trouble with.

**Plan What to Discuss and Cover in Each Period.** The teacher compares performance on an assignment across periods and adapts instruction (or what to cover in class) based on that period's performance. Sometimes the teacher covers only the topics that a period has the most trouble with; in other cases, the teacher might decide to discuss issues noticed from other periods in every class period.

**Display in Class Reports or Analytics from Software.** As students were working with the CL tutor, one teacher displayed anonymized class performance reports in front of the classroom, on a smart screen. The teacher aimed to support the students' learning and progress by seeing where they were compared to the other students in the class. In addition, displaying the report in class helped the teacher monitor the students' progress as the teacher walked around the class, while students were working with the tutor. The same teacher also displayed on the smart screen class analytics on students' performance generated from other software.

### 3.2.2 Decisions Regarding Individual Students or Groups of Students

**Decide Which Individual Students or Group of Students Need Special Attention.** The teacher identifies from the generated data individual students who have an issue with one or more concepts, have displayed the same misconception or error repeatedly, or are spending a lot of time but making little progress. The teacher records the individual students' names to work one-on-one with them. If the teacher notices that a group of students are having similar issues, the teacher might decide to work with them as a group.

**Determine the Focus of Intervention.** If the teacher does not know the reason why a student is having an issue, they spend time with that student trying to understand their problem(s). The teacher determines the focus of a mini-lecture or extra practice to help the student fix the issue and master the concept(s). The teacher will also call on the student during class time to prompt them to participate in discussion or problem solving for the concept(s) they are having trouble with. For groups of students, the teacher can decide to do a mini-lecture, or give practice worksheets, by differentiating intervention as to which student has to work with which exercise in the worksheet, based on individual issues identified.

**Show and Give Students Software Reports.** The teacher periodically shows, prints and gives students reports on their progress and performance over a given time period, in the CL tutor or other software used in the classroom. The teacher uses the data from these reports to update the students on their progress, what they still need to do, and what their grade is.

## 4 Breakdowns in Current Practices

From our interviews with the teachers, as well as from our data analysis, we noticed patterns of breakdowns in the current practices of generating and using data. We also noticed that the technology that some teachers use in the classroom is not always helpful, and can be inefficient.

**Teacher Adapts to Technology, Technology Does not Adapt to Teacher.** The CL tutor and other software provide more student data and reports than the teacher needs and can process. The teacher is selective in choosing among the provided reports, choosing only the data that is most useful to them. In addition, none of the technologies we observed provide data about misconceptions or student growth, which are hard to generate by hand. For example, one teacher used the Pennsylvania Value Added Assessment System to see students' growth from year to year. However, the teacher could use such reports only once per year, making it impossible to intervene in classes that the teacher would not be assigned to teach anymore. Another teacher said this about CL reports: "*It would actually be very useful [to see errors and misconceptions] because ... a lot of these reports I don't use frequently because it's not necessarily giving me what I need to know.*"

**Generating Data is Time Consuming and Effortful.** From grading student assignments to interacting with students on a class or individual level during and outside of class, the teacher continuously generates data on students. The teacher also spends time and effort in analyzing and drawing conclusions based on data from different sources, while differentiating the level of detail and instruction for the class or for individual students.

**Organizing, Integrating and Remembering Data from Different Sources is Challenging.** It takes time and effort to integrate data generated on paper with data from reports of tutors or other software. For example one teacher printed CL tutor reports and other software reports and organized them in a binder (Fig. 4). This teacher also put post-it notes on the binder and wrote things to remember on the printed reports, or highlighted in color particular students. Even without technology, we noticed that teachers integrate student data from different assignments and interactions with the students and, most of the time, keep track of this information in their heads.

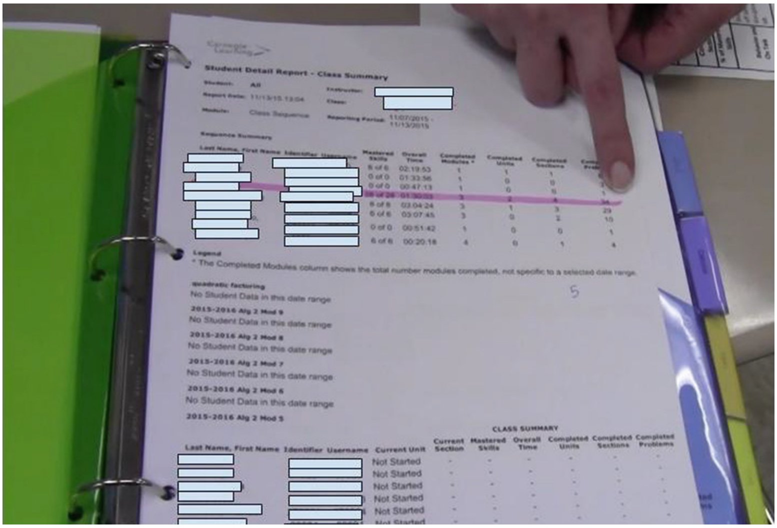



**Fig. 4.** Teacher prints and stores reports from CL tutor and other software in a binder offline. Student names and identifiers have been covered.

**Creating Materials for Intervention is Difficult.** The teacher has to spend time and effort to create or find the necessary materials for a mini-lecture or problems and exercises for a practice worksheet. One teacher used various online sites to find and give problems to students to practice for standardized tests. Another teacher looked for individual exercises the student got wrong in the CL tutor, to print and give it to the student to complete on paper.

# 5 Opportunities and Design Implications

From the Contextual Inquiry interviews and findings, we identified opportunities for technology, such as a teacher's dashboard, to address current breakdowns.

**Automate Processes the Teacher Does by Hand.** The detailed information on student mastery of concepts, performance and progress that teachers generate themselves can be provided by technology. This would save teachers time, effort and attention that can be used to help students in other ways.

**Adapt to Teacher Data Needs.** To be useful to the teacher, a new technology should provide data the teacher most needs in their instruction. This includes data that are difficult to generate by hand and that tutors or other software do not provide currently, but *could* provide, such as student misconceptions and growth over given periods of time, on the individual and class level.

**Help the Teacher Integrate Data from Different Sources.** Instead of the teacher having to remember and coordinate data they generate themselves from different assignments and data provided by tutors or other software, technology can help the teacher easily keep track of and manage this data.

**Suggest Materials for Intervention.** Teachers can receive suggestions from technology on materials and exercises to go over with students (individually or as a class), based on their performance with a topic. In addition, technology can create worksheets and assessments for the teacher by differentiating on the class or individual student performance. Technology should allow the teacher to access the problem or problems the student(s) got wrong and reassign it (or them) to the student(s).

**Provide Data on Hint Requests and Student Errors.** One teacher who used the CL tutor mentioned that they occasionally used the average hints and errors in the tutor reports to identify students who are goofing off or rushing through the problems, versus those who really need help. Hints and errors are important analytics that can help the teacher understand the performance of their students, and identify the need for intervention, while working with the tutor.

## 5.1 Towards the Design of a Teacher Dashboard for ITSs

In an ITS environment, where a lot of student data is produced by the system, a dashboard can provide the teacher with the necessary analytics and functionality to help them help their students learn better. Based on our findings of how teachers use data to drive instruction and help students on the class and individual student level, we have brainstormed and designed preliminary scenarios where a dashboard can be integrated in an ITS environment and help the teacher in this process.

**Teacher Dashboard for the Class Level.** Teachers could use this dashboard when preparing for the next lecture and deciding whether to move on to the next topic. In addition, the data provided by this dashboard would help the teachers identify the need for intervention by giving information on the class performance and progress in the ITS environment. The dashboard would help the teacher determine the focus of intervention, as well as suggest materials, such as example problems or practice worksheets for the class. Another scenario that teachers could use this dashboard for is when they quickly want to review where students' concept mastery stands, and whether a quick intervention or mini-lecture might be helpful. Teachers would use this dashboard when giving students a warm-up exercise at the beginning of class, or a short practice exercise at the end of a lecture. Lastly, the dashboard could provide teachers with real time data on students' performance during the time students are working with the ITS. Teachers would be able to project the dashboard on a wall or screen in class, and would better focus their time and attention on students who need it the most, while other students independently work with the tutor.

**Teacher Dashboard for the Individual or Group Level.** Teachers would use the information and analytics provided by this dashboard to give one-on-one attention and help to individual students or a group of students with similar issues and problems. The data provided by this dashboard would help the teacher identify the need for intervention, as well as the focus area(s), while providing the teacher with suggested practice problems.

## 6 Discussion and Future Work

A key assumption in our project is that a teacher dashboard will be more effective if it is designed with a deep understanding of how data about students' performance and learning can influence teacher decision-making. In this paper we investigate ways in which teachers generate and use data to drive and adjust their instruction. Through Contextual Inquiry interviews with 6 middle and high school teachers, we found that teachers use data to a surprising degree to inform their teaching, both to make decisions at the class level and to plan interactions with individual students. Further, the data they use (and often, generate themselves, by hand) can have a surprising amount of detail, as shown in Fig. 3. We also found that teachers use data provided by technology, when it is available. On the class level, teachers use this data to decide whether they need to spend more time on a certain topic and when to move to the next topic. In addition, teachers differentiate instruction across class periods focusing on each classes' specific needs and performance. Teachers who use technology in their classrooms make use of reports and analytics provided by the technology, again both on the class and individual student level. However, we also found that teachers have to adapt to technology and are selective in deciding which types of reports and data provided by such technology to use. An interesting finding is that teachers differentiate instruction on the individual student level. They spend time, effort and attention to identify what individual students need most help with, what issues they are having and how to help them remediate these issue(s).

Our findings provide novel insights into what data teachers generate and how they use it to help their students. To the best of our knowledge, this is the first study that investigates, through the use of Contextual Inquiry together with Interpretation

Sessions and Affinity Diagramming, how teachers use data in their day-to-day decision-making with or without technology. The findings may be useful for designers of dashboards and ITS more generally. Their import is not restricted to ITS, since the majority of teachers in the study did not use one with their students.

The next stage of our project is to use our results to inform the design of a teacher dashboard with student data collected from an ITS such as Lynnette [17, 28]. Focusing on specific use scenarios, the dashboard will take advantage of the rich analytics generated by the ITS, such as skill mastery, types of misconceptions, progress and time in the assignments, etc. Our findings will drive the decisions of what data is most important for the teacher in the given scenario and how it will be presented to the teacher in the dashboard in an easy-to-understand way. Continuing our user-centered design process, we will develop paper prototypes of the dashboard, which we will pilot and test with teachers. The ultimate product of our efforts will be a dashboard, fully integrated with CTAT/Tutorshop, our infrastructure for developing and deploying ITS [2]. Once it is fully implemented, we will conduct classroom studies to evaluate its effectiveness when used by teachers, in helping their students achieve better learning outcomes.

**Acknowledgments.** We thank Gail Kusbit, Carnegie Learning, Jae-Won Kim, and the teachers we interviewed for their help with this project. NSF Award #1530726 supported this work.

## References

1. Abel, T.D., Evans, M.A.: Cross-disciplinary participatory & contextual design research: creating a teacher dashboard application. Interact. Des. Archit. J. **19**, 63–76 (2013)
2. Aleven, V., McLaren, B.M., Sewall, J., van Velsen, M., Popsecu, O., Demi, S., Ringenberg, M., Koedinger, K.R.: Example-tracing tutors: Intelligent tutor development for non-programmers. Int. J. Artif. Intell. Educ. **26**, 224–269 (2016)
3. Ali, L., Hatala, M., Gasevic, D., Jovanovic, J.: A qualitative evaluation of evolution of a learning analytics tool. Comput. Educ. **58**(1), 470–489 (2012)
4. van Alphen, E., Bakker, S.: Lernanto: an ambient display to support differentiated instruction. In: Proceedings of the CSCL 2015 Conference on Computer Supported Collaborative Learning, pp. 759–760 (2015)
5. Anderson, J.R., Corbett, A.T., Koedinger, K.R., Pelletier, R.: Cognitive tutors: lessons learned. J. Learn. Sci. **4**(2), 167–207 (1995)
6. Arnold, K.E., Pistilli, M.D.: Course signals at Purdue: using learning analytics to increase student success. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 267–270. ACM (2012)
7. Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gasevic, D., Mulder, R., Williams, D., Dawson, S., Lockyer, L.: A conceptual framework linking learning design with learning analytics. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, pp. 329–338 (2016)
8. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann, San Francisco (1997)

9. Bull, S., Kay, J.: Open learner models. In: Nkambou, R., Bourdeau, J., Mizoguchi, R. (eds.) Advances in Intelligent Tutoring Systems. SCI, vol. 308, pp. 301–322. Springer, Heidelberg (2010)
10. Carnegie Learning. https://www.carnegielearning.com/
11. Hamilton, L., Halverson, R., Jackson, S.S., Mandinach, E., Supovitz, J.A., Wayman, J.C.: Using student achievement data to support instructional decision making, IES Practice Guide, NCEE 2009-4067, National Center for Education Evaluation and Regional Assistance (2009)
12. Heffernan, N.T., Heffernan, C.L.: The ASSISTments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. Int. J. Artif. Intell. Educ. **24**(4), 470–497 (2014)
13. Kamin, S., Capitanu, B., Twidale, M., Peiper, C.: A teacher's dashboard for a high school algebra class. In: Reed, R.H., Berque, D.A., Prey, J.C. (eds.) The Impact of Tablet PCs and Pen-based Technology on Education: Evidence and Outcomes, pp. 63–72. Purdue University Press, West Lafayette (2008)
14. Kelly, K., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., Soffer Goldstein, D.: Estimating the effect of web-based homework. In: Lane, H., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS, vol. 7926, pp. 824–827. Springer, Heidelberg (2013)
15. Kulik, C.-L.C., Kulik, J.A., Bangert-Drowns, R.L.: Effectiveness of mastery learning programs: a meta-analysis. Rev. Educ. Res. **60**(2), 265–299 (1990)
16. Kulik, J.A., Fletcher, J.D.: Effectiveness of intelligent tutoring systems: a meta-analytic review. Rev. Educ. Res. **86**(1), 42–78 (2015)
17. Long, Y., Aleven, V.: Mastery-oriented shared student/system control over problem selection in a linear equation tutor. In: Micarelli, A., Stamper, J., Panourgia, K., Krouwel, M. R. (eds.) ITS 2016. LNCS, vol. 9684, pp. 90–100. Springer, Heidelberg (2016). doi:10.1007/978-3-319-39583-8_9
18. Lovett, M., Meyer, O., Thille, C.: The open learning initiative: measuring the effectiveness of the OLI statistics course in accelerating student learning. J. Interact. Media Educ. **2008**(1), Art. 13 (2008)
19. Ma, W., Adesope, O.O., Nesbit, J.C., Liu, Q.: Intelligent tutoring systems and learning outcomes: a meta-analysis. J. Educ. Psychol. **106**(4), 901–918 (2014)
20. Maldonado, R.M., Kay, J., Yacef, K., Schwendimann, B.: An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 482–492. Springer, Heidelberg (2012)
21. Mazza, R., Dimitrova, V.: CourseVis: a graphical student monitoring tool for supporting instructors in web-based distance courses. Int. J. Hum. Comput. Stud. **65**(2), 125–139 (2007)
22. McLaren, B.M., Scheuer, O., Miksatko, J.: Supporting collaborative learning and e-discussions using artificial intelligence techniques. Int. J. Artif. Intell. Educ. **20**(1), 1–46 (2010)
23. Steenbergen-Hu, S., Cooper, H.: A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. J. Educ. Psychol. **105**(4), 970–987 (2013)
24. Steenbergen-Hu, S., Cooper, H.: A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. J. Educ. Psychol. **106**(2), 331–347 (2014)
25. van Leeuwen, A., Janssen, J., Erkens, G., Brekelmans, M.: Supporting teachers in guiding collaborating students: effects of learning analytics in CSCL. Comput. Educ. **79**, 28–39 (2014)
26. VanLehn, K.: The behavior of tutoring systems. Int. J. Artif. Intell. Educ. **16**(3), 227–265 (2006)

27. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educ. Psychol. **46**(4), 197–221 (2011)
28. Waalkens, M., Aleven, V., Taatgen, N.: Does supporting multiple student strategies lead to greater learning and motivation? Investigating a source of complexity in the architecture of intelligent tutoring systems. Comput. Educ. **60**(1), 159–171 (2013)
29. Woolf, B.P.: Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing e-Learning. Morgan Kauffman, Burlington (2010)

# Short Papers

# Assessing Learner-Constructed Conceptual Models and Simulations of Dynamic Systems

Bert Bredeweg[1(✉)], Jochem Liem[1], and Christiana Nicolaou[2]

[1] Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
B.Bredeweg@uva.nl, Jochem.Liem@gmail.com

[2] Department of Educational Studies, University of Cyprus, Nicosia, Cyprus
chr.nic@ucy.ac.cy

**Abstract.** Learning by conceptual modeling is seeing uptake in secondary and higher education. However, assessment of conceptual models is underdeveloped. This paper proposes an assessment method for conceptual models. The method is based on a metric that includes 36 types of issues that diminish model features. The approach was applied by educators and positively evaluated. It was considered useful and the derived grades corresponded with their intuitions about the models quality.

**Keywords:** Assessment · Conceptual modeling and simulation

## 1 Introduction

Acquiring knowledge by constructing and using models is seeing uptake in secondary and higher education [2]. Recently, the approach is applied in a novel way using *conceptual models* and accompanying tools, which allow modelers to develop and simulate conceptual representations of dynamic systems [1].

To implement modeling in classroom practice, formative and summative assessment techniques [3] for evaluating learner-constructed models are indispensable [5]. Assessment is one of the four vital parameters for science education, together with curriculum, instruction and professional development. However, the assessment of conceptual models is underdeveloped, hampering its usage. This means that there is a lack of criteria of what constitutes a good conceptual model. Consequently, it is difficult to give feedback to learners regarding the models they create. The problem is even more pressing when learning is self-regulated, and (groups of) learners develop their own unique models with different viewpoints, conceptualisations, and levels of abstraction. Comparison between learner-constructed models, and even comparison with a norm model, becomes impractical and inadequate for assessment.

This paper focusses on how assessment of *conceptual* models can be performed. The central idea is that learner-constructed conceptual models are rich representations, and as such provide evidence of learning. Particularly, the number of correctly modeled ingredients compared to the total number of model

Research co-funded by EU FP7, Project no. 231526, http://www.dynalearn.eu.



K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 357–362, 2016.
DOI: 10.1007/978-3-319-45153-4_27

ingredients (determined through a catalogue of modeling suboptimalities) can function as a measure of the modeling competence of the learner. This evidence can be identified, enumerated and scored by an assessment method and as such provide the basis for feedback, both formative and summative, and for learners and teachers. Hence, the question guiding the presented research is: What are the main components of an assessment method which can successfully evaluate diverse and different learner-constructed conceptual models?

## 2 Learner-Constructed Models

Consider the learner-constructed model shown in Fig. 1. It was created during a course on environmental science using DynaLearn (for details see [4], Chaps. 3 and 4). Learners were asked to choose a system based on their interest, pose a question about that system and develop a model that answers this question. There were no norm models. The only constraint was that at least two processes causing change in the system were modeled. The learners worked in pairs. Model assessment in the context of such a self-regulated learning activity is quite a challenge.

Let us start by interpreting the domain details shown in the diagram. The model represents a field of quinoa being irrigated using salt water. The amount of water absorbed by the quinoa is determined by the concentration of salt in the roots of the quinoa and the salinity of the earth near the roots of the quinoa. As water is absorbed, the quinoa grows and the yield increases.

There are no major issues with the representation of the physical structure of the system, although *Seeds* (and *Saponin*) can be considered superfluous. Quantities, on the other hand, can be improved. *Volume* of *Salt water* is positively influencing *Soil saturation*. However, causal dependencies of type $I-$ or $I+$ are used for processes, while in the model the dependency seems to be a proportionality, that is $P-$ or $P+$. Hence, this can be considered an incorrect causal relation (issue #20[1]) in the model. However, the model makes more sense if the volume quantity is interpreted as the irrigation process. Therefore, this issue is considered to affect the correctness of the model (validation, tagged $A$).

Quantity *Root zone salinity* refers to a mixture of notions including an entity and a quantity. As a result, it can be conceptually decomposed (issue #9). The simplest solution is to rename the quantity *Salinity*. Similarly, *Root salt concentration* can be conceptually decomposed (issue #9). The details in the model representing the physical structure of the system can be augmented by explicitly modeling the roots of the quinoa and indicating that these roots contain salt. This salt entity should have a quantity concentration.

The quantity spaces of *Root salt concentration* and *Root zone salinity* can be improved. There is no clear distinct behavior associated with reaching the landmark *Boundary* (issue #14). Consequently, this value and the value *Higher* can be removed. Secondly, the value *Higher* is vague (issue #15). That is, it is

[1] Our method identifies 36 issue types, each with a unique number.

context dependent (higher compared to what?). Renaming this value to whatever happens above the value *Boundary*, or removing the value, would resolve it.

Causality has 2 issues. First, quantity *Root zone salinity* is affected by both a positive influence (from *Water uptake*) and a positive proportionality (from *Soil saturation*). Mixing causality types is incorrect (issue #23). Either a quantity is affected by a process directly, or change propagates to this quantity. In this case the proportionality should be removed. Second, when there is no more water in the soil, there can be no more water uptake (which is modeled using a value correspondence between the magnitudes *Zero* of *Water update* and *Zero* of *Soil saturation*). However, for this to occur, *Water uptake* should decrease as *Soil saturation* decreases. This can be modeled using a positive proportionality from *Soil saturation* to *Water uptake*. This is missing in the model (issue #21).

There are 4 issues with inequalities and correspondences, all resulting in inconsistencies (issue #24) when simulating: value correspondence from *Volume* of *Salt water* to *Soil saturation*, from *Volume* of *Salt water* (derivative) to *Soil saturation* (derivative), and the two correspondences from *Water uptake* to *Growth*.

Finally, simulation has 2 issues (Fig. 2). First, quantity *Soil saturation* has no value (issue #32). Second, quantity *Root salt concentration* has the value *Plus* and is decreasing in state 3, but never reaches *Zero*. This is a so-called *dead-end* (issue #34), caused by an inconsistency.

## 3 Instrument for Assessing Conceptual Models

We propose to use *model features* that attest to the quality of a model (Table 1). These features are categorized into two verification categories. First, *formalism features* apply only to conceptual models developed in formalisms that allow for inferences, such as DynaLearn. These features can be assessed using the internal logic of the formalism (e.g., consistency). The second category, *domain features* apply to conceptual models generally, and rely on the human interpretation of the model to be assessed. For example, the model feature *conformance to ontological commitments* requires that a referent in the domain is represented using the correct model ingredient in the formalism (e.g., biomass should be represented as a quantity). We claim both features can be checked objectively. Algorithms can be created to automatically detect them and suggest corrections.

Next step is to determine which model characteristics can be used to actually measure the quality of a conceptual model in terms of formalism and domain features. Correctness, completeness, and parsimony can be used as such quality characteristics. *Correctness* indicates that a model is free from errors. *Completeness* means that everything of relevance is included in the model. *Parsimony* implies that the model does not include redundancies. The following sections identify model features that attest to these quality characteristics.

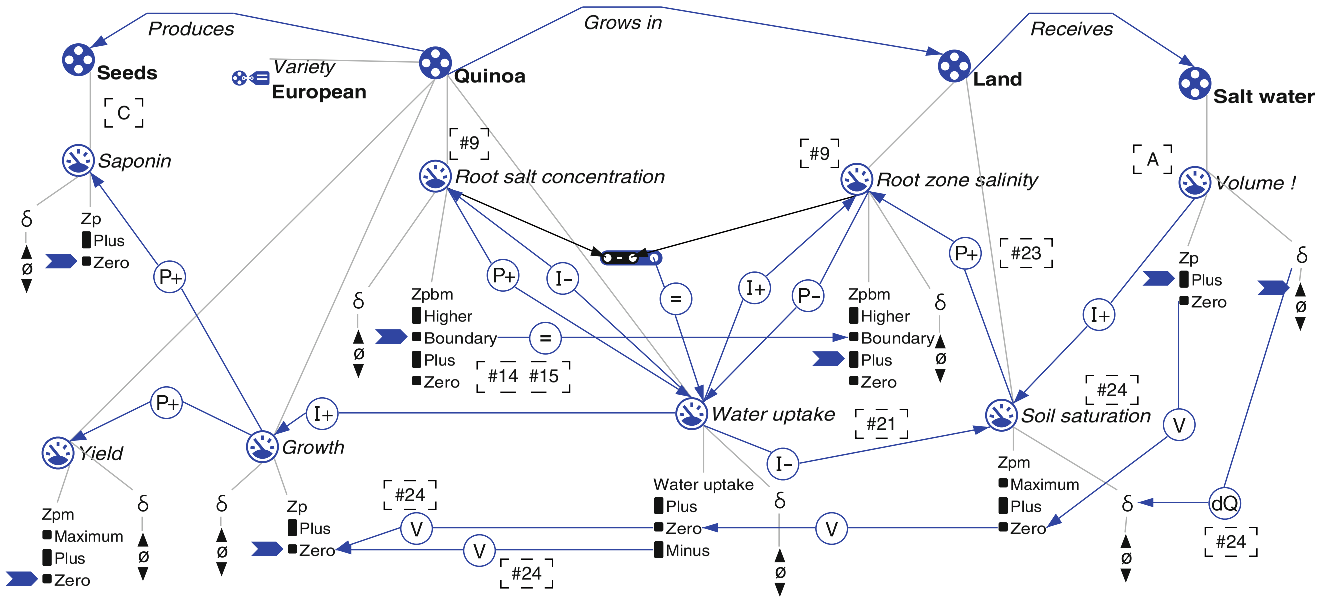



**Fig. 1.** Learner-constructed DynaLearn [1] model, using learning space 4, modeling the effects of watering quinoa using salt water. The amount of water absorbed (*Water uptake*, $WU$) by the quinoa is determined by the concentration of salt in the roots (*Root salt concentration*, $RSC$) of the quinoa and the salinity (*Root zone salinity*, $RZS$) of the earth near the roots of the quinoa ($RSC - RZS = WU$). As water is absorbed, the quinoa grows (*Water uptake* $I+$ *Growth*) and the yield increases (*Growth* $P+$ *Yield*). Particularly well-modeled is the so-called *equilibrium seeking mechanism* that determines the uptake of water, which consists of two negative feedback loops. The water uptake (if *Water uptake* $=$ *Plus*) decreases the salt concentration in the roots of the quinoa ($I-$), and increases the salinity of the soil surrounding the roots ($I+$). The water uptake decreases as the salt concentration in the root decreases ($P+$), and the water uptake also decreases as the salinity of the soil surrounding the roots increases ($P-$). Note, the layout has been changed by the authors. The model issue numbers (verification) and the validation issues (A: correctness, C: parsimony) are indicated in dashed boxes.

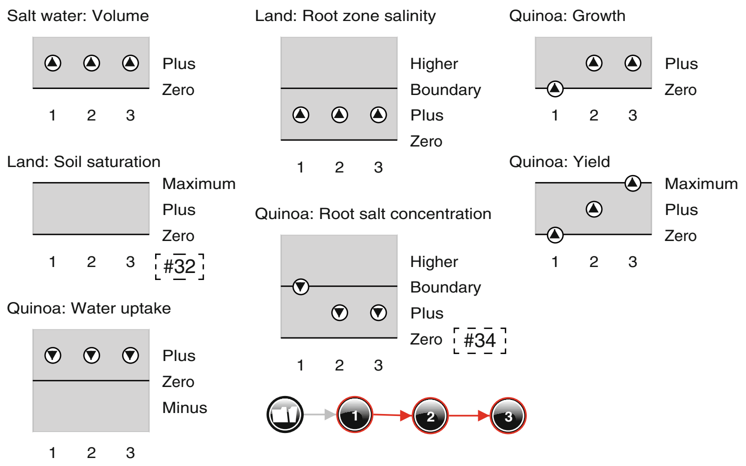



**Fig. 2.** The state-graph (4 connected circles) and value history (7 squares) of the quinoa model (Fig. 1). The model issues (#32 and #34) are indicated in dashed boxes.

**Table 1.** Model features that attest to quality characteristics of verification categories.

| Verification category | Quality characteristic | Model feature |
|---|---|---|
| Formalism | Correctness | Consistency |
| | Completeness | No unassigned variables |
| | Parsimony | Reasoning relevance |
| Domain representation | Correctness | Conformance to ontological commitments |
| | | Falsifiability |
| | Completeness | Conceptual decomposition |
| | | No missing representations |
| | Parsimony | No repetition |
| | | No synonyms |

## 4 Evaluating the Assessment Method

A pilot study was conducted with four evaluators who used the instrument to grade 34 models submitted by the student pairs in the course (two evaluators graded 9 models). The pilot focussed on whether the grades derived using the assessment method are comparable to grades that evaluators proclaim a model deserves. To this end, before having graded any models, the evaluators were asked to *intuitively* grade one set of models assigned to another evaluator (one assistant graded 2 sets). The instruction was to analyse each model for 5 min, write down the grade, and proceed to the next model.

The *agreement* between the intuitive and actual grades was calculated. For this the different evaluators are assumed equal, and therefore all assessment method grades are considered of one evaluator (34 grades), and all intuitive grades of another (34 + 10 = 44 grades) (data available via [4] Chap. 5, p. 140).

Typical statistical methods for inter-rater agreement (Cohen's kappa and Fleiss' kappa) cannot be used as they require a fixed number of mutually exclusive categories. IntraClass Correlation (ICC) and the Concordance Correlation Coefficient (CCC) can be used. Both were calculated, and both indicate strong agreement of about 0.89 ($r^{ICC} = 0.887$, 99 %-confidence interval: $0.765 < r^{ICC} < 0.947$, $r^{CCC} = 0.885$, 99 %-confidence interval: $0.767 < r^{CCC} < 0.945$). Suggesting the method's grades are acceptable.

Evaluators were able to detect model issues easily and only had difficulty in understanding one issue (#9. Ambiguous process rate *quantities*). This suggests that the assessment method is understandable and usable for evaluators. The evaluators required about 45 min per model to derive grades. As the model contributed 40 % of the final grade, 45 min was considered reasonable.

## 5 Conclusion and Discussion

We propose an assessment instrument based on a set of model features that attest to the quality of *conceptual* models. The model features address verification, and are categorized as formalism and domain features. The former apply only to conceptual models that allow for inferences, while the latter apply generally. The model features are further categorized as attesting to the quality characteristics correctness, completeness and parsimony. A pilot study using the assessment method suggests that the derived grades correspond to evaluators' intuition of what a model is worth. The assessment method proved understandable, and the time required to apply it is considered reasonable. A listing of all the issues in a model serves as both an argument why a particular grade was given and as valuable feedback for learners. As ongoing research we are investigating how the presented approach can be used as a real-time operating instrument, particularly for formative assessment, which requires automated detection of modeling issues.

## References


1. Bredeweg, B., Liem, J., Beek, W., Linnebank, F., Gracia, J., Lozano, E., Wißner, M., Bühling, R., Salles, P., Noble, R., Zitek, A., Borisova, P., Mioduser, D.: DynaLearn - an intelligent learning environment for learning conceptual knowledge. AI Mag. **34**(4), 46–65 (2013)
2. Goberta, J.D., O'Dwyer, L., Horwitz, P., Buckley, B.C., Levy, S.T., Wilensky, U.: Examining the relationship between students' understanding of the nature of models and conceptual learning in biology, physics, and chemistry. Int. J. Sci. Educ. **33**(5), 653–684 (2011)
3. Harlen, W., James, M.: Assessment and learning: differences and relationships between formative and summative assessment. Assess. Educ. Principles Policy Pract. **4**(3), 365–379 (1997)
4. Liem, J.: Supporting conceptual modelling of dynamic systems: a knowledge engineering perspective on qualitative reasoning, University of Amsterdam (2013). https://jochemliem.files.wordpress.com/2014/01/liem2013-thesisdigital.pdf
5. Songer, N.B., Ruiz-Primo, M.A.: Assessment and science education: our essential new priority? J. Res. Sci. Teach. **49**(6), 683–690 (2012)

# Learning Analytics Pilot with Coach2 - Searching for Effective Mirroring

Natasa Brouwer[1], Bert Bredeweg[2(✉)], Sander Latour[4], Alan Berg[3], and Gerben van der Huizen[2]

[1] Education Service Center, University of Amsterdam, Amsterdam, The Netherlands
[2] Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
[3] ICT-Services, University of Amsterdam, Amsterdam, The Netherlands
{N.Brouwer-Zupancic,B.Bredeweg,A.M.Berg}@uva.nl
[4] Perceptum B.V., Amsterdam, The Netherlands
Sander@perceptum.nl

**Abstract.** Coach2 project investigated usability and effectiveness of Learning Analytics in a group of Bachelor courses in the area of Computer Science. An advanced architecture was developed and implemented, including a standalone Learning Record Store for data storage and easy access to miscellaneous data, Machine Learning techniques for determining relevant predictors, and a dashboard for informing learners. The overall approach was based on mirroring, the idea that learners see themselves operating in the context of their peers. The results were informative in terms of pro's and con's regarding the design and approach. The treatment showed tendencies, but finding statistical significant results turned out difficult. This paper reports on the Coach2 project.

**Keywords:** Learning Analytics · Usability and effectiveness · Higher education · Mirroring

## 1 Introduction

Learning analytics concerns the process by which data generated by learners during learning activities is used to inform and advice learners about their behaviour with the goal to help them improve their learning and achieve better learning outcomes. Initial results on the potential of Learning Analytics have been reported [2,4,5,7], but it is also evident that Learning Analytics is still a challenge and in search of the appropriate procedures and techniques (cf. [6]).

As many higher education institutions, the University of Amsterdam (UvA) is interested in understanding and using Learning Analytics. Within that context the Coach2 project was formulated [1]. The overall goal of the project was to investigate the *usability* and *effectiveness* of Learning Analytics as an instrument to improve learning within the context of typical and regular ongoing courses. Additional foci included the wish to use only data generated within an actual course, and that the feedback towards learners should focus on mirroring (the


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 363–369, 2016.
DOI: 10.1007/978-3-319-45153-4_28

idea to show a learner's specific behaviour in the context of the behaviour of his or her peers). It was also deemed important to stay within the scope of the tools and Learning Management System (LMS) currently used during these courses, and learn about the potential and limitation thereof. Hence, a strong emphasis on working with data available from using Blackboard (the dominant LMS at the UvA), and the need to work with the technical infrastructure regarding tooling and educational software as currently deployed.

Figure 1 depicts the idea of the approach taken. Learners use educational tools and by doing so they generate data. This data is obtained and stored. Next, this data is processed, particularly using machine learning techniques to discover correlations and potentially predictors of successful learner behaviour. Finally, learning behaviour data are displayed in an informative way to the learner.

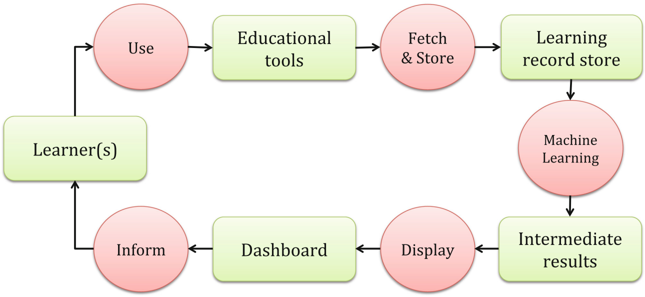



**Fig. 1.** Data containers and their high level activities implementing Learner Analytics.

For the technical realisation and the evaluation studies, three courses were selected from a bachelor programme on computer science. Each course had over 80 participants, and used a variety of tools and educational activities. Two of the courses worked with Blackboard; the third one did not. Using an informed consent, the learners were given the choice to participate in the evaluation study or decline. Next, the participating learners were randomly divided into two conditions, one with and one without a dashboard. In other words, a group with and a group without Learning Analytics. The teachers of the courses were informed about the study, and agreed to have their learners participate. However, the teachers were left outside the evaluation study. Hence, keeping them ignorant and thereby preventing unwanted effects because of their potential interferences.

## 2 Architecture and Technical Context

The Coach2 pilot architecture is shown in Fig. 2. The Coach2 pilot used a central Learning Record Store (LRS) as the secure web enabled location to capture and query the learners digital traces.[1] The protocol applied was xAPI. The LRS

[1] https://github.com/Apereo-Learning-Analytics-Initiative/Larissa.

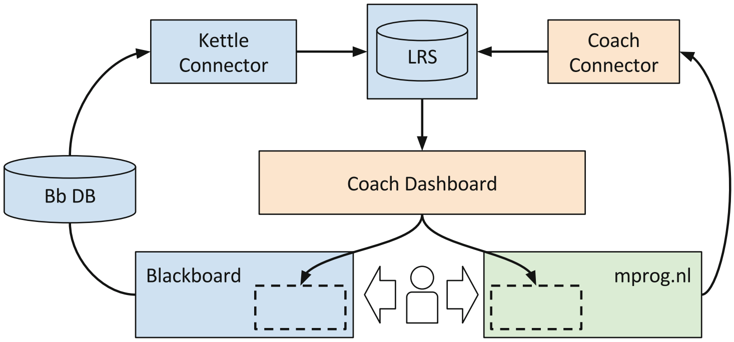



**Fig. 2.** Blue boxes denote institution-wide infrastructure, including the Learning Record Store, the Blackboard learning environment and its database, and the Kettle connector that exports Blackboard data to the LRS. Yellow boxes denote pilot-wide infrastructure maintained by the Coach2 project to use in the evaluation studies, including the Coach dashboard that was integrated into the learning environments and a Coach connector that provided an API for external sources to send events to, which would be exported to the Learning Record Store. The green box denotes the website of one of the courses that was used as their learning environment instead of Blackboard. (Color figure online)

was implemented by UvA ICT-Services. The motivation for this was to build indigenous expertise to understand in great detail the inner workings of the approach. The LRS was stress tested by Jmeter[2] an open source Java application and found to scale to 4 million records on one virtual machine. For the pilot the scalability was acceptable, however, greater usage would require improving the internal mechanisms for responding to querying. One of the significant lessons learned was that some xAPI queries are more expensive in resources such as CPU time than others. One approach to limit the impact is to define a specific set of queries that are allowable thus avoiding unnecessary resource consumption.

Filling the LRS with data was achieved by added an Extract Transform Load layer, which allows to pull in data from various systems and then convert to xAPI statements and pump events into the LRS (for details see Github[3]).

## 3 Dashboard and Data Processing

The developed DashBoard (DB) is shown in Fig. 3. It was presented inside the LMS for each of the three courses. By selecting specific study behaviour values on the left side, the probability of the values of study outcome metrics are updated on the right side. The hypothesis was that the DB enables learners to explore and reflect upon statistical relations between current study behaviour and future result, based on experiences of learners in the past. By visualising

[2] https://github.com/Apereo-Learning-Analytics-Initiative/LRSLoadTest.
[3] http://c-f-k.github.io/bb-kettle-lrs-tutorial/.

how the learner's study behaviour compares to that of peers, as well as whether that study behaviour correlates to study outcome in the past, it was expected that the DB provides an actionable tool for reflection.

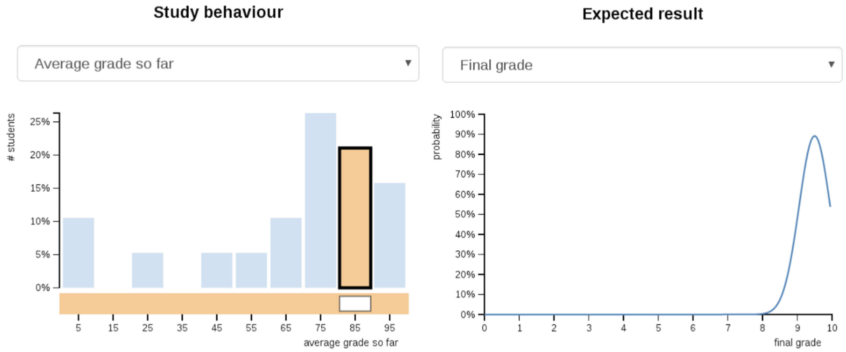



**Fig. 3.** Dashboard interface. The *barplot visualisation* (LHS) is used to visualize the values of a metric of study behaviour. The x-axis denotes the bin values and the y axis denotes the percentage of learners that have a value that falls into that bin for the specific metric. A bin can be selected (orange) by clicking on it, or by sliding the bar underneath the barplot to the desired bin. The bin in which the viewing learner is placed is selected at the beginning. The *bell curve visualisation* (RHS) is used to visualize the probability of each value of a metric of study result, given a selected value of a metric of study behaviour in the barplot. In other words, how likely it is that a certain end result is achieved based on the current state of behaviour. When a different bin is selected in the barplot, this curve is updated. The data is represented by mean and variance parameters. (Color figure online)

The study behaviour and expected results were approximated by quantitative metrics (Table 1), including (i) Input metrics (used to represent the current state of behaviour, e.g. time on task), (ii) Output metrics (represent the end result, e.g. exam grade), and (iii) In/Out metrics (can be used to represent either, e.g. running average grade). The DB provided insight in how certain values of an input metric related to certain values of an output metric. In the evaluation studies, depending on the course, a subset of the metrics (Table 1) were used.

When the dashboard was requested for viewing by a learner, he or she also selected an input metric to examine. The value history of that metric was transformed into the aggregated data necessary for the visualisation. For each aggregated value, the system calculated predicted aggregated values on each output metric. The metric's value history items were filtered to only contain data from learners of the same (and current) cohort of the viewing learner.

The data was divided into equally sized bins, with a fixed number of bins defined for the metric. The bins span from the lowest to the highest aggregated value. These steps resulted in an array of frequency bins where each bin denoted a range of metric values (e.g. average grade) and its frequency denoted the number of learners for which the metric value fell into that bin.

**Table 1.** Available metrics

| Metric | Type | Based on data |
|---|---|---|
| Average grade | In/Out | Intermediate grades from assignments and/or tests |
| Final grade | Output | Final course grade, based on exams and coursework |
| Pass rate | Output | Final course grade and the passing threshold |
| Blackboard activity | Input | The number of individual clicks in blackboard |
| Attendance | Input | Who was present at which lecture |
| Time spent on video | Input | Number of seconds spent playing an instructional video |
| Time spent on course site | Input | Estimate number of seconds in non-idle state on the site. This is based on time between user actions and tab focus |
| Time submitted before deadline | Input | Number of seconds between the submission time and the assignments deadline |
| Time before first attempt | Input | Number of seconds between the availability of the programming assignment and the first compilation of an attempt |

## 4 Evaluation Study

For each of the three courses data for the relevant variables (Table 1) were collected, and the following issues evaluated:

– Impact of the DB on the performance of learners, i.e. impact of the DB on the obtained grades. Evaluate if the percentage of successful learners was higher in the group which utilized the DB.
– Predictive value of the first achieved grades of each learner with regard to their entire performance during a course.
– The predictive value of cumulative grades of each learner obtained during a course (to predict learner performance at exams and of the entire course).
– Time spent on the LMS of a course, click behaviour, hand-in time of assignments before a deadline, watch time of videos and website paths were evaluated (if applicable) based on their correlation with the academic success/performance of learners during a course.

Correlations in data were found using the WEKA[4] visualization and correlation matrix functionality. The most notable (and surprising) result was that on average the learners in a DB condition had a (statistically significant) higher chance

[4] https://weka.wikispaces.com.

of successfully graduating for a course (79 % of the learners with a DB passed and 67 % of the learners without DB passed).

For one of the courses, more than half of the learners with no DB failed to pass the course, while 68 % of the learners who used the DB passed the course. On average, learners which on average scored low (below 4.5) for the first assignment of a course, also had a higher chance at scoring a low grade for the entire course. Learners with a high or average grade for their first assignment had similar results for their end grade as well. However, only for one course this result was statistically significant. The cumulative grades showed high potential as predictors as well, but this was dependent on the course and the amount of grades taken into consideration.

Time spent on the course LMS, click behaviour, hand-in time of assignments before a deadline, watch time of videos all had low correlations with the performance of learners (correlations were evaluated for each course if available).

## 5 Conclusion and Discussion

We have implemented and evaluated a Learning Analytics instrument, within the context of three bachelor courses in higher education. The instrument has the technical potential to scale and be applied to a much larger set of courses. However, the impact it has on learners and their behaviour is still unclear. The obtained results are preliminary, and further analysis is required.

On average, the learners in groups with DB seem to have better overall performance compared to the learners in groups without DB. However, it is unclear why and which aspects of the DB caused this influence on the performance of the learners, or if there were other confounding factors. The grade for the first assignment in each course can be considered relevant for predicting the performance of each learner during the entire course. Moreover, as further information of this sort accumulates (data related to cognitive behaviour), the predicted power quickly increases. Finally, there seem to be no significant correlations between learner activity in the LMS (e.g. click behaviour in Blackboard) and the expected output performance.

In further research we plan to include personal characteristics and motivation using the MSLQ questionnaire [3], as well as demographic data, and investigate how these can help to increase the accuracy of the prediction power of our Learning Analytics instrument and become a reliable and relevant tool for learners.

## References

1. Bredeweg, B., et al.: Coach2 project, UvAInform Learninganalytics. University of Amsterdam teaching innovation call (2015). http://starfish.innovatievooronderwijs.nl/project/603/
2. Chatti, M.A., et al.: A reference model for learning analytics. Int. J. Technol. Enhanced Learn. **4**(5/6), 318–331 (2012)

3. Chatti, M.A., et al.: Learner modeling in academic networks. In: Proceedings - IEEE 14th International Conference on Advanced Learning Technologies, ICALT 2014, pp. 117–121. Institute of Electrical and Electronics Engineers Inc. (2014)
4. Duval, E.: Attention please! Learning analytics for visualization and recommendation. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, LAK 11, p. 917. ACM, New York (2011). doi:10.1145/2090116.2090118
5. Park, Y., et al.: Development of the learning analytics dashboard to support student's learning performance. J. UCS **21**(1), 110–133 (2015)
6. Tempelaar, D.T., et al.: In search for the most informative data for feedback generation: learning analytics in a data-rich context. Comput. Hum. Behav. **47**, 157–167 (2015)
7. Verbert, K., et al.: Learning Analytics Dashboard Applications. Am. Behav. Sci. (2013). http://abs.sagepub.com/content/early/2013/02/27/0002764213479363.abstract

# Predicting Academic Performance Based on Students' Blog and Microblog Posts

Mihai Dascalu[1(✉)], Elvira Popescu[2], Alexandru Becheru[2], Scott Crossley[3], and Stefan Trausan-Matu[1]

[1] Faculty of Automatic Control and Computers, University "Politehnica" of Bucharest, 313 Splaiul Independenței, 60042 Bucharest, Romania
{mihai.dascalu,stefan.trausan}@cs.pub.ro

[2] Faculty of Automation, Computers and Electronics, University of Craiova, 107 Bvd. Decebal, Craiova, Romania
popescu_elvira@software.ucv.ro, becheru@gmail.com

[3] Department of Applied Linguistics/ESL, Georgia State University, 34 Peachtree St. Suite 1200, Atlanta, GA 30303, USA
scrossley@gsu.edu

**Abstract.** This study investigates the degree to which textual complexity indices applied on students' online contributions, corroborated with a longitudinal analysis performed on their weekly posts, predict academic performance. The source of student writing consists of blog and microblog posts, created in the context of a project-based learning scenario run on our eMUSE platform. Data is collected from six student cohorts, from six consecutive installments of the Web Applications Design course, comprising of 343 students. A significant model was obtained by relying on the textual complexity and longitudinal analysis indices, applied on the English contributions of 148 students that were actively involved in the undertaken projects.

**Keywords:** Social media · Textual complexity assessment · Longitudinal analysis · Academic performance

## 1 Introduction

Automated prediction of student performance in technology enhanced learning settings is a popular, yet complex research issue [1, 2]. The popularity comes from the value of the predictive information which can be used for advising the instructor about students at-risk, who are in need of more assistance [3]. More generally, automated methods offer instructors the ability to monitor learning progress and provide personalized feedback and interventions to students in any performance state [4]. In addition, individualized strategies for improving participation may also be suggested [3]. Furthermore, a formative assessment tool could be envisaged based on the automatic prediction mechanism [3], which has the potential to decrease instructors' assessment loads [4]. Finally, students' awareness can be increased by providing them prediction results and personalized feedback [4].


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 370–376, 2016.
DOI: 10.1007/978-3-319-45153-4_29

Performance prediction has been extensively studied in web-based educational systems and, in particular, in Learning Management Systems (LMS). This is due to the availability of large amounts of student behavioral data, automatically logged by these systems, such as: visits and session times, accessed resources, assessment results, online activity and involvement in chats and forums, etc. [2]. Thus, student performance prediction models based on Moodle log data have been proposed in multiple previous studies [5–7]. Additionally, log data from intelligent tutoring systems (ITS) have also been used for performance prediction [8]. In contrast, students' engagement with social media tools in emerging social learning environments has been less investigated as a potential performance predictor [9].

The current paper aims at analyzing students' contributions on social media tools (i.e., posts on blogs and Twitter) as potential predictors of academic performance. The context of the study is a collaborative project-based learning (PBL) scenario, in which students' communication and collaboration activities are supported by social media tools. Instead of relying only on quantitative usage data, similar to most previous studies, we explore the actual content of students' contributions by applying textual complexity analysis techniques. More specifically, we investigate how students' writing style in social media environments can be used to predict their academic performance. Multiple textual complexity indices (ranging from lexical, syntactical to semantic analyses [10, 11]) are used to create an in-depth perspective of students' writing style. We corroborate these findings with a longitudinal analysis performed on learners' weekly blog and microblog posts in order to obtain a more comprehensive view of academic performance prediction. The scale of our study is quite large, unfolding over the course of six years, as data is collected from six consecutive installments of the Web Applications Design (WAD) course comprising of 343 students. A preliminary study based on only one student cohort yielded encouraging results [12]; this paper is an extension of the pilot study, enriched also with longitudinal analysis of students' contributions.

Details about the study settings are presented in the following section, together with the data collection and preprocessing steps, as well as employed automated methods (textual complexity and longitudinal analysis indices). The results of our in-depth analysis are reported in Sect. 3, while conclusions are outlined in Sect. 4.

## 2 Methods

### 2.1 Data Collection and Preprocessing

Data was collected over 6 consecutive winter semesters (2010/2011 – 2015/2016), with 4th year undergraduate students in Computer Science from the University of Craiova, Romania. A total of 343 students, enrolled in the WAD course, participated in this study. A PBL scenario was implemented, in which students collaborated in teams of around 4 peers in order to build a complex web application of their choice. Several social media tools (wiki, blog, microblogging tool) were integrated as support for students' communication and collaboration activities; all student actions on these social media tools were monitored and recorded by our eMUSE platform [13].

For the current study, the collected writing actions used to assess students' writing styles consisted of their tweets, together with blog posts and comments. The yearly distribution of students and of their social media contributions is presented in Table 1. We focused only on the content written in English. This content was cleaned of non-ASCII characters and spell-corrected. Finally, only students who had at least five English contributions after preprocessing and who used at least 50 content words were considered in order to meet the minimum content threshold needed for our textual complexity analysis. A content word is a dictionary word, not considered a stopword (common words with little meaning - e.g., "and", "the", "an"), which has as corresponding part-of-speech a noun, verb, adjective or adverb. Thus, a total of 148 students were included in our analysis, having cumulatively 3013 textual contributions.

**Table 1.** Distribution of students and contributions per academic year.

| | Year 1 (2010-2011) | Year 2 (2011-2012) | Year 3 (2012-2013) | Year 4 (2013-2014) | Year 5 (2014-2015) | Year 6 (2015-2016) |
|---|---|---|---|---|---|---|
| Number of students | 45 | 48 | 56 | 66 | 53 | 75 |
| Number of blog posts & comments | 166 | 121 | 318 | 1074 | 451 | 479 |
| Number of tweets | 326 | 181 | 1213 | 1561 | 956 | 1233 |

### 2.2 Textual Complexity Evaluation

In order to evaluate text complexity, we used the *ReaderBench* framework [10, 11] which integrates a multitude of indices ranging from classic readability formulas, surface indices, morphology and syntax, as well as semantics. In addition, *ReaderBench* focuses on text cohesion and discourse connectivity, and provides a more in-depth perspective of discourse structure based on Cohesion Network Analysis (CNA) [14]. CNA is used to model the semantic links between different text constituents in a multi-layered cohesion graph [15]. We refer readers to [10, 11] for further information about these features.

### 2.3 Longitudinal Analysis

We used one week as timeframe, due to the schedule of the academic semester in which students had one WAD class per week. The total length of the considered time series is 16 weeks, including 14 weeks of classes and 2 weeks for the winter holidays. For each student, the number of weekly blog and microblog posts was computed in order to obtain his/her time series of social media contributions. The performed longitudinal analysis relies on a wide range of evolution indices including average & standard deviation of contributions, entropy, uniformity, local extreme points, and average & standard deviation of recurrence. We refer readers to [16] for further information about these features that were initially used for keystroke analysis.

## 3 Results

We split the students into two equitable groups: high performance students with grades ≥ 8, while the rest were catalogued as low performance students. The indices from *ReaderBench* and from the longitudinal analysis that lacked normal distributions were discarded. Correlations were then calculated for the remaining indices to determine whether there was a statistical ($p < .05$) and meaningful relation (at least a small effect size, $r > .1$) between the selected indices and the dependent variable (the students' final score in the course). Indices that were highly collinear ($r \geq .900$) were flagged, and the index with the strongest correlation with course grade was retained, while the other indices were removed. The remaining indices were included as predictor variables in a stepwise multiple regression to explain the variance in the students' final scores in the WAD course, as well as predictors in a Discriminant Function Analysis [17] used to classify students based on their performance.

Medium to weak effects were found for *ReaderBench* indices related to word entropy, number of verbs, prepositions, adverbs, and pronouns, the number of unique words, number of named entities per sentence, and average cohesion between sentences and corresponding contributions measured with Latent Dirichlet Allocation [10] (see Table 2).

**Table 2.** Correlations between *ReaderBench* and longitudinal analysis indices, and course grade.

| Index | $r$ | $p$ |
|---|---|---|
| Word entropy | .416 | <.001 |
| Time series entropy | .378 | <.001 |
| Average verbs per sentence | .323 | <.001 |
| Average cohesion (LDA) between sentences and corresponding contribution | −.274 | <.010 |
| Average unique words per sentence | .270 | <.001 |
| Average prepositions per sentence | .264 | <.010 |
| Time series local extremes | .236 | <.010 |
| Average adverbs per sentence | .236 | <.010 |
| Average pronouns per sentence | .250 | <.010 |
| Average named entities per sentence | .189 | <.050 |

We conducted a stepwise regression analysis using the ten significant indices as the independent variables. This yielded a significant model, $F(3, 143) = 17.893$, $p < .001$, $r = .521$, $R^2 = .272$. Three variables were significant and positive predictors of course grades: word entropy, time series entropy and average verbs in sentence, denoting a higher activity and participation for high performance students. These variables explained 27 % of the variance in the students' final scores for the course.

The stepwise Discriminant Function Analysis (DFA) retained the same three variables as significant predictors of course performance (*Time series entropy* had the highest standardized canonical discriminant function coefficient), and removed the remaining variables as non-significant predictors. These three indices correctly allocated 108 of the 148 students from the filtered dataset, $\chi^2(\mathrm{df} = 3, n = 148) = 43.543$ $p < .001$, for an accuracy of 73.0 % (the chance level for this analysis is 50 %). For the leave-one-out cross-validation (LOOCV), the discriminant analysis allocated 105 of the 148 students for an accuracy of 70.9 % (see the confusion matrix reported in Table 3 for results). The measure of agreement between the actual student performance and that assigned by the model produced a weighted Cohen's Kappa of .457, demonstrating moderate agreement.

**Table 3.** Confusion matrix for DFA classifying students based on performance

| | | Predicted Performance Membership | | Total |
|---|---|---|---|---|
| | | Low | High | |
| Whole set | Low | 48 | 23 | 71 |
| | High | 17 | 60 | 77 |
| Cross-validated | Low | 48 | 23 | 71 |
| | High | 20 | 57 | 77 |

## 4 Conclusions

This paper investigated how students' writing style on social media tools, corroborated with the time evolution of their posts, can be used to predict their academic performance. Textual complexity and longitudinal analyses were performed on the blog and microblog posts of 148 (out of the total 343) students engaged in a project-based learning activity during 6 consecutive installments of the Web Applications Design course.

The analyses indicated that students who received higher grades in the course had greater word entropy, used more verbs, prepositions, adverbs, and pronouns, produced more unique words, and more named entities. Additionally, students who received higher grades had lower inner cohesion per contribution, indicating more elaborated paragraphs that represented a mixture of different ideas in the context of each contribution. The time series variables denote a more uniform distribution, with weekly fluctuations in terms of participation, which is normal for students that were more actively involved in using the social media tools. Three of these variables (word entropy, time series entropy and average verbs in sentence) were predictive of performance in both a regression analysis and a DFA.

The results are promising as several significant correlations and statistical models were identified in order to predict academic performance (i.e., course grades) based on textual complexity and longitudinal analysis indices. Additional experiments that will consider the learning style of each student, as well as an equivalent textual complexity model for Romanian language, are underway in order to augment the depth of our analyses. This will enable the consideration of a higher sample of students from the total of 343 course participants and will increase the power of the applied mechanisms.

**Acknowledgments.** This work was supported by the FP7 208-212578 LTfLL project, the 644187 EC H2020 RAGE project, and a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2604.

## References


1. Baker, R.S., Yacef, K.: The state of educational data mining in 2009: A review and future visions. J. Educ. Data Min. **1**(1), 3–17 (2009)
2. Romero, C., López, M.I., Luna, J.M., Ventura, S.: Predicting students' final performance from participation in on-line discussion forums. Comput. Educ. **68**, 458–472 (2013)
3. Yoo, J., Kim, J.: Can online discussion participation predict group project performance? investigating the roles of linguistic features and participation patterns. Int. J. Artif. Intell. Educ. **24**, 8–32 (2014)
4. Xing, W., Guo, R., Petakovic, E., Goggins, S.: Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. Comput. Hum. Behav. **47**, 168–181 (2015)
5. Calvo-Flores, M.D., Galindo, E.G., Jiménez, M.P., Piñeiro, O.P.: Predicting students' marks from Moodle logs using neural network models. Curr. Dev. Technol. Assist. Educ. **1**, 586–590 (2006)
6. Romero, C., Ventura, S., Espejo, P.G., Hervás, C.: Data mining algorithms to classify students. In: 1st International Conference on Educational Data Mining, pp. 8–17. Quebec, Canada (2008)
7. Zafra, A., Ventura, S.: Predicting student grades in learning management systems with multiple instance genetic programming. In: 2nd International Conference on Educational Data Mining, pp. 309–319. Cordoba, Spain (2009)
8. Pardos, Z.A., Heffernan, N.T., Anderson, B., Heffernan, C.L.: The effect of model granularity on student performance prediction using bayesian networks. In: Conati, C., McCoy, K., Paliouras, G. (eds.) UM 2007. LNCS (LNAI), vol. 4511, pp. 435–439. Springer, Heidelberg (2007)
9. Giovannella, C., Popescu, E., Scaccia, F.: A PCA study of student performance indicators in a Web 2.0-based learning environment. In: 13th IEEE International Conference on Advanced Learning Technologies (ICALT 2013), pp. 33–35. IEEE, Beijing, China (2013)
10. Dascalu, M.: Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence, vol. 534. Springer, Cham (2014)
11. Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., Nardy, A.: Mining texts, learner productions and strategies with Reader Bench. In: Peña-Ayala, A. (ed.) Educational Data Mining: Applications and Trends, pp. 335–377. Springer, Cham, Switzerland (2014)
12. Popescu, E., Dascalu, M., Becheru, A., Crossley, S.A., Trausan-Matu, S.: Predicting student performance and differences in learning styles based on textual complexity indices applied on blog and microblog posts – a preliminary study. In: 16th IEEE International Conference on Advanced Learning Technologies (ICALT 2016). IEEE, Austin, Texas (in press)
13. Popescu, E.: Providing collaborative learning support with social media in an integrated environment. World Wide Web **17**(2), 199–212 (2014)
14. Dascalu, M., Trausan-Matu, S., McNamara, D.S., Dessus, P.: ReaderBench – automated evaluation of collaboration based on cohesion and dialogism. Int. J. Comput. Support. Collaborative Learn. **10**(4), 395–423 (2015)
15. Trausan-Matu, S., Dascalu, M., Dessus, P.: Textual complexity and discourse structure in computer-supported collaborative learning. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 352–357. Springer, Heidelberg (2012)

16. Allen, L.K., Jacovina, M.E., Dascalu, M., Roscoe, R., Kent, K., Likens, A., McNamara, D.S.: {ENTER}ing the time series {SPACE}: uncovering the writing process through keystroke analyses. In: 9th International Conference on Educational Data Mining (EDM 2016). International Educational Data Mining Society, Raleigh, NC (in press)
17. Klecka, W.R.: Discriminant analysis. Quant. Appl. Soc. Sci. Ser, **19**. Sage Publications, Thousand Oaks, CA (1980)

# Take up My Tags: Exploring Benefits of Meaning Making in a Collaborative Learning Task at the Workplace

Sebastian Dennerlein[1(✉)], Paul Seitlinger[1], Elisabeth Lex[1], and Tobias Ley[2]

[1] Graz University of Technology, Graz, Austria
{sdennerlein,paul.seitlinger,elisabeth.lex}@tugraz.at
[2] Tallinn University, Tallinn, Estonia
tley@tlu.ee

**Abstract.** In the digital realm, meaning making is reflected in the reciprocal manipulation of mediating artefacts. We understand uptake, i.e. interaction with and understanding of others' artefact interpretations, as central mechanism and investigate its impact on individual and social learning at work. Results of our social tagging field study indicate that increased uptake of others' tags is related to a higher shared understanding of collaborators as well as narrower and more elaborative exploration in individual information search. We attribute the social and individual impact to accommodative processes in the high uptake condition.

**Keywords:** Collaborative learning · Meaning making · Uptake · Social tagging

## 1 Introduction

Leveraging social technologies at work enables professionals to collaboratively learn and solve ill-defined problems based on mediating artefacts [6] such as annotated resources: e.g. a team receives a challenging project, for which its members explore supplementary resources, upload them annotated with tags and description and engage in a reciprocal annotation process until the problem is understood and an appropriate solution is found. These mediating artefacts reflect the shared meaning negotiated in a collaborative knowledge building effort [9]. Digital negotiation requires combining each other's knowledge or expertise, reciprocally: i.e. taking up the socially shared meaning and building on top of it by manipulating the mediating artefact. This process leads to a composition of interrelated interpretations of meaning and enables two workers, small groups or whole organizations to achieve more than alone [7].

The underlying mechanism, called meaning making (MM), represents the essence of collaboration [8]. MM stresses the interactive and reciprocal nature of negotiation processes and the fact that meaning resides in the social realm. It can manifest itself in manifold ways in sociotechnical systems ranging from more explicit forms of negotiation such as collaborative writing to more implicit forms such as social tagging. Recent empirical studies in CSCL confirm that collaboratively building shared meaning is an inherent and inseparable part of individual learning. In studying a group of university students using a social tagging system (STS), [3] found, for example, that individual


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 377–383, 2016.
DOI: 10.1007/978-3-319-45153-4_30

learning is dependent on collective processes. Among groups, where agreement was reached more quickly about the use of tags, individuals also learned better. [1] discovered the dependency, as well, while studying navigation behaviour in a STS based on coevolution's internalization and externalization. In particular, they figured out that collective knowledge reflected in the strength of associations in a tag cloud takes effect on navigation and results in incidental learning in form of a change of the individual strength of associations in an internal test.

We, therefore, assume that engagement in MM also leads to an internally shared understanding of the collaborators, i.e. an alignment of their individual understanding [7]. Via those internalization and externalization processes, collaborators, artefacts and interpretations coevolve in a constant dynamic MM process: i.e. interpretations of collaborators become manifest in artefacts, which in turn shape their interpretations leading to a higher shared understanding of them and a more elaborated meaning. A central concept in MM is 'uptake', a term used for the interaction with others' interpretations in terms of understanding and doing something further with them [9]. High uptake indicates intensive engagement with the diverse accumulated meanings in a sociotechnical system and implies parallel social stimulation. This way, uptake suggests benefits for collaborative and individual learning: on social level (H1), uptake is expected to lead to a higher shared understanding of collaborators due to mutual stimulation; via this stimulation, uptake is expected to cue new ideas when exploring the Web, thereby, improving information search on individual level (H2).

Empirical studies (e.g. [3] & [1] reported above) have shown collaborative learning influencing individual learning convincingly. These studies, however, have not considered the extent of engagement with shared meaning and not explored effects on shared understanding. Besides, there is less evidence on benefits of MM in a workplace learning context, where learning is embedded into current work activities and typically happens in a self-regulated manner. Therefore, the purpose of the current paper is to explore effects of these uptake events on the individual and team in the working context. To test the hypotheses, we conducted a field study with a STS at the workplace allowing for uptake via the interaction with others' tags in a tag cloud.

## 2 Method

We carried out a social tagging study at the workplace lasting 4 weeks. Participants (N = 17) were recruited from Tallinn University, Graz University of Technology and Know-Center GmbH: 4 females and 13 males with an average age of 31.5 years (SD = 5.5) and computer (n = 11) or cognitive science (n = 6) background.

Professionals were asked to collaboratively explore web resources as basis for writing a state of the art for a project proposal about the topic *'Digital, Physical, and Socio-political Design Ideas to enhance the Exchange and Creation of Knowledge at Work.'* They were especially encouraged to explore different ideas (e.g. 'rotating desktop assignments') to shed light on the topic from different perspectives. They were also asked to consider others' contributions as cues to become aware of new perspectives. The task required to collect and tag 4 links or documents per week in a STS called

KnowBrain [2] and to explore other resource by means of a tag cloud. When adding resources to KnowBrain, participants were prompted to select themes (sub-topics derived from the exploration topic) from a multiple choice list to enable the thematic classification of the web queries before tagging them. The eight themes were 'Gamification & Playfulness', 'Inspiration Sources & Techniques', 'Collaboration Technologies', 'Personalization Services', 'Augmented Reality', 'Interior Design', 'Wellbeing & Health' and 'Socializing'.

We measured uptake by the extent to which a user reuses tags introduced by others, the 'social' tags. The number of clicked, unique social tags in the tag cloud, hence, defined the uptake rate. All activities in KnowBrain were recorded in log files. To assess the internal knowledge, we used association tests (AT; word fluency) [4] including the eight search themes as stimuli. To study benefits of uptake, a median split with respect to uptake was applied to differentiate between participants reusing more or less unique social tags in the tag cloud ($U_{high}$ vs. $U_{low}$ condition). For the exploration of benefits on social level, i.e. higher shared understanding (H1), the number of overlapping associations between the ATs was computed for both conditions. For the exploration of benefits on individual level, i.e. improved information search (H2), search was characterized by the number of explored re-sources and the rate at which users explored new themes during search (search costs).

## 3 Results

### 3.1 Social Level - Shared Understanding

H1 assumes higher shared understanding in terms of the intersection of associations in ATs for the $U_{high}$ than the $U_{low}$ condition. To exclude pre-existing differences between both conditions, we computed a comparison of means at $t_0$ obtaining no difference: $t(13) = -0.09$, *n.s.* To understand differences at $t_1$, a weighted graph was created, whereas the nodes correspond to the $n$ participants and a tie was created between two nodes if they shared an association. The number of overlapping association between nodes is reflected in the tie strength. In other words, we created an $n \times n$ weighted adjacency matrix to visualize social networks that reflect the amount of shared understanding. Finally, we computed density and degree centrality of the networks.

Figure 1 depicts both social networks of shared understanding. Visually analysing them, it seems that the $U_{high}$ compared to the $U_{low}$ network is more interconnected and includes stronger relations (more shared associations) pointing towards a higher shared understanding. Only outlier is Mary and Joseph's relation with 12 shared associations, which could be due to parallel offline-collaboration at work. SNA confirms the observed difference in interconnectivity and reveals a higher density for the $U_{high}$ ($D = 1.00$) than the $U_{low}$ ($D = 0.89$) network: i.e. participants clicking on more unique social tags in the tag cloud have more edges to others due to overlaps in their association tests. As well, there is a difference in heavy weight edges reflected in a higher averaged node degree centrality (respecting edge count & weight) [5] for the $U_{high}$ (*deg.* = 14.95) than the $U_{low}$

(*deg.* = 11.72) network: i.e. $U_{high}$ participants have more higher weighted edges due to more overlapping associations. A comparison of means validates the difference as tendentially significant: $U(15) = 56$, $p = <.10$.

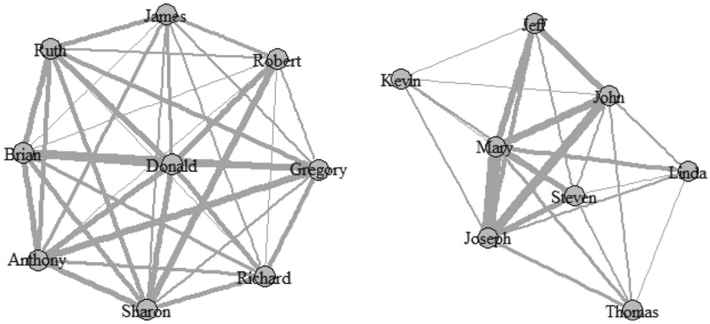



**Fig. 1.** $U_{high}$ (left) & $U_{low}$ (right) networks. Edge width is number of shared associations in AT.

### 3.2 Individual Level - Information Search

H2 assumes improved information search in terms of more explored resources and a faster exploration of themes during search in the $U_{high}$ condition. To quantify the latter search costs, we extracted the sequence of collected resources for each user and determined for each position $i$ in her resource sequence the number of unique theme combinations $n_i$ explored up to this point in time. Afterwards, we performed a regression of $n_i$ on $i$ and used the resulting slope $k$ as an average estimate of the users' rate of theme exploration. Finally, the categorical predictor uptake was included to explore whether the theme exploration is faster in the $U_{high}$ than the $U_{low}$ condition.

Figure 2 presents the average $n_i$ for a sequence of $i = 2–9$ resources for both conditions. Contrary to our expectation, it reveals a linear relationship with a larger slope (lower search costs) for the $U_{low}$ condition. For instance, in order to explore four theme combinations, $U_{low}$ participants needed to collect about 5 resources, while $U_{high}$ participants needed to collect about 7 ($U_{low}$: $n_5 = 4.25$, $SD = 0.71$; $U_{high}$: $n_7 = 4.33$, $SD = 1.24$). To derive estimates of the varying search costs, we performed a linear regression of $n_i$ on the two predictors $i$ and *condition* ($U_{low}$ vs. $U_{high}$). In particular, we applied the following regression model: $n_i = \beta_0 + \alpha X_0 + \beta_1 i + \beta_2 X_0 i + \varepsilon$ (1), where $X_0$ takes on the values 0 or 1, if the corresponding resource was collected by a participant of the $U_{low}$ or the $U_{high}$ condition. 130 data points entered the linear regression[1], explaining about 70 % of variance in the number of themes explored $n_i$ (adjusted $R^2 = 0.69$, $p < .001$). It yielded a highly significant effect for the predictor $i$ ($t = 8.60$, $p < .001$) and – in line with expectations – a highly significant interaction $\beta_2 X_0 i$ between this continuous and the categorical predictor *condition* ($t = -0.30$, $p < .001$). However, contrary to our

[1] Three participants (N = 17) collected not more than 8 resources and one only 6, resulting in 13(users) * 8(positions) + 3(users) * 7(positions) + 1(users) * 5(positions) = 130 data points.

expectations, the rate of theme exploration (slope) amounts to $\beta_1 = 1.09$ under the $U_{low}$ condition (intercept: $ß_0 = 0.53$), and declines to a rate of $\beta_1 + \beta_2 = 0.79$ under the $U_{high}$ condition ($\alpha = 0.47$; $\beta_2 = -0.30$; intercept: $ß_0 + \alpha = 1.00$).

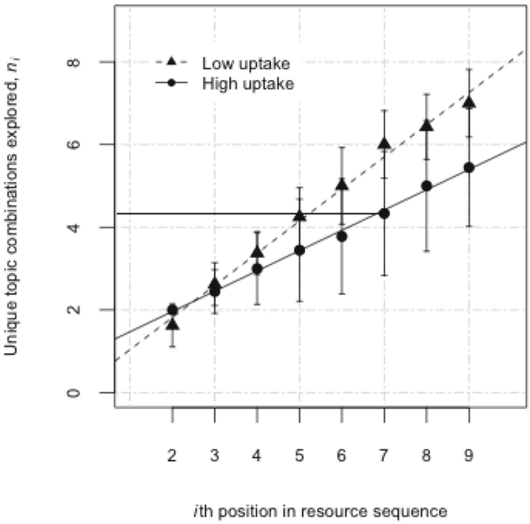



**Fig. 2.** Search Costs – average number of unique theme combinations $n_i$ explored at a given position $i$ in a resource sequence. *SD*s are indicated by error bars. A dashed and a solid line represent the linear regression of $n_i$ on $i$ for participants of the $U_{low}$ and the $U_{high}$ condition.

Moreover, more efficient search for $U_{high}$ should also be reflected in the number of explored resources. We found a correlation between uptake and explored resources ($r_{spearman} = 0.51(N = 17), p < .05$): i.e. the more unique social tags are clicked in the tag cloud, the greater is the number of explored resources. To validate correlation results, we computed a comparison of means that resulted in an affirmative significant difference between $U_{high}$ and $U_{low}$ condition as far as the exploration of resources is concerned: $M_{high} = 15.44$ (*SD* = 3.50), $M_{low} = 10.75$ (*SD* = 3.99), $t(14) = 2.56$, $p < .05$.

## 4 Discussion & Future Work

This social tagging study explored the social and individual benefits of engagement in MM based on uptake. High uptake of others' tags had a twofold effect: 1. Increase of shared understanding indicated by higher overlaps in collaborator's conceptual knowledge in ATs & 2. Narrower and more elaborative search indicated by a slower theme exploration with more considered resources. On the one hand, uptake seems to lead to a higher shared understanding of co-workers. Taking up others' tags and receiving parallel social stimulation could result in irritations and adaptations, called accommodative processes [1]. They specify internalization and externalization processes of coevolution and trigger the differentiation of underlying cognitive structures. Over time, these structures align establishing shared understanding. On the other hand, results indicate that uptake has an ambivalent effect on information search leading to more explored resources at the expense of higher search costs. This could be explained by the extent to which the search theme is narrow or broad. We assume social stimulation and respective accommodative processes to trigger an elaboration of a narrow theme (/limited theme combinations) and the related cognitive structures, which becomes manifest in a large number of semantically similar resources: i.e. a small rate at which new themes are explored. Since search costs measure the broadness of search via the assessment of explored theme combinations over time, this kind of search behaviour yields worse results. Therefore, extensive uptake might have led to more explored resources, but to increased search costs. In conclusion, the degree of uptake or engagement in MM, the "trialogicality" [6], seems to play a crucial role for experiencing benefits in individual and collaborative learning. Future work will consider the thematic focus of uptake and the role of assimilative processes, i.e. the repeated instantiation of existing cognitive structures, to better understand the effects of uptake onto search costs. For example, each reused social tag could be categorized by topics and weighted by the usage frequency to infer on the depth of elaboration of search themes. Furthermore, we will qualitatively validate and deepen the assumptions on professionals' tagging behaviour. Shedding light on MM and its underlying mechanisms is going to improve the design of collaborative working and learning systems as well as the structuring of pedagogical and workplace scenarios.

**Acknowledgment.** The work is funded by Know-Center GmbH (COMET Program managed by AT Research Promotion Agency FFG), Austrian Science Fund (FWF; Grant Project: 25593-G22) and EU-IP Learning Layers (Grant Agreement: 318209).

## References

1. Cress, U., Held, C., Kimmerle, J.: The collective knowledge of social tags: Direct and indirect influences on navigation, learning, and information processing. Comput. Educ. **60**(1), 59–73 (2013)
2. Dennerlein, S., Theiler, D., Marton, P., Rodriguez, P.S., Cook, J., Lindstaedt, S., Lex, E.: KnowBrain: an online social knowledge repository for informal workplace learning. In: Conole, G., et al. (eds.) EC-TEL 2015. LNCS, vol. 9307, pp. 509–512. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24258-3_48
3. Ley, T., Seitlinger, P.: Dynamics of human categorization in a collaborative tagging system: how social processes of semantic stabilization shape individual sensemaking. Comput. Hum. Behav. **51**, 140–151 (2005)
4. Jonassen, D.H., Beissner, K., Yacci, M.: Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge. Knowledge Creation Diffusion Utilization. Lawrence Erlbaum, Hillsdale (1993)
5. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths. Soc. Netw. **32**(3), 245–251 (2010)

6. Paavola, S., Hakkarainen, K.: From MM to joint construction of knowledge practices and artefacts: A trialogical approach to CSCL. In: Proceedings of CSCL, pp. 83–92 (2009)
7. Stahl, G.: Group cognition: Computer support for building collaborative knowledge. MIT Press, Cambridge (2006)
8. Stahl, G., Koschmann, T., Suthers, D.D.: Computer-supported collaborative learning: An historical perspective. Computer **4**, 409–426 (2006)
9. Suthers, D.D.: Technology affordances for intersubjective MM: A research agenda for CSCL. ijCSCL **1**(3), 315–337 (2006)

# Consistency Verification of Learner Profiles in Adaptive Serious Games

Aarij Mahmood Hussaan[1(✉)] and Karim Sehaba[2]

[1] IQRA Universirty, Main Campus, Karachi, Pakistan
aarijhussaan@iqra.edu.pk
[2] Université de Lyon, CNRS, Université Lyon 2, LIRIS, UMR5205, 69676 Lyon, France

**Abstract.** This article addresses issues of consistency verification of learner profiles in adaptive serious games. More precisely, our research objective is to propose models and tools that allow the user (learner, teacher or expert, depending on the context of application) to create coherent profiles consistent with domain knowledge. Our approach has been conceived and developed in the context of the platform GOALS. GOALS, as Generator Of Adaptive Learning Scenarios, is an online platform which allows the generation of learning scenarios, keeping into account the educational and entertaining aspects of serious games. For this, the knowledge in GOALS is organized into three layers: the domain concepts, the pedagogical resources, and the game resources. The profile is represented by a set of couples in the form <attribute, value>, where *attribute* corresponds to a concept, and *value* represents the learner competence in that concept. The profile is initialized by the user. During the game session, the profile is updated automatically according to dependencies among different domain concepts. In order to verify the learner profiles validity, we use a rule-based system which verifies, for every type of relation between concepts, the values between the source and the target concept. In this article, we present the formalization of our approach, as well as, its evaluation.

**Keywords:** Adaptive serious games · User profiles · Consistency profile · Scenario generation

## 1 Introduction

Adaptive educational systems provide learners with a personalized learning experience, according to their needs, preferences and skills. They use a learner model to keep track of an individual learner's skills, needs and preferences. These models are then used to dynamically adapt learner's learning experiences, thus, facilitating learning [2].

The quality of adaptation provided by the system greatly depends upon the consistency of the information kept in the learner profiles. The information is


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 384–389, 2016.
DOI: 10.1007/978-3-319-45153-4_31

consistent if it adheres to the semantics of the domain knowledge. Otherwise, then the learner profile will be termed as inconsistent.

The inconsistency of a profile can be caused by many sources. The instructor can, mistakenly, initialize the learner profile incorrectly, the learner can himself/herself update his/her profile incorrectly, or the processes for automatically updating the learner profile are not correct. Therefore, the need is to provide the users with the models and tools to verify the consistency of the learner profile.

To address this problem, we propose a rule-based approach that checks the consistency of the profile. Using this approach, the instructor, or the domain expert, can define the rules that must be respected when updating values in the learner profile. These rules will be defined for every kind of pedagogical relation, because every relation has different semantics about the flow of information between concepts.

We have implemented our approach in GOALS, Generator of Adaptive Learning Scenarios [7]. GOALS is an online platform, where instructors or domain experts can design the domain knowledge in the form of concepts and the relations between those concepts. They can also initialize and maintain the learner profiles. GOALS then generate adaptive pedagogical scenarios according to each individual learner profile. The learner profile can get updated manually by the instructor and/or automatically based on interactions between the learner and GOALS. While updating the learner profile, the instructor/domain expert can verify the consistency of the updated values.

The rest of the article is organized as follows: the next section presents the related works. Section 3 presents our approach for the verification of the learner profile. In Sect. 4, we present the integration of our approach in GOALS. We conclude in Sect. 5 with our conclusions and discussion.

## 2 Literature Review

The GUMS system, [5] which is based on Prolog, is aimed at providing a set of services for the maintenance of assumptions about the user beliefs. GUMS does not draw assumptions itself. Instead, it accepts and stores new facts about the user which are provided by the application system, verifies the consistency of a new fact with the currently held assumptions by trying to deduce the negated fact from the current assumptions, informs the application system about recognized inconsistencies, and answers queries of the application concerning its current assumptions about the user.

[1] describe and evaluate a two-stage personalized information retrieval system. Their CASPER system used classification to personalize the jobs according to the user profile. However, the knn algorithm, which they use, suffer from noise in the user profile for smaller values of k. They used a higher number of k to reduce the effect of noise in their user profiles.

These methods have some limitations. They require a certain expertise in logic programming. The representation of the rules is not straight forward, and creating them requires mastery of declarative programming.

Other approaches are based on Bayesian Networks (NB) like [8] that provides a flexible method to present differentiated trust and combine different aspects of trust. A BN-based trust model is presented for a file sharing peer-to-peer application. [3] uses BN to handle uncertainty in the user models. Authors also present the work done on the ANDES ITS. [4] proposes a model and an architecture for designing intelligent tutoring system using BN. BNs are used to assess user state of knowledge and preferences, in order to suggest pedagogical options and recommend future steps. [6] combines the overlay model with Bayesian networks to infer user knowledge from evidence collected during the user interaction with the user.

The use of Bayesian network(BN) for user modeling also have some limitations. In a BN, there could be only one type of relation. Furthermore, the semantics of this relation are defined by the Bayes Rules. Therefore, we cannot have semantically different relations in a BN. Also, creating and maintaining a BN is not a trivial task. Although, BNs could be learned automatically using machine learning techniques. The quality of the learned BN relies heavily upon both the quantity and quality of the data. Both of which could be unavailable during learner's model conception.

In the next section, we present our approach for profile verification.

## 3 Verification of Profile

There are different types of pedagogical relations in our domain model. Therefore, the information between concepts flows differently according to different pedagogical relations. Consequently, different types of validations are required for different relations to maintain consistency among learners profile.

In order to address this issue, we propose to attach a set of rules with each type of relation. These rules will determine whether the values, to be updated in the learner profile, are valid or not. The obvious advantage of this approach that it will not limit the instructor or domain expert in the types of relation s/he wants to have in the domain knowledge. The responsibility of maintaining consistency will remain attached to every particular relation type. Since, each relation will maintain consistency individually, it will result that all the values in the learner profile will remain consistent.

In our previous works, we have defined a relation R as follows: R = $< C_{From}$, T, RC+>, where $C_{From}$ is the origin concept of the relation and T is type of relation defined as T = $<$ Name, Description, $F_{Type} >$, where Name is the name of the relation, Description is the description of the relation, and $F_{Type}$ is the function used to calculate the dependencies of the concept $C_{From}$ to the concept $C_{To}$ linked via this relation.

*RC* is relation of concepts defined as RC $=< C_{To}$, F, Value $>$, where $C_{To}$ is target concept of the relation, the direction of relation is from $C_{From}$ to $C_{To}$. $F$ is function that calculates the value used by $F_{Type}$. If the function $F$ is absent, then Value is used by $F_{Type}$ to calculate the dependencies between the concepts of this relation. The function $F_{Type}$ is used to propagate the information in the graph and is used to update the learners profile.

For validation, we augment the definition of $R$ with another function $F_{Update}$. The function $F_{Update}$ takes a value $v$ and a learner $l$ as an input and returns a boolean $b$ as output. The value $v$ defines the value to be updated in $C_{To}$, $l$ defines the learner profile in question and $b$ returns whether the updating is valid or not. The mapping $F_{Update}$: $v * l \rightarrow b$ depends upon the set of rules that are defined in $F_{Update}$,. We are defining here the rules to create rules. The instructor or domain expert can use a strategy of his/her choice to implement the different function.

Every time a user wants to update the values in a learner profile, s/he can verify whether the values are consistent or not. This verification will be done by applying the $F_{Update}$, function for all the values in the learner profile and the user will be notified of the results of the function.

Let us demonstrate the above mentioned concepts through an example. Suppose we have a relation of type *PreRequisite*. This relation states that if a concept A is a prerequisite of concept B, then it is mandatory for the learner to have sufficient mastery of concept A before studying concept B. Using our proposed models, this relation will be modeled as follows: $R_{PreRequisite}$ <B, PreRequisite, $F_{Update}$, RC+ >, RC = <A, $F_{PreReq}$, 60>. This means that a learner need to have a mastery of 60 % in concept A before accessing concept B. The $F_{Update}$ could be defined by the following set of rules:

1. Find all the concept that are in relation of type *PreRequisite* that are in relation with the concept $C_{From}$;
2. Verify whether the learner $l$ has more than the required competence in all the concepts found in the first step;
3. If the learner $l$ has the required competence in all the concepts return True and update the value of $C_{From}$;
4. If not, then highlight the concepts violating the rules and return False.

Now, if a learner $l$ has a competence of 40 % in concept A, and the instructor wants to put 100 % as the competency of $l$ in concept B. Then this will not be allowed by the function $F_{Update}$ of the $R_{PreRequisite}$. Step 2 of $F_{Update}$ will not pass and according to step 4 the function $F_{Update}$ will return false.

## 4 Validation in GOALS

To validate our approach in a real world setting, we implemented it in GOALS. We tested our implementation in the CLES project which aims to develop a serious game environment designed to diagnose and/or treat cognitive disabilities for children. Thus, the developed game contains different mini games according to various cognitive troubles. These games are designed to diagnose/treat specific cognitive troubles in their users.

CLES is an ideal project for testing because it contains large number of concepts and the even larger number of different pedagogical relations between those concepts. More precisely, there are 8 main concepts, and these concepts are further divided into subconcepts, totaling up to around 40 concepts. Furthermore, there are around 45 relations between the concepts. CLES also has users in thousands. Here we present some basic modelling of the project CLES, as the real model is quite large.

The eight main concepts are: Perception, Attention, Vision-Spatial, Memory, Oral Language, Written Language, Logical Reasoning, and Mixed Competencies. These concepts have sub-concepts. For example, Perception is further divided into: visual perception and auditory perception. These concepts have many pedagogical relations among themselves. For example, the Has-Parts relation between Perception and its sub-functions: Visual perception and Auditory perception. Similarly, Visual perception is "Prerequisite" of written Language.

CLES also have learner/user profile of thousands of persons in situation of a cognitive disability. Each has different profile values. The profiles have been introduced by a speech and language expert. To conduct the experiments in GOALS with the project CLES, we defined an experimentation protocol as follows: Firstly, the expert starts by entering multiple learner profiles that do not contain any inconsistencies. Secondly, we modify the profiles, introduced in the first step, to introduce deliberate inconsistencies in the profiles. Thirdly, we verify whether GOALS, using our approach, can detect the inconsistencies introduced in the 2nd step.

We followed the protocol for our experimentation. First, we had many profiles in GOALS that were introduced by a speech and language expert associated with project CLES. Given in Table 1 are some of the profiles and their values.

The columns of Table 1 indicate the concepts and the rows indicate the values of the different profiles concerning those columns. The cells that have a values of the format X/Y, indicates the values (X) entered by the expert, and the erroneous values (Y) introduced by us deliberately. Then following the 2nd step of the protocol, we introduced deliberate inconsistencies in the profiles. These errors were introduced to verify whether our approach is detecting the errors.

Then following the 3rd step of the protocol we checked whether the inconsistencies were detected by GOALS. The last two columns indicate the number of errors introduced, and the number of errors found respectively.

For example, following 1st step of the protocol, if we consider profile 1, then the value of the concept 1 introduced by the expert is 100, and, following the 2nd step of the protocol, the erroneous value introduced by us is 0. According to the domain model of project CLES, Perception is a *PreRequisite* of the concept Memory. Hence, a scenario, where the learner does not have any capacity of Perception and some capacity of Memory, is inconsistent. Following the 3rd step of the protocol, we checked whether GOALS will be able to detect it or not, and we found that this inconsistency is successfully detected by GOALS.

## 5 Conclusion

In this article, we presented an approach for detecting inconsistencies in the learners profile. We argued that the failure to detect them may result in a suboptimal learning experience for the learner. We showed that the existing approaches

**Table 1.** The learner profiles, with the consistent values and inconsistent values.

| Profiles | Perception | Attention | Visual Spatial | Memory | Raisonnement Logique | Langage Ecrit | Lecture | Oral Language | Orthographe | Gnosis | Memory Grill | Sequential Memory | Working Memory | Visual Memory | Errors Introduced | Errors Caught |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P1 | 80/00 | 80 | 80/0 | 80 | | | | | | | | | | | 1 | 1 |
| P2 | 100/10 | 100/10 | 40 | 100 | 100 | 40 | 40 | 50 | 30 | | | | | | 2 | 2 |
| P3 | 80/80 | 80 | 30 | 80 | 80 | 80/0 | | 30/0 | | 80 | | | | | 2 | 2 |
| P4 | 100/100 | 60 | 40 | 70 | 70 | 70 | | | | 100 | 40/0 | 40/0 | | 40 | 2 | 2 |
| P5 | 100/100 | 100 | 100 | 100 | 100 | 100/0 | | | | | 100 | 100 | 100 | | 1 | 1 |
| P6 | 100/10 | 100/10 | 100 | 80 | 65 | 65/0 | | | | | 75 | 70 | 60 | | 3 | 3 |

have some limitations. Afterwards, we demonstrated our rule based approach for detecting inconsistencies in the learner profile. We proposed to attach with every pedagogical relation a function that verifies, for every profile, whether the value to be updated will introduce any inconsistency or not. We applied our approach in GOALS and tested it on the project CLES. We discussed the experimentation protocol, as well as, the results of the experimentation.

## References


1. Bradley, K., Rafter, R., Smyth, B.: Case-based user profiling for content personalisation. In: Brusilovsky, P., Stock, O., Strapparava, C. (eds.) AH 2000. LNCS, vol. 1892, pp. 62–72. Springer, Heidelberg (2000)
2. Brusilovsky, P., Karagiannidis, C.: The benefits of layered evaluation of adaptive applications and services. In: Evaluation of Adaptive (2001)
3. Conati, C., Gertner, A., Vanlehn, K.: Using Bayesian networks to manage uncertainty in student modeling. User Model. User-Adap. Inter. **12**(4), 371–417 (2002). http://www.springerlink.com/index/UH10863178301011.pdf
4. Gamboa, H., Fred, A.: Designing intelligent tutoring systems: a Bayesian approach. In: Enterprise Information Systems III (Fred 1994), p. 146 (2002)
5. Kobsa, A.: Generic user modeling systems. User Model. User-Adap. Inter. **11**(1–2), 49–63 (2001)
6. Nguyen, L., Do, P.: Combination of Bayesian network and overlay model in user modeling. In: Allen, G., Nabrzyski, J., Seidel, E., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2009, Part II. LNCS, vol. 5545, pp. 5–14. Springer, Heidelberg (2009)
7. Sehaba, K., Hussaan, A.M.: GOALS: a platform for the generation of adapive pedagogical scenarios. Int. J. Learn. Technol. **8**(13), 224–245 (2013)
8. Wang, Y., Vassileva, J.: Bayesian network-based trust model. In: Proceedings - IEEE/WIC International Conference on Web Intelligence, WI 2003, pp. 372–378 (2003)

# MoodlePeers: Factors Relevant in Learning Group Formation for Improved Learning Outcomes, Satisfaction and Commitment in E-Learning Scenarios Using GroupAL

Johannes Konert[1(✉)], Henrik Bellhäuser[2], René Röpke[3], Eduard Gallwas[3], and Ahmed Zucik[3]

[1] Beuth University of Applied Sciences Berlin, Berlin, Germany
johannes.konert@beuth-hochschule.de
[2] Department of Psychology, University of Mainz, Mainz, Germany
bellhaeuser@uni-mainz.de
[3] TU Darmstadt, Darmstadt, Germany
roepkix@gmail.com, egallwas@gmail.com, zukic07@gmail.com

**Abstract.** High-scale and pure online learning scenarios (like MOOCs) as well as blended-learning scenarios offer great possibilities to optimize the composition of learning groups working together on the assigned (or selected) tasks. While the benefits and importance of peer learning for deep learning and improvement of e.g. problem-solving competency and social skills are indisputable, little evidences exist about the relevant factors for group formation and their combination to optimize the learning outcome for all participants (in all groups). Based on the GroupAL algorithm, MoodlePeers proposes an plugin solution for Moodle. Evaluated in a four-week online university mathematics preparation course MoodlePeers proved significant differences in submission rate of homework, quality of homework, keeping up, and satisfaction with group work compared to randomly created groups. The significant factors from personality traits, motivation and team orientation are discussed as well as the algorithmic key functionality behind.

**Keywords:** Group formation · Learning outcome · Learning goal alignment · Peer learning · Personality traits · Motivation · Expectation · Optimization

## 1 The Group Formation Problem: Importance of Peer Learning in a Network Learning Environment

The potential of peer education, especially peer assessment, peer feedback and peer collaboration, has long been reported and recognized [1]. Especially for improvement of problem solving competency, peer collaboration in small learning groups is valuable for all group members. Ideally, they amend each other in prior knowledge and have a similar attitude to the expected result. Consequently, beside the factors of the learning scenario (e.g. group-size and suitable open-format tasks) the characteristics of the


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 390–396, 2016.
DOI: 10.1007/978-3-319-45153-4_32

learners themselves and their matching are essential for fruitful group collaboration. In the further course we discuss the foundations from educational psychology about relevant matching criteria and give a brief overview over existing algorithms. The focus of the remaining sections is on the plugin-solution *MoodlePeers* for the *Learning Management System* (LMS) Moodle (https://moodle.org/) which assists in the algorithmic creation of optimized learning groups. Finally, a user study with 510 participants is described. The results lead to recommendations for learning group formation.

## 2 Related Work

**Foundations in Educational Psychology on Relevant Criteria.** If a group is not properly formed, individual members start solo attempts to solve the given tasks alone, the group degenerates and motivation of other members decreases [2]. More details about negative aspects can be found in [3]. Concerning proper matching criteria, demographic characteristics, such as gender, age, or educational level have been shown to be less important for team performance than deep-level composition variables from psychological tests, such as personality factors and attitudes [4]. Humphrey et al. [5] argued that *extraversion* would be a relevant matching criterion that should be distributed heterogeneously, as *extraversion* is associated with leadership. On the other hand, *conscientiousness* should be distributed homogeneously, as it is necessary for teams to adjust their goals. As Bell [6] described, *team orientation* (i.e. the preference for teamwork) should be distributed homogeneously, so that team members agree on the degree to which they cooperate. Nederveen Pieterse et al. [7] showed that homogeneity in *motivation* and *goal orientation* is associated with higher team performance. General mental ability and *prior knowledge* have been argued to be important for team performance as well [8]. However, the interplay of these factors remains an open question.

**Algorithmic Approaches to Group Formation.** Comprehensive discussions about algorithmic approaches to learning group formation have been published in [9]. Concerning input data, approaches exist that derive (incomplete) data about learners' goals and characteristics from interaction with e-learning systems (e.g. LMS). Such approaches train their models to form groups for specific learning scenarios and settings. The main benefit is the abolition of extra effort (for students) to fill out electronic surveys. On the contrary, little is known about the relation of interaction features with group learning quality [10]. MoodlePeers uses a different approach by generating dynamic questionnaires based on the desired learning scenario, like in [11]. Concerning algorithmic design, semantic approaches and logical solvers can be used to find appropriate group constellations, e.g. [12]. The main disadvantage is the rarely available semantic data basis for various learning domains. Finally, a variety of non-linear optimization approaches exist. Only few of them are designed to support homogeneous and heterogeneous to match criteria at the same time, e.g. [11]. Additionally, for various formal learning scenarios, it is essential to optimize all groups equally good to ensure fairness for all students. Besides GroupAL, to the best of our knowledge, no other algorithmic approach considers this yet. Thus, MoodlePeers uses GroupAL [9].

# 3 MoodlePeers: LMS-Plugins for Optimized Learning Outcomes

The MoodlePeers group formation plugin for Moodle LMS was created in an agile process over the last two years and released as Open Source code[1]. The relevant criteria data is collected using standardized and internationally as reliable proved questionnaires (see Sect. 4). As such the following results about group work performance are expected to be better transferable to other course topics, scenarios and cultural contexts. Figure 1 (left) shows a part of the questionnaire provided to students when they click on the group formation activity. Questions on prior knowledge can be given by the teacher as shown in Fig. 1 (right). When the group formation task is done (usually in seconds to minutes after the submission deadline for questionnaires is reached) students see their group members via Moodle group displaying and within their tab *group assignment* next to their questionnaire tab. The groups can be used for any group activity in Moodle.

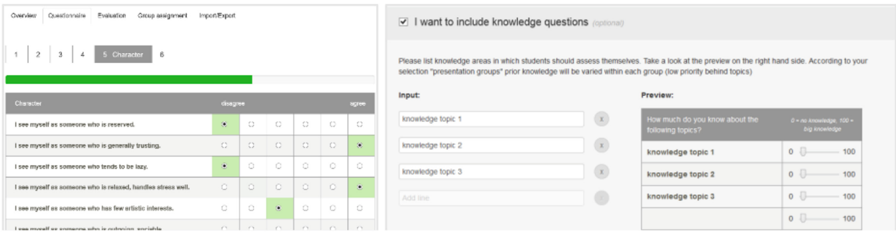



**Fig. 1.** (left) Student interface, part of the group formation activity's questionnaire, (right) Teacher interface, setup-part of group formation (preview aside)

# 4 Evaluation Design and Scenario

Based on the related work findings, this contribution investigates the following two hypotheses mainly: **(H1)** A learning group formation based on a thoroughly selected set of personal characteristics (criteria) significantly *improves individual learning outcomes* (like quality of solutions to assignments); **(H2)** A learning group formation based on a thoroughly selected set of personal characteristics (criteria) significantly *impacts values of commitment* (like drop-out rate, daily time investment, individual satisfaction with the learning group and course).

For evaluation an online mathematics preparation course was used, where prospective students of mathematically oriented fields of study (e.g. computer science and mathematics) could voluntarily take part in. During the four weeks of the course (07.09.–04.10.2015), students had access to the course structure via Moodle that included instructions and tests. Via MoodlePeers participants were asked to fill in a demographic questionnaire at the beginning of the preparation course including the following

[1] See https://github.com/moodlepeers/. The acceptance in the Moodle plugin repository (http://moodle.org/plugins/mod_groupformation/) is currently in progress.

psychological measures: Big Five personality traits extraversion (het.), neuroticism, conscientiousness (hom.), social agreeableness (het.), and openness for new experiences (het.) [13], a questionnaire of team orientation (hom.) [14], and a questionnaire of motivation (hom.) [15]. Furthermore, participants rated their prior knowledge (het.) in all six chapters of the preparation course (in parenthesis indicated when used for homogeneous or heterogeneous matching later). Before the actual group formation, the sample was equally and randomly divided into experimental half and control half. Within each, learning groups of 5 persons were formed. While for the experimental half the GroupAL algorithm was used, the control half was grouped by a random algorithm. Participants were not informed about any difference.

Five-hundred and ten participants filled in the demographic questionnaire and were assigned into 51 experimental groups and 51 control groups. Due to a loss of data of 154 students, we only collected questionnaires of 356 students (267 male, 89 female; 216 in experimental groups, 140 in control groups). Learning groups were informed about their group members and received access to a bulletin board for group communication. Weekly assignments provided open-ended, complex math problems. Participants were asked to work on the problems individually and then to discuss within their learning groups. The groups handed in one agreed solution. Two tutors rated the solutions independently on a 3-point scale (0 = not sufficient; 1 = sufficient; 2 = outstanding). By summing up the points for all four weekly assignments we calculated an overall score for the assignments (0 to 8 points). Furthermore, participants had access to a mathematics pre-test at the beginning of the preparation course that provided feedback on their individual competencies, and also to a mathematics post-test at the end. The final evaluation sheet included questions on average daily time investment (hours), satisfaction with group members (6-point Likert scale), and satisfaction with the preparation course (6-point Likert scale).

## 5 Results

Due to technical problems of the Moodle server (not related to the *MoodlePeers* plugin), the preparation course suffered from a serious dropout rate that also affected participation in the evaluation study. While 254 participants took part in the pre-test, only 50 participants also used the post-test. Evaluation sheets were filled in by 55 participants. The dropouts impaired statistical testing in several cases and limited the generalization of results. However, a chi-squared test revealed that mere participation in the post-test differed significantly between groups ($X^2 = 4.957$; $df = 1$; $p = .026$) after controlling for participation in the pre-test. This can be interpreted as a sign that participants in algorithmically formed groups *persevered longer* in the preparation course and showed a smaller dropout rate. As depicted in Fig. 2 (left) the same trend could be observed in the weekly assignments. While in the groups formed by *MoodlePeers*, participation rates were approximately between 20% and 30%, in the control groups the participation rates fell from approximately 12 % in week 1 to 0% in the last week. When calculating the overall score for the assignments, we found a significant difference ($t = 6.079$, $df = 336.6$, $p < .001$) with the experimental groups ($M = 1.32$; $SD = 1.64$) outperforming the control groups ($M = 0.51$; $SD = 0.83$). Therefore, not only the *dropout rates* were lowered, but also the *quality of solutions to assignments* was improved.

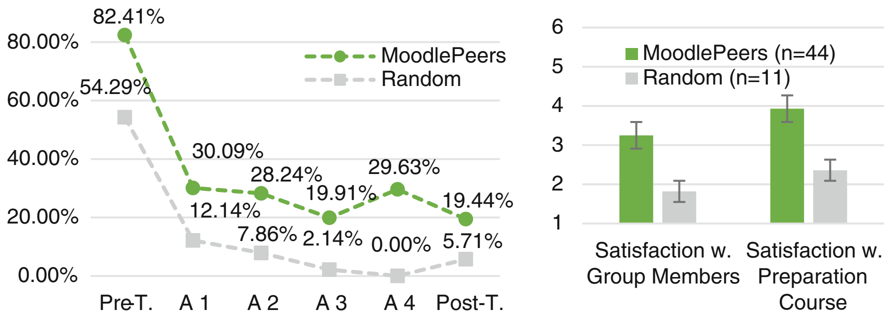



**Fig. 2.** (left) Participation rates in the pre-test, the four weekly assignments (A1-A4), and the post-test, (right) satisfaction with group members and with the preparation course in general for experimental condition (MoodlePeers groups) and control condition (random groups)

The final evaluation sheet revealed that participants in the experimental groups were significantly *more satisfied* with the selection of group members ($t = 3.645$, $df = 27.3$, $p < .001$) and also with the preparation course in general ($t = 2.892$, $df = 14.6$, $p = .011$) compared to control groups (Fig. 2 right). Furthermore, participants in the experimental groups reported a higher daily time investment, but this difference marginally missed statistical significance ($t = 1.724$, $df = 25.8$, $p = .097$).

## 6 Interpretation, Conclusion and Outlook

The results of group formation with the *MoodlePeers* plugin for Moodle are promising: We were able to demonstrate that considering personality traits for group formation can have a statistically significant effect on *self-reported satisfaction*, *perseverance*, and *performance* when compared to random matching. This finding is even more substantiated by the rigorous experimental design in which participants were blind to the manipulation. Our findings cover all stages of this theoretical model: We found improved satisfaction with the group members as well as with the preparation course in general and also participants in the experimental groups persevered until the end of the preparation course to a higher percentage despite the technical problems. While we were not able to test the effect on the mathematics post-test due to too few individuals in the control groups remaining at the end, we did find quality of the weekly assignments to be higher in the experimental groups. Therefore, we find our hypotheses to be confirmed: The group formation as performed by the *MoodlePeers* plugin improved both group learning performance (hypothesis 1) as well as values of commitment (hypothesis 2). As the group formation algorithm in the present study used a variety of matching criteria (prior knowledge, personality factors, motivation, and team orientation), it remains an open question whether a different selection of criteria would improve the results further. Thus, our next user study will investigate this in more detail. Concerning the broader field of algorithmic support for learning group formation, a support for minority protection would prevent learning groups with group members being alone with their characteristic (e.g. gender or spoken language). Next, indicators to detect inactive groups have to be found and solutions for re-grouping.

**Acknowledgements.** This work has partly been funded by the TU Darmstadt Quality Program for excellent teaching. Additionally, special thanks to the co-workers on this interdisciplinary work, especially Regina Bruder, Annette Glathe, Christian Hoppe, Steffen Pegenau, Christoph Rensing, Diana Seyfarth, Marcel Schaub, Klaus Steitz, and Nora Wester (all TU Darmstadt).

## References

1. Damon, W.: Peer education: the untapped potential. J. Appl. Dev. Psychol. **5**, 331–343 (1984)
2. Michaelsen, L.K., Fink, L.D., Hall, A.: Designing effective group activities: lessons for classroom teaching and faculty development. In: DeZure, D. (ed.) To Improve the Academy: Resources for Faculty, Instructional and Organizational Development. New Forums, Stollwater, OK (1997)
3. Srba, I., Bielikova, M.: Dynamic Group formation as an approach to collaborative learning support. IEEE Trans. Learn. Technol. **8**(2), 173–186 (2014)
4. Harrison, D.A., Price, K.H., Gavin, J.H., Florey, A.T.: Time, teams, and task performance: changing effects of surface- and deep-level diversity on group functioning. Acad. Manag. J. **45**, 1029–1045 (2002)
5. Humphrey, S.E., Hollenbeck, J.R., Meyer, C.J., Ilgen, D.R.: Trait configurations in self-managed teams: a conceptual examination of the use of seeding for maximizing and minimizing trait variance in teams. J. Appl. Psychol. **92**, 885–892 (2007)
6. Bell, S.T.: Deep-level composition variables as predictors of team performance: a meta-analysis. J. Appl. Psychol. **92**, 595–615 (2007)
7. Nederveen Pieterse, A., van Knippenberg, D., van Ginkel, W.P.: Diversity in goal orientation, team reflexivity, and team performance. Organ. Behav. Hum. Decis. Process. **114**, 153–164 (2011)
8. Horwitz, S.K.: The compositional impact of team diversity on performance: theoretical considerations. Hum. Resour. Dev. Rev. **4**, 219–245 (2005)
9. Konert, J., Burlak, D., Steinmetz, R.: The group formation problem: an algorithmic approach to learning group formation. In: Rensing, C., de Freitas, S., Ley, T., Muñoz-Merino, P.J. (eds.) Proceedings of the 9th European Conference on Technology Enhanced Learning (EC-TEL), pp. 221–234. Springer, Berlin, Graz, Austria (2014)
10. Zheng, Z.: A dynamic group composition method to refine collaborative learning group formation. In: Proceedings of the 6th International Conference on Educational Data Mining (EDM). pp. 360–361 (2013)
11. Cavanaugh, R., Ellis, M.: Automating the process of assigning students to cooperative-learning teams. In: Proceedings 2004 American Society for Engineering Education Annual Conference & Exposition (2004)
12. Isotani, S., Mizoguchi, R.: Theory-driven group formation through ontologies. Int. J. Comput. Collab. Learn. **4**, 445–478 (2009)
13. Rammstedt, B., John, O.P.: Kurzversion des big five inventory (BFI-K): entwicklung und validierung eines ökonomischen inventars zur erfassung der fünf faktoren der persönlichkeit. Diagnostica **51**, 195–206 (2005)

14. Hosiep, R., Paschen, M.: Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung - 6 Faktoren. Hogrefe, Göttingen (2012)
15. Rheinberg, F., Vollmeyer, R., Burns, B.D.: FAM: Ein Fragebogen zur Erfassung aktueller Motivation QCM: A questionnaire to assess current motivation in learning situations. Diagnostica **47**, 57–66 (2001)

# Towards a Capitalization of Processes Analyzing Learning Interaction Traces

Alexis Lebis[1,2(✉)], Marie Lefevre[2], Vanda Luengo[1], and Nathalie Guin[2]

[1] Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, Paris, France
{alexis.lebis,vanda.luengo}@lip6.fr
[2] Université de Lyon, CNRS, Université Lyon 1, LIRIS, UMR5205, Lyon, France
{alexis.lebis,marie.lefevre,nathalie.guin}@univ-lyon1.fr

**Abstract.** Analyzing data coming from e-learning environments can produce knowledge and potentially improve pedagogical efficiency. Nevertheless, TEL community faces heterogeneity concerning e-learning traces, analysis processes and tools leading these analyses. Therefore, analysis processes have to be redefined when their implementation context changes: they cannot be reused, shared nor easily improved. There is no capitalization and we consider this drawback as an obstacle for the whole community. In this paper, we propose an independent formalism to describe analysis processes of e-learning interaction traces, in order to capitalize them and avoid these technical dependencies. We discuss both this capitalization and its place and effects in the iterative learning analysis procedure.

**Keywords:** Learning analytics · Capitalization · Analysis process · Interaction traces · Operator · Analysis tool · e-learning

## 1 Introduction

The e-learning is defined by the use of digital environments that can be networked. Its aim is to reinforce the construction of knowledge by learners. These environments can produce data that relate the interaction of users among themselves (e.g. private messages, forum, etc.), with the system or even with resources. Thereafter, we name *traces* this data of learning interaction. These traces can be considered as knowledge warehouses since their analysis brings knowledge out of them. However, there is no solution to easily share, enrich or reuse such analysis processes of interaction traces nor the knowledge they are producing. Consequently, when the implementation context changes (e.g. analysis tool, formats of data used), analysis processes have to be reworked, frequently from scratch. Thus, in such a situation, TEL community cannot have an effective awareness of what is existing or what is redundant.

In this paper, we introduce our approach to bring capitalization for analysis processes. It relies on a formalism for describing analysis processes of learning interaction traces independently of analysis tools, which aims to avoid technical specificities. Moreover, we discuss the place of such analyses capitalization in the iterative learning analysis procedure, and the potential actors involved in it.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 397–403, 2016.
DOI: 10.1007/978-3-319-45153-4_33

## 2 Related Works

**Analysis Processes of Interaction Traces.** An analysis process of traces is the use of operations, made by operators, over data in order to produce knowledge (e.g. indicators) addressing needs [1]. These analyses can be classified according to expected knowledge as descriptive (describe what happened), predictive (determine prospects), diagnostic (understand why something happened) or prescriptive (identify best decisions) [4]. Moreover, their inner steps have be widely covered too [4,8], and three steps can be identified as recurrent across different fields: preprocessing, analysis of relevant data and post-processing. Other steps are more specific to the TEL field such as publication of results or reuse of data [8], giving us clues regarding capitalization needs.

Some works consider an analysis process as an organized and fixed combination of operators [6]. It can be seen as a "black box" and be reused in another analysis process [5]. This property led to methods that reinforce the importance of capitalization, such as the *discovery with models* method where previous developed models are used as components for other analyses [2].

**Analysis Tools.** TEL community has at its disposal a variety of cross-field analysis tools, like RapidMiner[1] or R[2], and specialized solutions. For instance, UnderTracks (UT) takes into consideration data and operators life cycle within analysis [6]. We can also cite Usage Tracking Language (UTL), which calculates and describes indicators by mapping data coming from heterogeneous traces into more generic ones expressed in XML [3]. All these tools can be classified into three categories [6]: data storage, data analysis (like R) and both data storage and analysis (like UT). Our work concerns only the tools designed for analyses.

**Capitalization.** Since analysis tools implement operators that are strongly dependent of data formalism in order to be computed, they are poorly permissive. As a result, some works suggest to work with a more generic data formalism before making any analysis, like Caliper Analytics[3] or UTL. These tools map cross-origin data into a regulated formalism, allowing reproducible analyses. But the analysis capitalization is not guaranteed: they are done in a specific tool and produce specific formatted data. Such analysis processes cannot be shared and reused as they are in other tools.

For all we know, capitalizing analysis processes has not been worked out in TEL community. Despite the fact that some works aim to share results of analyses [8] or customised operators [6], they are mainly tool specific and there is no federation between these tools. *Ipso facto*, TEL community is confronted to the difficulty of being aware of what already exists, involving re-implementation of pre-existing analyses. However, non TEL specific works go in this capitalization direction, like Predictive Model Markup Language (PMML)[4]. PMML aims to share predictive and machine learning models, trained or not, between free and non-free analysis tools: this is a clue for us that the need of capitalization is real.

[1] https://rapidminer.com/.
[2] http://www.revolutionanalytics.com/.
[3] https://www.imsglobal.org/activity/caliperram.
[4] http://dmg.org/pmml/v4-1/GeneralStructure.html.

## 3 Preliminary Assumptions

As shown in the previous section, there is, to our knowledge, no effective and easy way to capitalize analysis processes of e-learning traces. Thus we focus on how analysis processes of traces and their inner components can be described in such a way that they are not related to a specific analysis tool. To do so, we based our work over three assumptions. Firstly, (A1) we assume that designing analysis processes is a cognitive task and is realized by manipulating the meaning of data instead of specific values of them. Indeed, Rosch expresses the fact that the cognition is made *via* categories playing the role of cognitive reference models instead of elementary instances [7]. Secondly, (A2) we assume that since this design is a cognition process, specificities of analysis tools are not taken into consideration. Thus, an analysis process can be regarded as a set of elementary operations. Finally, as our state of the art suggests it, (A3) an analysis process can be seen as a non linear ordered succession of operations, taking inputs and producing outputs: this brings up an important sequential property for ordering.

## 4 The Capitalization of Analysis Processes

### 4.1 Where to Capitalize in the Iterative Learning Analysis Procedure?

We notice in the literature three main steps concerned by the capitalization of analysis processes. They are (S1) selection of relevant data and consideration of context constraints, (S2) preparation of the analysis and (S3) implementation of the analysis. From users' point of view, two roles are mainly involved in these steps: the e-learning tool expert and the analyst.

The step about the selection of relevant technical information (S1) is often implicit in the literature but since it requires practical knowledge, we consider it as an independent step [9,10]. The e-learning tool expert is involved during this step. He/she has expertise about the technical context of needs, like the pedagogical domain, the pedagogical platform, learning traces produced, data subjects. Thus, this expert makes the needs more concrete by communicating these information and can eventually detect some inconsistencies or limits.

The preparation of the analysis step (S2) is realized by the analyst. This role is played for instance by data miners, statisticians or researchers. Thanks to his/her expertise in the analysis field and information obtained from S1, the analyst designs the analysis in order to address needs. This implies setting up its limits as well as its strategies, defining which data is pertinent or even how pedagogical domain specificities should be used. Hence, this is a complex step which requires a strong interactivity with experts in order to correctly understand and exploit the context of the analysis.

The outcome of this two previous steps is the analysis step (S3), nearly always realized by the analyst. Accordingly to our state of the art, many papers are concerned about analysis methods in several domains (e.g. EDM, LAK). In any case, the objective is to produce knowledge addressing needs (e.g. dropout rates in a MOOC). S4 is strongly bound with S2 and S3, providing the possibilities to refine the supplied information, the analysis and even the needs.

Finally, we suggest that capitalizing analysis processes, through a capitalization step, can occur at two moments: S2 or S3, when they are designed (e.g. drew on paper), and S4 once they are implemented inside analysis tools. In both cases, this capitalization should be made by describing analysis processes with the formalism presented in the Sect. 4.3, which is not constrained by technical tools specificities. The description can be realized by the analyst due to his/her analysis expertise, and also by the e-learning tool expert since designed analysis processes are pertinent to be capitalized. As we can note here, one of our proposition's strengths is its integration to the analysis procedure without modifying it: we enrich it and provide potential supports to actors involved.

## 4.2 Independence Using Operators

According to the non-linear assumption (A3), in order to describe a designed or implemented analysis process independently from analysis tools, all its inner operators must also be described independently. Thus, we represent an independent operator as the concept conveyed by semantic equivalent operators, implemented in different analysis tools. For example, let us consider a temporal filter. The way it is implemented, as well as the way of using it, differ between analysis tools. However, the underlying concept is to apply a filtering over a time: this is what is represented by such an independent operator *Temporal Filter*.

The cognition assumption (A1) implies that independent operators, and then *independent analysis process* (IAP), do not process data directly: data are designed with concepts instead of specific instances of concepts. Hence they only manage data concepts, for keeping track of what is available at each step of the analysis (see Fig. 1). Data are conceptualized under a notion of *type of traced elements* (TTE) representing the concept they convey. For instance, if a student made an event at 11:02 am, then 11:02 am is a datum's value and the TTE is *time*. Besides, a *temporal filter operator* will use as input *time*, not 11:02 am.

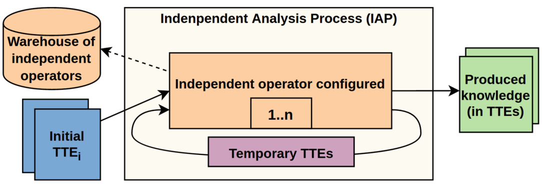



**Fig. 1.** Representation of the description of an independent analysis process (IAP).

Hence, IAP are adaptive concerning initial data requirements because they are not based upon values themselves but upon TTEs, offering capitalization abilities. Also, since the produced knowledge is also expressed with TTEs, the description of a process through independent operators potentially grants a greater semantic and insights of it. Accordingly, this ensures that, for any IAP, if given data match the prerequisites TTEs, then expected knowledge can be obtained.

## 4.3 Meta-Models of Our Approach

We present in this section how both IAP and independent operators have to be represented in order to allow the capitalization of analysis processes.

Firstly, the independent operator meta-model (see Fig. 2, right part) has been obtained by an iterative process of identification and comparison between operators in different analysis tools such as UnderTracks, RapidMiner or Weka (http://www.cs.waikato.ac.nz/ml/weka/). It describes how an operator has to be constructed in order to be independent of technical specificities. Moreover, it describes how input TTEs will evolve when applying operator on them, according to processing behavior rules *OutputSheet*. For instance, the rule for a clustering operator can be to create a new TTE, representing new groups.

Furthermore, independent operators require few properties in order to exist per se, such as the number of input and output TTEs (*NbInputs* and *NbOutputs*) and the number of parameters (*NbParameters*). They are not directly specified because otherwise, these operators would be constrained before use. Independent operators also require information on which analysis tools are able to implement them, given by *TargetPlatform*. Consequently, it is possible to produce indications for the implementation of an IAP in a specific tool.

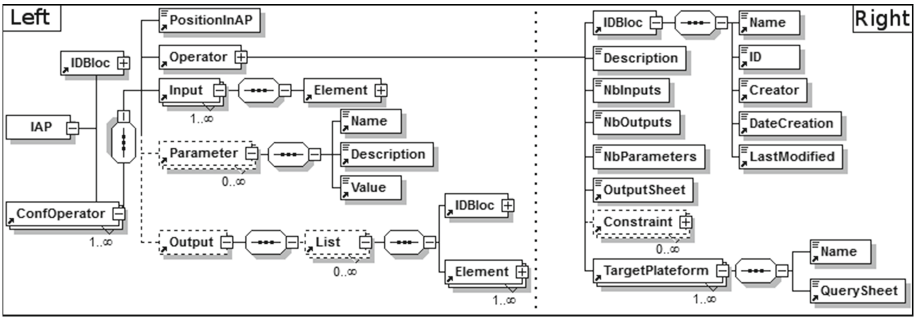



**Fig. 2.** Meta-model of IAP at the left and independent operator at the right.

Secondly, the IAP meta-model shown in Fig. 2, left part, describes an analysis process. It respects the A3 assumption stating that an analysis is a non linear combination of operators with *ConfiguredOperator* which is an ordered step in the IAP: a triplet *(Inputs, Operation, Outputs)*. The inputs are TTEs that will be processed by an independent operator, producing eventually some output TTEs according to the rules of its *OutputSheet*. Then outputs of such operators can be used as the inputs of other ones. This chaining is guaranteed by the partial order property *PositionAP*. Consequently, a *ConfiguredOperator* is reflexive, antisymmetric and transitive, enabling to reliably capitalize an analysis. Hence, a relationship between the expected knowledge and initial TTEs is set up. Moreover, an IAP can be entirely used in another one if the knowledge produced by the first one fits initial TTEs requirements of the second one. This combination offers great perspectives about conception of new capitalized analysis processes.

## 5 Discussion and Future Works

We have implemented our approach in a web-based prototype[5] to test its viability through experimentations with 6 subjects used to work in the TEL domain. Results strongly encourage that analysis processes can be described as suggested in this paper. However, experimentation also shows that there is a lack of semantic power concerning TTEs and a lack of feedback available during the description.

Our future works will focus about how supporting actors of the analysis procedure, using capitalized analysis processes. Firstly our efforts will be focused on driving a meaningful description of analysis processes, from a TEL point of view, such as which elements are able to discriminate analysis processes and enrich them. This will help to establish an effective and informative warehouse of IAPs. Secondly, we aim to enable the reuse of these independent analysis processes according to analysis tools and traced data. We assume that analyses will then be more accessible with more supports and insights (e.g. producing relevant instructions). Furthermore this reuse of IAPs can lead to interesting interoperability perspectives between the analysis tools available in the community.

**Acknowledgement.** This work has been supported by the HUBBLE project (ANR-14-CE24-0015).

## References

1. Baker, B.M.: A conceptual framework for making knowledge actionable through capital formation. doctorial dissertation, University of Maryland (2007)
2. Baker, R.S., Yacef, K.: The state of educational data mining in 2009: a review and future visions. JEDM **1**(1), 3–17 (2009)
3. Choquet, C., Iksal, S.: Usage tracking language: a meta language for modelling tracks in tel systems. In: Proceedings of ICSOFT 2006, pp. 133–138. INSTICC (2006)
4. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI Magazine **17**(3), 37 (1996)
5. Jeong, H., Biswas, G.: Mining student behavior models in learning-by-teaching environments. In: Proceedings of the 1st International Conference on Eucational Data Mining, pp. 127–136 (2008)

[5] Experimentation materials at: http://liris.cnrs.fr/~alebis/iogap.html.

6. Mandran, N., Ortega, M., Luengo, V., Bouhineau, D.: DOP8: merging both data and analysis operators life cycles for technology enhanced learning. In: Proceedings of the LAK 2015, pp. 213–217. ACM (2015)
7. Rosch, E.H.: Natural categories. Cogn. Psychol. **4**(3), 328–350 (1973)
8. Stamper, J.C., Koedinger, K.R., Baker, R.S.J., Skogsholm, A., Leber, B., Demi, S., Yu, S., Spencer, D.: Managing the educational dataset lifecycle with datashop. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS, vol. 6738, pp. 557–559. Springer, Heidelberg (2011)
9. Tsantis, L., Castellani, J.: Enhancing learning environments through solution-based knowledge discovery tools: forecasting for self-perpetuating systemic reform. J. Spec. Educ. Technol. **16**(4), 39 (2001)
10. Volle, M., Malinvaud, E.: Le métier de statisticien. Economica (1984)

# Improving Usage of Learning Designs by Teachers: A Set of Concepts for Well-Defined Problem Resolution

Anne Lejeune[1,2(✉)], Viviane Guéraud[1,2], and Nadine Mandran[1,2]

[1] Univ. Grenoble Alpes, Grenoble, France
[2] CNRS, LIG-MeTAH, Grenoble, France
{anne.lejeune,viviane.gueraud,nadine.mandran}@imag.fr

**Abstract.** Appropriating a learning design remains difficult for teachers. Our research works directed by "co-design learning with teachers", led us to propose a generic set of concepts for situations engaging learners to solve a well-defined problem. It puts forward the benefits of integrating into the learning design both observables on the effective learning situation and teachers' professional knowledge explicating the learning design components. Implementation in FORMID is further discussed.

**Keywords:** Learning design · Teachers · Share and re-use · Problem-solving · Professional knowledge

## 1 Introduction

Despite the great amount of learning design initiatives, there is definitely a long way toward effective share and re-use of learning designs as it was theoretically advocated is still current, as reflected through [1, 2].

The term "learning design" has various acceptations (Dobozy, E. cited in [1 p. xiii]). In this paper we use the term "scenario" in place of "learning design" for referring to the result of the design process, that is, the machine readable representation of a TEL activity as it has be defined in [3]. Consequently, we will use the term "scenario-model" for referring to the structured set of concepts in compliance of which a scenario is modelled (i.e. the educational modelling language).

Our proposal attempts to make scenarios more familiar for teachers by tackling the obstacles under a conceptual perspective. It addresses the concepts which should compose a machine readable scenario-model with the aim of facilitating teachers understanding, reusing and monitoring of scenarios which engage learners to solve individually a well-defined problem. This contribution builds upon our previous research and development works, among those, the FORMID project which is a web-based environment - developed in collaboration with teachers - for the design, operationalization and monitoring of TEL activities [4, 5].

This paper is as follows: we first browse what generally miss in scenarios which hampers their appropriation by teachers before focusing on the particular case of scenarios engaging learners in solving a well-defined problem; next we propose the PROF-K set of concepts, which should be included in a scenario-model for answering


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 404–409, 2016.
DOI: 10.1007/978-3-319-45153-4_34

appropriation issues; an instantiation and implementation in the context of the FORMID project is then presented; before concluding, the FORMID instantiation is evaluated following a Design-Based Research methodology [6].

## 2 Rationale

When teachers prepare a learning activity, they are used to drawing on their discipline knowledge as well as on the curricula and also on their consciousness of their learners' abilities or difficulties. Considering the central role that teachers should keep, it is inconceivable to bypass their needs for planning or monitoring TEL scenarios.

The planning of any type of learning activity mobilizes the teachers' professional knowledge from the most generic level, such as curriculum, pedagogy, to the most specific one, i.e. classroom/student level [7]. The scenario is thereby supposed to convey this teaching professional knowledge, which, for purposes of sharing and reusing, must be described with commonly shared terms, and connected to real and contextualized practices. Studies revealed that teachers found difficulties to identify relevant scenarios in a specific subject matter due to: the lack of descriptions of the studied knowledge, the lack of references to the concerned program and the lack of elicitation of the designers' intentions or else of the underlying pedagogical strategy [9]. The fact remains that the practitioners' expertise is more often described apart from the scenario [8].

Monitoring TEL activities implies that teachers are aware of the on-going pedagogical activities for inferring the learning process [5, 10]. Awareness' means are built on traces collected throughout the learning session. The main approach consists in collecting, transforming and displaying the activity traces in order to provide educational practitioners with information on the on-going activity [11]. The FORMID project approach stands out by including in the scenario-model observables which allow the teacher-designer to define what he wants to observe [4, 5, 12]. The traces harvested at runtime are thereby naturally meaningful.

In the case of activities leading to solve a well-defined problem, the scenarios must be described with a fine level of granularity. In the light of the above, we hypothesize that for helping teachers to appropriate such scenarios, the scenario-model should contain formal concepts for capturing the teachers' expertise and their observation needs.

## 3 Prof-K

Modelling a scenario for well-defined problem-resolution supposes to formally describe the problem (P), its resolution (R) and the desired level of guidance and feedbacks (F). PROF-K adds two requirements: (1) formally modelling observables (O) on the on-going activity and (2) include concepts formalising the professional knowledge (K) which grounds the P/R/O/F elements.

The problem (P) is the "calling card" of the scenario which should allow teachers to clearly identify if this scenario is suitable for its students. To that end, the following concepts should be modelled: the educational system for which the problem makes sense, the instruction level at which the problem is studied, the discipline and the sub-discipline, the general topic, the problem statement and the knowledge at stake. Last, the different resources made available to learners should be informed by the scenario-designer about their role and their relevance in the context of this scenario.

Different methods can be applied to solve most of the problems, among them some are expected at a given educational context, others not. A method is also at the origin of potential recurrent errors or misconceptions acknowledged by teachers. Resolution (R) should model the expected solution(s) of the problem, the expected/alternative method(s) and the common errors with reference to the educational context.

An observable (O) can be considered as a predefined fine-grained mean of awareness on the on-going activity, whose detection mechanisms are formally described. Its modelling requires teaching proficiency for explicating its intent in order to make it understandable by other teachers. It can be classified into quantitative categories (e.g. duration measures, common error) or qualitative ones (e.g. invalid usage of a law).

The term Feedback (F) covers all automated notices sent to learners according their progress throughout the on-going activity. Different types of feedbacks can be modelled (e.g. simple guidance messages, additional clues). Feedbacks deliverance should be explained by the scenario designer according to the desired level of scaffolding, and should be classified by level of guidance and reified by a content description (Fig. 1).

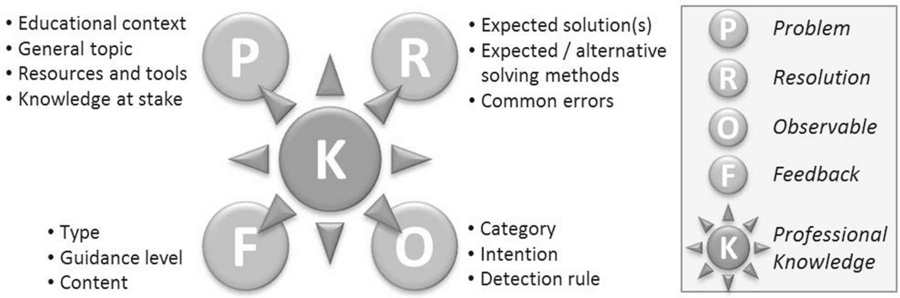



**Fig. 1.** PROF-K: a set of concepts for well-defined problem scenarios

## 4 Instantiation in FORMID

FORMID [4, 5, 12] is based both on authoring principles and a learning design approach and composed of three web-based complementary tools covering design, execution and monitoring of scenarios engaging learners to individually solve well-defined problems by manipulating external solving devices (e.g. simulation, micro-world, tangible device).

The machine readable model for each FORMID scenario is described as follows. Each scenario concerns a *disciplinary field* to which belong the problems to be solved; the *land* of the educational system; the *instructional level;* the general *topic* of the problems; the *design year*; the *external problem-device* used (**P** elements).

Each scenario is divided into one or more *steps*. A step corresponds to a problem to be solved. For each step a scenario is composed of the following elements*: the step instructions* (wording and other advices); the *state* in which the problem-solving device has to be when a learner starts the step; the *validation rule* (a logical expression whose evaluation is triggered by a learner validation request); the *noteworthy situations* (logical expressions which are continuously evaluated at runtime); the *feedbacks* intended for learners (messages displayed in the learner interface either when he or her requests the current step validation, or when a noteworthy situation is detected).

Applied to the proposed set of concepts, *step instructions* and *initial state* of the problem-solving device belong also to the **P** group. The *validation rule* and the *noteworthy situations* which may reflect a valid solving method represent the problem resolution **R**. All *noteworthy situations* are **O** elements as well as the current state of the problem-solving device when a learner requests the step validation. In the **F** class, we find *feedbacks* delivered when a noteworthy situation is detected and the success or failure messages.

The **K** of the proposed generic model is distributed among these different elements. It is formalized by concepts derived from the Anthropological Theory of the Didactic (ATD), which originates from the didactics of mathematics [13]. ATD studies the human activities and their interrelations with knowledge inside the institution in which they take place. It provides a general epistemological model in terms of praxeologies (T, $\tau$, $\theta$, $\Theta$) whose the components are: a *type of tasks* (T) - what one has to do, a *technique* ($\tau$) - a way to achieve tasks of the given T type, a *technology* ($\theta$) - a discourse used to describe, justify, explain, and produce the technique $\tau$ and a *theory* ($\Theta$) - a discourse that plays the same role towards technology that technology does towards technique.

In the FORMID scenario-model, relationships are established between a scenario step (i.e. a problem) and a "typical task" that instantiates a type of task T; this typical task is related to its expected "solving method" instantiating $\tau$ (recursively composed of type of tasks); a solving method is related to "knowledge at stake" elements instantiating $\theta$. The typical task description is completed by "usual mistakes" which instantiate an erroneous technique or technology known as commonly applied by students according to the educational context. The noteworthy situations intended to reflect common errors are linked to one or several "usual mistakes", each of them relating to the typical tasks susceptible to be at their origin.

Every knowledge characterization is available from an external database informed by educational experts in the scenario discipline.

## 5 Evaluation of the Instantiation

The experimental method is based on Design-Based Research as defined in [6, pp. 6–7] and more specifically, we followed a process based on iterative cycles of analysis, design, implementation and redesign. In this research context, it is necessary to combine different approaches. Qualitative approach is deployed to explore and understand the ground and quantitative approach is deployed to quantify results that qualitative approach has identified.

Our first objective was to assess whether the teachers monitoring FORMID sessions were building upon the observables their inferences about the students' work. The second one was to assess the relevance of ATD representations for capturing the implicit teachers' expertise. Three experiments have been conducted.

The first one involved teachers who were the designers of a scenario to be monitored. The second one involved teachers who were not the designers but who benefited of its rationale provided in a paper form. For these two experiments, we used the method of users' tests. Teachers were previously interviewed about their monitoring practices and all their thinks aloud where registered during their use of the FORMID monitoring tool in order to carry out a diagnosis of the students' progress. At the end, teachers were interviewed and their speeches were transcribed. Then we classified the verbatim (parts of the speeches) into three categories: (1) "Declarative verbatim", which corresponds to a simple observation of what a learner or a group of learners is doing; (2) "Interpretative verbatim", which corresponds to the representations of how a learner or a group of learners works; (3) "Diagnosis verbatim", which expresses inferences made by teachers on the domain-knowledge mobilized by a learner at a given time of the activity. Equivalent results can be drawn from these experiments. All categories are represented, which indicates that, thanks to the observables described in the scenario, tutors are able to describe the students' work, to represent themselves their progression and to diagnose the mobilized knowledge, in the case of a single student or of a group of students.

For the last experiment, we used a "focus group" method, which is well suited to confront different points of view. The focus group involved experienced teachers who participated to the FORMID project since its starting. Teachers were first given a summary of the different praxeological concepts (cf. 4, p. 4), and then they discussed among themselves the relevance of these concepts to model the professional knowledge they were accustomed to express in their paper documentation. Finally, we debated with the involved teachers the appropriate use and naming of these concepts with a view to their inclusion in the scenario-model. The conclusions of the focus-group confirmed the adequacy of the ATD concepts for representing the teachers' professional knowledge as described in Sect. 4.

## 6 Conclusion

That teachers might still be under-considered in regards to learning design solutions is for us a real waste. In this paper we tried to contribute to a better taking into account of their teaching knowledge and practices.

After studying what could, from a conceptual point of view, hamper their appropriation of existing scenarios for purposes of share, reuse and monitoring, we proposed a set of concepts enabling teachers to better understand an existing scenario in the case of specific activities engaging learners to solve a well-defined problem.

Our proposal has been instantiated upon the FORMID scenario-model and results of the experiments confirmed our hypotheses. From now the authoring tool has been enhanced and provides the teachers-designers with means for integrate shared professional knowledge information in the scenarios they design.

## References


1. Maina, M., Craft, B., Mor, Y: The Art and Science of Learning Design, p. 252 Sense Publishers, Rotterdam (2015)
2. McKenney, S.: Toward relevant and usable TEL research. In: Maina, M., Craft, B., Mor, Y. (eds.) The Art and Science of Learning Design, pp. 65–74. Sense Publishers, Rotterdam (2015)
3. Koper, R.: Current research in learning design. J. Educ. Technol. Soc. **9**(1), 13–22 (2006)
4. Guéraud, V., Cagnat, J.M.: Automatic semantic activity monitoring of distance learners guided by pedagogical scenarios. In: Nejdl, W., Tochtermann, K. (eds.) Proceedings of the 1st European Conference on Technology Enhanced Learning, EC-TEL 2006 Crete. LNCS, vol. 4277, pp. 476–481. Springer, Heidelberg (2006)
5. Guéraud, V., Adam, J.M., Lejeune, A., Dubois, J.M., Mandran, N.: Teachers need support too: FORMID-observer a flexible environment for supervising simulation-based learning situations. In: Proceedings of Workshop ISEE 2009 in AIED 2009, Brighton, pp. 19–28 (2009)
6. Wang, F., Hannafin, M.J.: Design-based research and technology-enhanced learning environments. Educ. Technol. Res. Dev. **53**(4), 5–23 (2006)
7. Shulman, L.S.: Those who understand: knowledge growth in teaching. Educ. Res. **15**(2), 4–14 (1986). http://www.jstor.org/stable/1175860. Accessed 30 Mar. 2016
8. Warburton, S., Mor, Y.: Configuring narratives, patterns and scenarios. In: Maina, M., Craft, B., Mor, Y. (eds.) The Art and Science of Learning Design, pp. 93–104. Sense Publishers, Rotterdam (2015)
9. Pernin, J.-P., Emin, V., Guéraud, V.: ISiS: an intention-oriented model to help teachers in learning scenarios design. In: Dillenbourg, P., Specht, M. (eds.) EC-TEL 2008. LNCS, vol. 5192, pp. 338–343. Springer, Heidelberg (2008)
10. Gonon, P., Leroux, P.: Designing tutoring activity - an extension of two EMLs, based on an organizational model of tutoring. In: Proceedings of the 10th International Conference on Advanced Leaning Technologies, ICALT 2010, Sousse, pp. 217–221. IEEE Press (2010)
11. Marty, J.C., Carron, T., Pernelle, P.: Observe and react: interactive indicators for monitoring. Int. J. Learn. Technol. **7**(3), 277–296 (2012)
12. Lejeune, A., Guéraud, V.: Embedding observation means into the learning scenario: authoring approach and environment for simulations-based learning. In: Proceedings of the 12th, 9th International Conference on Advanced Leaning technologies, ICALT 2012, Rome, pp. 273–275. IEEE Press (2012)
13. Chevallard, Y.: Readjusting didactics to a changing epistemology. Eur. Educ. Res. J. **6**(2), 131–134 (2007)

# Immersion and Persistence: Improving Learners' Engagement in Authentic Learning Situations

Guillaume Loup[1(✉)], Audrey Serna[2], Sébastien Iksal[1], and Sébastien George[1]

[1] Université Bretagne Loire, Université Du Maine, EA 4023, LIUM, 72085 Le Mans, France
{guillaume.loup,sebastien.iksal,sebastien.george}@univ-lemans.fr
[2] Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR 5205, 69621 Lyon, France
audrey.serna@insa-lyon.fr

**Abstract.** According to the recent technological advances, a new type of digital learning games has emerged. These games integrate virtual worlds persistence and immersion devices allowing the learners to experience more authentic and rich situations. Several studies highlighted their pedagogical value, knowledge transfer and learners' engaged-behaviors. In this paper, we draw the characteristics of these learning games based on the integration of new technologies according to two characteristics: immersion and persistence. To investigate the impact of such technological components, we developed a game and evaluated it in ecological conditions. Four groups of fifteen high school students played the game through two testing conditions: two groups used a prototype allowing only classical interactions limited on usual devices, while the two other groups used a prototype integrating persistent and immersive interactions using Oculus Rift vision. All the interactions were recorded and their analysis suggests more engaged behaviors from students using the immersive and persistent prototype.

**Keywords:** Immersion · Persistence · Serious game · Mixed reality · Engaged behaviors · Trace-based analysis · Digital epistemic game

## 1 Introduction

For many years, teaching tools known under the term "epistemic games" [1] were experienced. More recently, the digital age has led to the evolution of these epistemic games toward "Digital Epistemic Games" (DEG) [2], but too limited to conventional techniques such as Web, e-mail or video conferencing [3]. In the meantime, new types of games known as "pervasive" (or, augmented games, mixed reality games or mobile computing games) have succeeded in fully exploiting advanced technologies. These pervasive games have opened up new perspectives in the field of education, increasing the number of stimuli by both a physical experience in reality and a social and immersive experience, and then improving motivation.

In this paper, we investigate this new concept of pervasive digital epistemic game where technologies can favor authentic situations using immersive and persistent elements. After exposing related work both on digital epistemic games and pervasive


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 410–415, 2016.
DOI: 10.1007/978-3-319-45153-4_35

technologies, we present our study aiming at understanding the impact of immersion and persistence on the learners' engagement.

## 2 Related Work

### 2.1 Digital Epistemic Games

Digital epistemic games belong to "serious games" or "learning games" category because they promote learning. A serious game can be considered as a digital epistemic game if it:

– proposes the solving of non-deterministic problems [4], as in Clim@ction[1],
– concerns the solving complex problems [5], as in Digital Zoo[2],
– relies on multidisciplinary activities [6], as in Urban Science[3],
– supports the learner in a realistic and authentic context [4], as in Clim@ction,
– is based on an "epistemic framework" [1], that is to say when the learner must conduct its business with the skills, methods, knowledge and values of the professional he embodied [7], as in Science.net[4].

These characteristics have the ambition to offer to the student both practical and theoretical learning in order to be able to mobilize and transfer knowledge and skills in many situations [5].

### 2.2 Pervasive Learning Games

A pervasive game is a game having one or more salient features that expand the contractual magic circle of play socially, spatially or temporally [8].

**Game Environment Expansion by Transmedia Use.** Alternate Reality Games (ARG) aim to offer learners the opportunity to solve collectively various problems while confronting with the real world by exchanging SMS, forum, blog or phone calls, and also adding physical moving [9]. Unlike a mixed reality game that overlays virtual to reality (or vice versa) on the same interface, an ARG allows alternating between sessions in the digital world and game phases requiring actions in the real world under a coherent scenario.

**Spatially Expansion by Immersion.** Mixed Reality (MR) was defined as a continuum that connects the real world and the virtual world [10]. The objective is to enrich a situation based on the real or add realism in a virtual environment. This mix can be achieved using many technologies such as screens, cameras, see-through glasses, mobile interfaces, tactile or tangible interfaces.

---

1 Clim@ction: http://eductice.ens-lyon.fr/EducTice/recherche/jeux/jpael/climaction/2011-2012/.
2 Digital Zoo: http://edgaps.org/gaps/projects/digital-zoo-2/.
3 Urban Science: http://edgaps.org/gaps/projects/urban-science/.
4 Science.net: http://edgaps.org/gaps/projects/science-net/.

Studies on the integration of mixed reality elements into educational applications have highlighted their potential, mainly to improve the anchoring of learning and the positioning of learners in relative authentic situations [11].

**Temporally Expansion by Persistence.** Extending temporal aspects of a game rely on persistence concept, widely used in famous MMORPG[5] such as World of Warcraft or Second Life. This games offer to thousands of people, the possibility to interact together within a virtual world. According to Gonzalez *et al.* [12], a virtual world is defined as a persistent computer-simulated environment allowing large number of users, who are represented by avatars interacting in real-time with each other at the simulated environment.

Applied to TEL, this concept can be used for improving learners' engagement. For instance, the learning virtual world called MMOLE[6] allows the development of simulation activities and stimulate active participation.

# 3 Impact of Immersion and Persistence on Learners' Engagement

## 3.1 Study Design

In this study, we investigate a new concept of games named pervasive digital epistemic games (PDEG). We designed and evaluated a PDEG, in which the learning situation authenticity is enhanced by pervasive technologies, bringing more immersion and persistence.

**Table 1.** Main differences between DEG and PDEG prototypes

| | DEG | PDEG | Pervasive Characteristics |
|---|---|---|---|
| Teaser | Oral presentation | Video on the synopsis at the beginning of the first session | Immersion |
| Programming | Resume from the last position recorded | The program continues to run between 2 sessions | Persistence |
| Planetary rover speed | 2,5 m/s | 0,25 m/s | Immersion |
| Observation Interface | Sonar (2D Interface) | Laser (3D Interface) & stereoscopic vision by Oculus | Immersion |
| Planetary rover report on email adress | None : Need to read the logbook | Email giving the characteristics of the last action at the end of the program on the personal address | Immersion Persistence |

[5] MMORPG: Massively Multiplayer Online Role-Playing Game.
[6] Massively Multilearner Online Learning Environments.

The purpose of our study was to analyze the impact of PDEG in terms of motivation and engagement on learners. To do so, we compared of a group of learners using the DEG prototype with another group of learners using the PDEG prototype. The two prototypes used to program a rover on a new planet, analyze the results and debate.

The experimentation was conducted in a high school in two classes of STI2D[7] formation. 57 students aged between 16 to 18 years. Two groups of 13 and 14 students, achieved the experimentation in pervasive condition (i.e. using the PDEG prototype). Two groups of 15 students each, carried out the experimentation in non-pervasive condition (i.e. using the DEG prototype).

The first prototype with DEG characteristics offers classical WIMP interactions. The second prototype with PDEG characteristics offers a more authentic situation (Table 1).

### 3.2 Results

**Motivation Evaluation.** After the final session, participants were asked to fill a survey made of 25 questions. This survey was composed of three parts. The first allowed the identification, the second was related to the practice and perception of video games, the last one was based on The Situational Motivation Scale [13]. 50 students answered to the questionnaire: 27 for the non-pervasive group, 23 for the pervasive group.

The results of Levene's test allow observing that we have a homogeneous variance and a t-test was performed for each measure. There was no difference in the scores for pervasive ($M = 4.11$, $SD = 1.40$) and non-pervasive groups ($M = 4.70$, $SD = 1.72$) concerning the intrinsic motivation; $t(48) = 1.089$, $p = 0.281$. There was no difference in the scores for pervasive ($M = 3.06$, $SD = 1.21$) and non-pervasive groups ($M = 3.32$, $SD = 1.58$) concerning the autodetermination; $t(48) = 0.632$ $p = 0.531$. There was no difference in the scores for pervasive ($M = 4.94$, $SD = 1.51$) and non-pervasive groups ($M = 4.81$, $SD = 1.52$) concerning the extrinsic motivation; $t(48) = -0.149$, $p = 0.882$. There was no difference in the scores for pervasive ($M = 3.57$, $SD = 0.98$) and non-pervasive groups ($M = 3.35$, $SD = 1.15$) concerning the amotivation; $t(48) = -0.797$, $p = 0.429$.

The focus group conducted after experiment suggested that all the students were already very motivated to use a serious game rather than a traditional session, and the difference between the two prototypes seems to be secondary for them.

**Engagement Evaluation.** In order to convert several high-level indicators related to engagement, a trace generator to record information such as the selection of each menu, the content of the programs implemented and the results of the sensors registered by the rover have been included in the game prototype.

According to Bouvier *et al.* proposition [14], which is mainly based on the Self-Determination Theory [15], we considered three types of engaged-behaviors in our analysis: environmental, self and action engaged-behaviors. To measure each type, we defined 6 indicators. This information was calculated for each group and each learning session. One indicator was directly integrated in the prototypes and the others used logs. We used UTL [16] to process our logs.

[7] STI2D: Science and Technology of Industry and Sustainable Development.

The environmental engagement indicator has shown that the pervasive group used more different data sources in order to evaluate the rover progression.

The self-engaged-behaviors indicator demonstrates better results from learners in the pervasive groups (60.51 % of coverage against 46.77 % for the non-pervasive group).

Finally, action-directed engagement results demonstrate clearly that the pervasive groups want to ensure the validity of their program before submission (more simulations but less submissions) while the non-pervasive groups seem to have a “trial and error” approach (lots of submissions, less simulations).

## 4 Discussion and Conclusion

These first results show that PDEG seems to favor engaged-behaviors of learners during the learning activity in comparison to more classical DEG. Obviously, the data samples being low, the interpretation of these indicators cannot be directly generalized.

The experiment protocol, consisting of two conditions in which different participants used several prototypes can also limit conclusions. However, prototypes were designed to have the same functionalities. We ensured that the integration or the reduction of technologies for immersion and persistence (according to PDEG or DEG prototype) did not affect usefulness and usability of the whole game.

As a conclusion, PDEG offers the ability to add immersion and persistence positively impacting learners’ engagement. We still remain to determine in detail how the distribution of immersion and persistence influences this engagement. However, this question seems as difficult as the one regarding the distribution between pedagogy and playful for the serious game.

**Acknowledgements.** These experiments were made possible by funding from the French National Research Agency (ANR-13-APPR-0001).

## References

1. Shaffer, D.W.: Epistemic frames for epistemic games. Comput. Educ. **46**, 223–234 (2006)
2. Hatfield, D., Shaffer, D.W.: Press play: designing an epistemic game engine for journalism. In: Proceedings of the 7th International Conference on Learning Sciences, pp. 236–242. International Society of the Learning Sciences (2006)
3. Boots, N.K., Strobel, J.: Equipping the designers of the future: best practices of epistemic video game design. Games Cult. (2014)
4. Sanchez, E., Jouneau-Sion, C., Delorme, L., Young, S., Lison, C., Kramar, N.: Fostering epistemic interactions with a digital game. A Case Study about Sustainable Development for Secondary Education
5. Shaffer, D.W., Gee, J.P.: before every child is left behind: how epistemic games can solve the coming crisis in education. In: WCER Working Paper No. 2005-7. Wisconsin Center for Education Research (2005)
6. Salmani Nodoushan, M.A.: The Shaffer-Gee perspective: can epistemic games serve education? Teach. Teach. Educ. **25**, 897–901 (2009)

7. Shaffer, D.W., Hatfield, D., Svarovsky, G.N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A.A., Mislevy, R.: Epistemic network analysis: a prototype for 21st-century assessment of learning. Int. J. Learn. Media. **1**, 33–53 (2009)
8. Montola, M.: Exploring the edge of the magic circle: Defining pervasive games. In: Proceedings of DAC. p. 103 (2005)
9. Kim, J., Lee, E., Thomas, T., Dombrowski, C.: Storytelling in new media: the case of alternate reality games, 2001–2009. In: First Monday, vol. 14 (2009)
10. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. **77**, 1321–1329 (1994)
11. Egenfeldt-Nielsen, S.: Overview of research on the educational use of video games. Digit. Kompet. **1**, 184–213 (2006)
12. González, M.A., Santos, B.S.N., Vargas, A.R., Martín-Gutiérrez, J., Orihuela, A.R.: Virtual worlds. opportunities and challenges in the 21st century. Procedia Comput. Sci. **25**, 330–337 (2013)
13. Guay, F., Vallerand, R.J., Blanchard, C.: On the assessment of situational intrinsic and extrinsic motivation: the situational motivation scale (SIMS). Motiv. Emot. **24**, 175–213 (2000)
14. Bouvier, P., Sehaba, K., Lavoué, E.: A trace-based approach to identifying users' engagement and qualifying their engaged-behaviours in interactive systems: application to a social game. User Model. User-Adapt. Interact. **24**, 413–451 (2014)
15. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. **55**, 68–78 (2000)
16. Iksal, S.: A declarative and operationalized language for learning systems analysis. In: UTL (2011)

# STI-DICO: A Web-Based ITS for Fostering Dictionary Skills and Knowledge

Alexandra Luccioni[1(✉)], Jacqueline Bourdeau[2], Jean Massardi[1], and Roger Nkambou[1]

[1] Université du Québec à Montréal, Montréal, Canada
alexandra.vorobyova@gmail.com
[2] Télé-Université, Montréal, Canada

**Abstract.** A major issue in introducing new technological tools in the classroom is that the teachers who are meant to use them often do not receive the necessary training. This is the case of electronic dictionaries, which are seldom used by both students and teachers, despite their benefits for improving vocabulary development and academic achievement [14]. We propose to address this issue with STI-DICO, an Intelligent Tutoring System (ITS) to help French teachers-in-training acquire both the linguistic knowledge and the practical skills needed to successfully use electronic dictionaries [7]. ITS are advanced intelligent learning environments aiming at providing learners with adaptive tutoring services, relying on a cognitive diagnostic to adapt to learners' knowledge states at each step of the learning process, based on a formal modeling of the knowledge domain [11]. In this paper, we describe our design-based approach to STI-DICO, the first iterations of which have resulted in the development of a repository of linguistic and meta-linguistic skills, paired with an ontology of lexical concepts and supported by a series of authentic learning activities, all created with the active participation of experts in the field.

**Keywords:** Intelligent tutoring systems · Linguistic skills · Electronic dictionaries · French language learning · Knowledge representation · Cognitive diagnostic

## 1 Introduction

Dictionaries play an important role in vocabulary development, which has been shown to be a key indicator of academic achievement [14]. In recent years, electronic dictionaries have emerged, offering new functionalities, search functions and a dynamic interface, and resulting in a paradigm shift that has fundamentally changed the process of dictionary use [5]. Furthermore, the ability to use an electronic dictionary has been defined as a top-priority skill by the Ministry of Education of Québec at both primary and secondary school levels [8]; despite this, studies have shown that electronic dictionaries are seldom used by both students and teachers, mostly due to the fact that neither group has received the proper instruction [6, 7].

To address this lack of dictionary skills, a widespread opinion among the authors in the domain is the need to teach these skills explicitly [15, 18]. This involves targeting


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 416–421, 2016.
DOI: 10.1007/978-3-319-45153-4_36

both the practical skills mobilized during dictionary use (e.g. deciding on the appropriate form of the look-up item) as well as the underlying theoretical knowledge (e.g. recognizing a word's antonym). However, dictionary training is far from straightforward, with the risk of proposing rote learning exercises involving a single skill and/or a single dictionary, which is useful in a limited application context, but does not foster the development of far-reaching linguistic and cognitive skills, and often does not help learners in applying the mastered skills in real-life situations [3, 7].

Our project aims to create STI-DICO, an Intelligent Tutoring System (ITS) targeting the new generation of teachers to equip them with the practical skills and theoretical knowledge they need for an appropriate use of electronic dictionaries in the classroom. To carry out this project, we have adopted an iterative methodology, Design-Based Research (DBR) [13], each iteration bringing both progress in system and in terms of theoretical knowledge. The iterations of this project are the following: (1) providing a repository of core dictionary skills and knowledge based on existing studies on dictionary usage, supported by an ontology of lexical concepts; (2) developing a series of situated learning tasks, linking each task with the skills it targets; (3) evaluating the tasks via a Think Aloud protocol [2] to determine different learner profiles and learning paths; (4) developing the STI-DICO interface using Open edX, an E-learning platform, supported by adaptive back-end components; and (5) evaluating STI-DICO with future French teachers to validate its performance.

This short paper describes some of the preliminary results of our iterative approach. It is organized as follows: Sect. 2 presents our unique dictionary skill repository and its evaluation by experts in linguistics. Section 3 describes the authentic learning tasks we have developed and their empirical testing. Section 4 describes the nature and architecture of STI-DICO, as well as further steps in its development.

## 2 Representing Dictionary Use

The successful consultation of a dictionary is a complex process requiring the simultaneous mobilization of multiple skills and concepts, the entirety of which has yet to be described. While studies exist regarding the steps of effective dictionary consultation and the skills needed [5, 9], as well as regarding the underlying cognitive reasons behind consultation errors [10, 18], there has not been, to our knowledge, a complete representation of the practical skills and theoretical concepts needed for successful dictionary consultation. This gap was therefore the heart of our project – we started with the creation of a comprehensive repository of all of the skills and knowledge mobilized during dictionary consultation.

In order to ensure our correspondence with the educational context in which we are situated, we started the repository creation process with an overview of the requirements of the Ministry of Education with regards to both teachers and students [8]. We coupled this with an in-depth analysis of existing studies in dictionary usage [5, 6, 9, 10, 14, 18], drawing parallels between dictionary consultation steps and the skills they solicit. We cross-referenced this initial repository with GTN, an ontology of lexical knowledge [16], allowing us to anchor the skills and steps involved in dictionary consultation using lexical concepts from the ontology.

The skill repository we created is composed of 125 skills and knowledge items, each linked to one or several of 25 lexical concepts extracted from the GTN. It is composed of a series of interconnected databases representing a different level of knowledge, starting from the concepts taken from the GTN ontology, each linked with its corresponding lexical knowledge, lexical skills, and dictionary skills. The research methodology that we have chosen, DBR, emphasizes the collaboration with practitioners from the domain in order to ensure the cohesion of the research project and its application context [13]. We therefore evaluated the totality of our repository with three experts from the fields of linguistics, lexicology, and didactics. The results of the evaluation were very encouraging and the data processing of the evaluators' suggestions enabled us to improve our definitions and add new skills. Furthermore, suggestions given by one of our evaluators led us to restructure the repository to emphasize the link between dictionary skills and the situations that use them resulting in the creation of sets of authentic learning tasks aimed at fostering these skills, which we describe in Sect. 3.

## 3 Authentic Learning Tasks in STI-DICO

In terms of learning activity design, we adhere to the authentic learning paradigm, which advocates the development of learning activities and situations with strong links to learners' everyday contexts, thereby supporting them in applying the skills acquired when needed [3]. Since our target learners are future French teachers, we returned to analyzing the Ministry of Education documents [8] to identify tasks involving dictionary consultation and separated them into 4 types of tasks: (1) reading, (2) writing, (3) text improvement and (4) text correction. We then indexed each of the tasks identified with the skills and knowledge from our repository that we believe are mobilized during the task, thereby creating holistic representations of each task and its linguistic foundations and dictionary skills, covering various contexts of dictionary usage and consultation.

While the tasks that we have selected are based on ministerial documents and correspond to authentic situations that our learners will face in the classroom, it is essential within the DBR methodology to validate the links that we have established between the tasks and the skills. Since these tasks are mostly carried out "behind closed doors", i.e. silently during the reading or writing process, we designed an evaluation using a Think Aloud protocol [2] to empirically validate the skills and concepts that the tasks mobilize. This experimentation is a novel way of examining the process of dictionary consultation, inspired by existing studies in dictionary consultation which asked participants to identify steps they followed post-hoc [10, 18]. But it is the first time, to our knowledge, that a variety of tasks requiring dictionary consultation are tested with a think aloud protocol, granting us an unprecedented view into the cognitive processes behind dictionary consultation.

In order to represent a variety of learner levels, we selected 6 participants, separating them into 3 groups (novice, intermediate and advanced) based on a pre-experiment questionnaire regarding dictionary usage. Subsequently, we asked each participant to carry out 7 dictionary consultation tasks while verbalizing their thought processes and actions. During the experiment, we recorded audio data of participants' verbalizations, synchronized with screen recordings of their actions, as well as a post-experiment interview to further elucidate their cognitive processes.

The Think Aloud experimentation was completed in June 2016, and the transcription and encoding of the results of the recordings is currently underway. Following results analysis, we will verify the indexation of the skills and tasks to assure its cognitive coherency, comparing the mental processes enumerated by our participants with the theoretical skills and concepts attributed to each task.

In the next iteration of our project, these tasks will be used as the basis for designing the learning activities in STI-DICO, coupling authentic tasks with more theoretical exercises to develop particular fundamental concepts. These activities will be based on existing courses in language didactics and supported by feedback provided by the system. The learning activities will be deployed via a Web-based architecture that uses a learning management platform to deliver content to students. We describe the functional prototype of this architecture in the following section.

## 4 STI-DICO Architecture

Intelligent Tutoring Systems have been successful in raising student performance and have been deployed on a large scale in schools and on the Web for a variety of topics [11]. In recent years, there have been a number of proposals to integrate new technologies and approaches to ITS development, including dividing ITSs into separate services and distributing them across multiple systems and using existing learning environments as ITS interfaces [1, 12]. This provides new opportunities for user adaptation and experimentation, exploiting the popularity of existing tools to gather data and provide tutoring support at a larger scale while enhancing the accessibility of courses that provide adaptive tutoring behavior.

For the prototype of STI-DICO, we consulted with experts from computer science and AIED to implement a modular, Web-based ITS architecture which integrates Open edX, an open-source LMS platform with a back-end tutoring architecture using the LTI (Learning Tools Interoperability) standard [4] (see Fig. 1). We based our architecture on that developed for a pilot project which illustrated the feasibility of connecting an ITS back-end with an Open edX front end [1]. We chose this architecture because it enables us to create custom JavaScript problems for more complex our learning activities and to utilize the existing Open edX exercise templates for more simple exercises, all the while providing us with a high degree of freedom in the creation and evaluation of our ITS [7].

In order to ensure STI-DICO's adaptability, we have implemented the double-loop adaptation proposed by Van Lehn [17], involving an outer loop that selects the next learning activity based on the learner's knowledge state and an inner loop that determines the behavior of the system within the learning activity [7]. These two rule engines are embedded within our architecture along with the domain module, a formal representation of the skill and knowledge repository described in the previous section, and the student module, represented by a series of databases which store data regarding learning

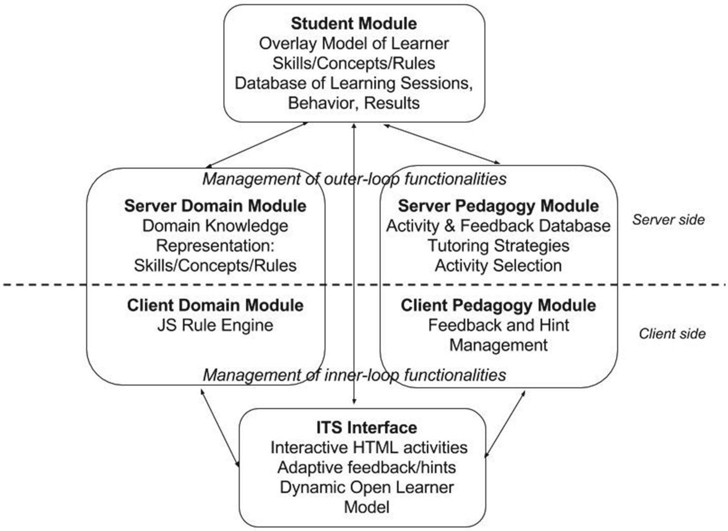



**Fig. 1.** STI-DICO architecture

sessions and learner. We have currently developed a functional prototype of STI-DICO on a small scale, with 20 adaptive activities created using custom HTML templates in Open edX, based on authentic learning tasks, and indexed with the skills and concepts they evaluate (see Fig. 1).

## 5 Conclusion

STI-DICO is an innovative project aimed at creating an ITS to help future French teachers acquire the fundamental linguistic knowledge and practical dictionary skills that they need to meet Ministry standards and to help them transfer this knowledge to their students. Our ITS integrates existing components, such as an ontology of lexical concepts [16], results from empirical research on dictionary use usage [5, 9, 10], and authentic activities in dictionary consultation using an iterative DBR methodology [7]. The repository of the skills and knowledge involved in the dictionary process represents an innovative modeling of this phenomenon, and is at the heart of STI-DICO, enabling the ITS to adapt content and activities to meet the needs of its learners. Several iterations of the project have already been designed and evaluated, with the next iteration of evaluation results expected in mid-2016, in which the results of a Think Aloud protocol will enable us to improve the authenticity of the learning activities provided by our system and assure its pertinence for the target audience [7].

Finally, our implementation approach, integrating a LMS front-end interface with ITS core services, is a promising path for ITS development because it permits the exploitation of the scalability and ease of use of LMS along with the adaptive guidance and tutoring intelligence of ITS. If this integration is carried out successfully, this could provide ITS with a springboard towards their usage on a larger scale both inside and outside of the classroom and for different domains of learning.

## References


1. Aleven, V., et al.: The beginning of a beautiful friendship? Intelligent tutoring systems and MOOCs. In: Conati, C., Heffernan, N., Mitrovic, A., Verdejo, M. (eds.) AIED 2015. LNCS, vol. 9112, pp. 525–528. Springer, Heidelberg (2015)
2. Ericsson, A.K., Simon, H.A.: Protocol Analysi: Verbal Reports As Data. MIT Press, Cambridge (1993)
3. Herrington, J., Oliver, R.: An instructional design framework for authentic learning environments. Educ. Technol. Res. Dev. **48**(3), 23–48 (2000)
4. IMS Global Learning Consortium: Learning Tools Interoperability, http://www.imsglobal.org/lti/ (2012)
5. Lew, R.: From paper to electronic dictionaries: evolving dictionary skills. In: Lexicography and Dictionaries in the Information Age. Selected Papers from the 8th ASIALEX International Conference, pp. 79–84. Airlangga University Press, Surabaya (2013)
6. Lew, R., Galas, K.: Can dictionary skills be taught? The effectiveness of lexicographic training for primary-school-level Polish learners of English. In: Proceedings of the XIII EURALEX International Congress. Universitat Pompeu Fabra, Barcelona (2008)
7. Luccioni, A., Nkambou, R., Massardi, J., Bourdeau, J., Coulombe, C.: STI-DICO: a web-based system for intelligent tutoring of dictionary skills. In: Proceedings of the 25th International Conference on the World Wide Web, International World Wide Web Conferences Steering Committee, pp. 923–928. Republic and Canton of Geneva, Switzerland (2016)
8. Ministère de l'Éducation du Québec: Programme de Formation de L'école Québécoise du Niveau Préscolaire et Primaire. MEQ, Québec (2006)
9. Nesi, H.: The specification of dictionary reference skills in higher education. In: Dans Hartmann, R.R.K. (ed.). Dictionaries in Language Learning. Recommendations, National Reports and Thematic Reports from the Thematic Network Project in the Area of Languages, Sub-Project 9: Dictionaries, pp. 53–67. Freie Universitat Berlin, Berlin (1999)
10. Nesi, H., Haill, R.: A study of dictionary use by international students at a British university. Int. J. Lexicography **15**(4), 277–306 (2002)
11. Nkambou, R., Bourdeau, J., Mizoguchi, R.: Studies in Computational Intelligence. Advances in Intelligent Tutoring Systems, vol. 308. Springer, Heidelberg (2010)
12. Nye, B.D.: AIED is splitting up (Into Services) and the next generation will be all right. In: Proceedings of the Workshops at the 17th International Conference on Artificial Intelligence in Education (2015)
13. Reeves, T., Herrington, J., Oliver, R.: Design research: a socially responsible approach to instructional technology research in higher education. J. Comput. High. Educ. **16**(2), 97–116 (2005)
14. Scott, J., Nagy, B., Flinspach, S.: More than merely words: redefining vocabulary learning in a culturally and linguistically diverse society. In: What Research Has to Say About Vocabulary Instruction, pp. 182–210 (2008)
15. Tremblay, O., Anctil, D., Vorobyova, A.: Utiliser le dictionnaire efficacement: une compétence à développer. Formation et Profession **21**(3), 95–98 (2013)
16. Tremblay, O., Polguère, A.: Une ontologie des savoirs linguistiques au service de la didactique du lexique. In: Actes du Congrès Mondial de Linguistique Française (2014)
17. VanLehn, K.: The behavior of tutoring systems. Int. J. Artif. Intell. Educ. **16**, 227–265 (2006)
18. Wingate, U.: Dictionary use—the need to teach strategies. Lang. Learn. J. **29**(1), 5–11 (2004)

# PyramidApp: Scalable Method Enabling Collaboration in the Classroom

Kalpani Manathunga(✉) and Davinia Hernández-Leo

ICT Department, Universitat Pompeu Fabra, Barcelona, Spain
{kalpani.manathunga,davinia.hernandez}@upf.edu

**Abstract.** Computer Supported Collaborative Learning methods support fruitful social interactions using technological mediation and orchestration. However, studies indicate that most existing CSCL methods have not been applied to large classes, means that they may not scale well or that it's unclear to what extent or with which technological mechanisms scalability could be feasible. This paper introduces and evaluates PyramidApp, implementing a scalable pedagogical method refining Pyramid (aka Snowball) collaborative learning flow pattern. Refinements include rating and discussing to reach upon global consensus. Three different face-to-face classroom situations were used to evaluate different tasks of pyramid interactions. Experiments led to conclude that pyramids can be meaningful with around 20 participants per pyramid of 3–4 levels, with several pyramids running in parallel depending on the classroom size. An underpinning algorithm enabling elastic creation of multiple pyramids, using control timers and triggering flow awareness facilitated scalability, dynamism and overall user satisfaction in the experience.

**Keywords:** Computer-Supported collaborative learning · Pyramid/snowball collaborative learning flow pattern · Large groups · Classroom

## 1 Introduction

Multiple findings from educational research highlight the importance of active learning [1]. In particular, sound collaborative learning methods foster rich social interactions between students leading to fruitful learning. Provision of technological means to support collaboration has enabled new or enhanced learning scenarios [2]. Technologies can mediate social interactions; facilitate orchestration regarding coordination requirements (e.g., group distribution); monitor interactions for regulation. Yet, despite the potential technologies, effective CSCL methods that favour equal, meaningful interactions between students -sometimes referred as macro-scripts [3, 4]-, have been mostly applied upon small groups of students [5].

Recently, popularity and social impact of open educational settings such as Massive Open Online Courses (MOOCs) have driven more research interests around scalable pedagogies [6] and urge to build up pedagogical methods based on active learning approaches fostering social interactions [7]. Unstructured discussion through forums and social media helps [7], but its potential effectiveness is limited compared to what can be achieved by more structured CSCL approaches [3]. The need for active learning


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 422–427, 2016.
DOI: 10.1007/978-3-319-45153-4_37

in large classroom settings has been acknowledged for over three decades [8]. However, actual teaching practice in large classrooms is still broadly based on lecturing with passive participation of students. Only few remarkable initiatives have offered technological solutions to facilitate active learning in large classroom, based on collective polls or self-organized backstage interactions [9, 10]. However, there are no approaches that extrapolate sound macro-scripts methods that structure the collaborative learning flow for effectiveness in terms of fostering individual accountability, positive interdependence and meaningful interactions between students [3, 4].

Direct application of collaborative learning methods that work well with small groups appears to be challenging in both massive virtual learning contexts and large synchronous classes due to lack of scalable aspects or practical challenges hindering sensible implementation of CSCL methods. Practical challenges include time and teachers' load limitations if learning outcomes of all groups should be measured in a large classroom or implications of flexibility issues with large and varying number of students [2]. This research work aims at exploring to what extent or with which technological mechanisms, scalability of relevant collaborative learning methods could be feasible. In particular, the paper studies the Pyramid (a.k.a. Snowball) method, which intuitively suggest reasonable scalability potential. Pyramid method, formulated as a Collaborative Learning Flow Pattern (Pyramid CLFP), can be particularized and reused with multiple epistemic tasks and educational levels [4]. A Pyramid flow starts with individual proposals being discussed and iteratively joined into larger-groups till a common consensus is reached upon at the global level. Such scenarios foster individual participation and accountability (equal opportunity for all, yet with singular contributions) and balanced positive interactions (opinions of all members count) in a collaborative knowledge-oriented negotiation process. The specific research question that guides this research is how can the Pyramid flow pattern be technologically supported to serve as a scalable method for collaboration in the classroom?

To tackle the question, we have iteratively proposed, prototyped and evaluated refinements of the Pyramid CLFP. Initial refinements propose a method using peer rating promoting integrated consensus reaching accompanied by discussion. The technological implementation of this is "PyramidApp" tool. Main challenges identified at initial iterations referred to *scalability* and *dynamism*. With *scalability* we mean capability to elastically accommodate growing numbers of participants while maintaining pedagogical and practical effectiveness. As *dynamism* we mean the ability to keep activity progression while preserving enthusiasm and usability. To achieve *scalability* and *dynamism*, iterative refinements of the method incorporate an algorithm implementing flow creation, flow control and flow awareness rules. A first evaluation of PyramidApp in three real-class contexts offers insights about its scalability prospects and suitability of the proposed rules.

## 2 PyramidApp Method and Algorithmic Rules

PyramidApp is a technological solution implementing a scalable method applying Pyramid CLFP principles. Individuals propose their option (i.e. can be a question or an answer for a given task, seeking active comprehension) and PyramidApp forms small groups to share ideas about suggested options, to clarify and negotiate before rating the options. Highly rated options are promoted to upper levels and groups grow larger (by joining previous smaller groups) following a pyramid/snowball structure. Rating and discussion propagate till the final level until the complete group reaches upon a global consensus on the best options. Everyone, including the educator, see finally selected options to which the educator comments. In large classes, educators do not have sufficient time to comment each individual's option (questions or answers), but can attend for an agreed selection of options more feasibly. At the same time, all students have the chance to express and discuss their ideas and to critically reflect and assess peer's contributions, with positive benefits in their negotiation skills and knowledge building process. Initial prototypical pyramid implementations were developed and evaluated in rounds to iterate behavioural rules for the algorithm behind the method to address *scalability* and *dynamism* issues (Fig. 1). *Flow creation* rules allow *scalability* by automatically adapting the pyramid structure based on the number of joining participants providing an elastic mechanism of multiple pyramid creation. *Flow control rules* lead to *dynamism* by preventing potential blocking within flow progression if participants leave (due to any reason: unexpected situation or technological problem). Parameter values presented in Fig. 1 are shown with recommended values that are configurable if preferred. Those estimations were acquired from the initial evaluations. For example, it was observed that when number of level increases participants absolve enthusiasm, longer timing values lead to boredom, very high satisfaction percentages may froze pyramid branches or lead to higher waiting time. *Flow awareness* rules (e.g.: progression level, group members, timing notifications and selected options along the flow) elevate learner engagement and usability.

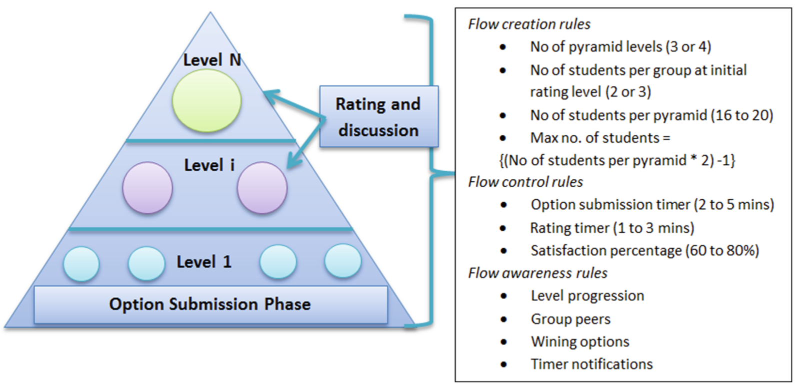


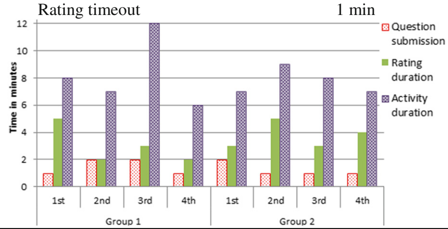

**Fig. 1.** Schema of PyramidApp: parameters for flow creation, control and awareness rules

## 3 Evaluation

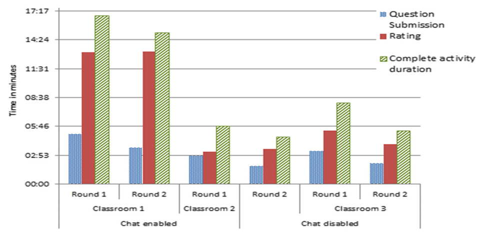
PyramidApp's algorithmic rules were evaluated for potential *scalability* (ability to accommodate an increase in the number of participants) and *dynamism* (ability to keep activity progression) while maintaining pedagogical and practical effectiveness

**Table 1.** Pyramid configurations and timing aspects across three experimental settings

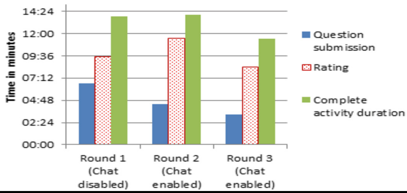
| Experimental setting | | Pyramid configurations | Values | Discussion |
|---|---|---|---|---|
| Higher education setting | 3rd year class<br>(N=23)<br>(1 round) | No of levels<br>Students per group in second level<br>Question submission timeout<br>Rating timeout | 4<br>2<br>3 mins<br>2 mins | Not Available |
| *Task(3d year)*: Propose a potential exam question<br>*Task (1st year) :* Propose questions for peer presentations | 1st year class, two groups<br>(N=22)<br>(8 rounds) | No of levels<br>Students per group in second level<br>Question submission timeout<br>Rating timeout | 3 or 4<br>2<br>2 mins<br>1 min | Not available |




| Experimental setting | | Pyramid configurations | Values | Discussion |
|---|---|---|---|---|
| Secondary school setting<br><br>*Task* : Ask any doubt or question about "HTML" and "Scratch" concepts that had already been taught in the ICT class | Classroom 1<br>(N=21)<br>(2 rounds) | No of levels<br>Students per group in second level<br>Question submission timeout<br>Rating timeout | 4<br>2<br>3 mins<br>2 mins | Enabled |
| | Classroom 2<br>(N=20)<br>(2 rounds) | No of levels<br>Students per group in second level<br>Question submission timeout<br>Rating timeout | 4<br>2<br>3 mins<br>2 mins | Disabled |
| | Classroom 3<br>(N=10)<br>(2 rounds) | No of levels<br>Students per group in second level<br>Question submission timeout<br>Rating timeout | 3<br>2<br>3 mins<br>2 mins | Round1-enabled<br>Round2-disabled |




| Experimental setting | | Pyramid configurations | Values | Discussion |
|---|---|---|---|---|
| Vocational training setting<br><br>*Task :* Propose question about<br>a) future career opportunities,<br>b) curriculum,<br>c) suggest outdoor activity | N=43<br>(3 rounds) | No of levels<br>Students per group in second level<br>No of pyramids<br>Question submission timeout<br>Rating timeout | 4<br>3<br>2<br>3 mins<br>2 mins | Round1-disabled<br>Round2-enabled<br>Round3-enabled |



alongside enthusiasm and usability. The evaluation also seeks which configurations of the method (values for parameters in rules and to use discussion or not) achieve satisfactory scalability, dynamism and overall impact. Across all experimental settings, 80 % satisfaction percentage was maintained and deliberately enabled or disabled chat to observe different behaviour. Table 1 explains the nature of each experimental setting, tasks given and pyramid configuration parameters with values. Graphs illustrate timing aspects across three settings.

Proposed Pyramid refinement satisfactorily accommodates groups of up to 20 students in a single pyramid and several pyramids can run in parallel in large classrooms (two pyramids in the vocational training setting) facilitating late comers to join on-going activities or managing drop-outs. Each Pyramid flow results a single outcome that can be commented by the educator. A classroom of 100 students will have five outcomes, which can be feasibly addressed. Pyramids of 3 or 4 levels can maintain satisfactory engagement. Flow control rules like timers facilitated dynamism by preserving activity progression. Depending on the context, a pyramid activity can take between 5 to 16 min. Table 2 shows structural *scalability* and adapted *dynamism*.

**Table 2.** Dynamism and scalability preservation in Pyramid flows within three settings

<table>
<tr><th>Experiment Setting</th><th>Structural aspects</th><th colspan="4">Comments /Observations</th></tr>
<tr><td>Higher education setting</td><td>Timeouts: 5<br>Late logins: 1</td><td colspan="4">Pyramid flows were not frozen or interrupted thanks to flow control mechanisms (satisfaction percentage and timers) maintaining dynamism and flow creation rules enabling a scalable inclusion of students joining late.</td></tr>
<tr><td>Secondary school Setting</td><td>Late logins: 7</td><td colspan="4">Flow control rules (timers and satisfaction percentage) allowed smooth flows irrespective of multiple timer expirations maintaining dynamism. Students were enthusiastically and participating in discussions and rating.</td></tr>
<tr><td rowspan="3">Vocational training Setting</td><td></td><td>Round 1</td><td>Round 2</td><td>Round 3</td><td rowspan="3">2 pyramids were enacted without interruptions.</td></tr>
<tr><td>Desired pyramid sizes</td><td>20 & 16</td><td>20 & 16</td><td>20 & 16</td></tr>
<tr><td>Final pyramid sizes</td><td>20 & 19</td><td>20 & 27</td><td>20 & 22</td></tr>
</table>

As with any pedagogical method, its effectiveness depended on the context (e.g., classroom atmosphere) and proposed epistemic tasks (e.g., need of active comprehension). Along with dynamism, flow awareness rules contributed to preserve enthusiasm and usability. Rating, viewing winning options and levelling up in pyramid offering gaming effect were perceived with more than 85 % satisfaction across all experiments.

## 4 Conclusion

Diverse educational contexts raise requirements for active pedagogical methods like collaborative learning to be applied with large numbers of students using reasonable time durations. In the paper we identified scalability and dynamism as the key requirements to be addressed by such methods and their technological implementations. We studied a refined implementation (PyramidApp) of the Pyramid flow, addressing these issues, incorporating flow creation, flow control and flow awareness rules. Results suggest suitability of the mechanisms behind the method and open new perspectives that are worth further exploring with diverse epistemic tasks, contexts, larger classroom settings and other challenging settings such as massive open courses.

**Acknowledgements.** Special thanks to participants from Escola Forestal De Sta. Coloma De Farners, Oak House School and Engineering School of UPF Barcelona. This work has been partially funded by the Spanish Ministry of Economy and Competitiveness (TIN2014-53199-C3-3-R; MDM-2015-0502).

## References

1. Bonwell, C.C., Eison, J.A.: Active learning: creating excitement in the classroom. In: ERIC Clearinghouse on Higher Education. The George Washington University, Washington, D.C. (1991)
2. Dillenbourg, P., Tchounikine, P.: Flexibility in macro-scripts for CSCL. J. Comput. Assist. Learn. **23**(1), 1–13 (2007)
3. Dillenbourg, P.: Split where interaction should happen, a model for designing CSCL scripts. In: Gerjets, P., Kirschner, P.A., Elen, J., Joiner, R. (eds.) Instructional design for effective and enjoyable computer-supported learning. Knowledge Media Research Centre, Tuebingen (2004)
4. Hernández-Leo, D., Villasclaras-Fernández, E.D., Asensio-Pérez, J.I., Dimitriadis, Y., Jorrín-Abellán, I.M., Ruiz-Requies, I., Rubia-Avi, B.: COLLAGE: a collaborative learning design editor based on patterns. J. Educ. Technol. Soc. **9**(1), 58–71 (2006)
5. Manathunga, K., Hernández-Leo, D.: Has research on collaborative learning technologies addressed massiveness? J. Educ. Technol. Soc. **18**(4), 357–370 (2015)
6. Ferguson, R., Sharples, M.: Innovative pedagogy at massive scale: teaching and learning in MOOCs. In: Rensing, C., de Freitas, S., Ley, T., Muñoz-Merino, P.J. (eds.) EC-TEL 2014. LNCS, vol. 8719, pp. 98–111. Springer, Heidelberg (2014)
7. Rosé, C.P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., Sherer, J.: Social factors that contribute to attrition in MOOCs. In: Proceedings of the First ACM Conference on Learning @ Scale Conference, pp. 197–198 (2014)
8. Gibbs, G.: Teaching more students: discussion with more students. Headington, Oxford (1992)
9. Herreid, C.F.: “Clicker” cases: introducing case study teaching into large classrooms. J. Coll. Sci. Teach. **36**(2), 43–47 (2006)
10. Gehlen-Baum, V., Pohl, A., Weinberger, A., Bry, F.: Backstage – designing a backchannel for large lectures. In: Ravenscroft, A., Lindstaedt, S., Kloos, C.D., Hernández-Leo, D. (eds.) EC-TEL 2012. LNCS, vol. 7563, pp. 459–464. Springer, Heidelberg (2012)

# From Idea to Reality: Extensive and Executable Modeling Language for Mobile Learning Games

Iza Marfisi-Schottman, Pierre-Yves Gicquel(✉), Aous Karoui, and Sébastien George

Université Bretagne Loire, Université du Maine, LIUM-EA 4023, Le Mans, France
{iza.marfisi,pierre-yves.gicquel,aous.karoui,
sebastien.george}@univ-lemans.fr

**Abstract.** Mobile Learning Games (MLGs) show great potential for education, especially in fields that deal with outdoor learning activities such as archaeology or botany. However, the number of MLGs currently used remains insignificant. This is partly due to the fact that the current authoring tools are based on modeling languages that only allow creating very specific and rigid types of MLGs. In this paper, we therefore propose an extensive modeling language for MLGs. This model was designed, with the help of botanical experts, in order to cover the variety of MLG types they would like for their field trips. This modeling language uses high-level concepts, such as game activities and points of interest on a map that can therefore be used by teachers in any domain. Finally, we discuss how scenarios, described with this language, can be automatically transformed into executable web applications.

**Keywords:** Mobile learning game · Game-Based learning · Serious games · Modeling language · Scenario

## 1 Introduction

Mobile Learning Games (MLGs) have proven their efficiency in various domains of education. *Gaius' Day*, for example, is used by history teachers, during archaeological outings Egnathia in Italie [1]. *Frequency1550* is used to teach high-school students about history in medieval Amsterdam [2] and *Skattjakt* is a MLG to promote activity while visiting Swedish castle [3].

However, even though MLGs show great potential, very few are actually used by teachers. Yet, there are several authoring tools that enable teachers to create their own MLGs without any computer skills. Why aren't these tools more used by teachers?

In this paper, we unravel the mystery by analyzing the needs of a group of botanists, who wish to create MLGs for a natural park. This work is part of the ReVeRIES[1] project, which aims at using mobile technologies to help humans recognize the trees in a fun and motivating way. In the following section of this paper, we present a design experiment where botanist and learning game experts design MLG scenarios. We then use current authoring tools to implement these scenarios. Our

[1] visited in April 2016, visited in April 2016.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 428–433, 2016.
DOI: 10.1007/978-3-319-45153-4_38

conclusion is that state of the art authoring tools do not allow to express these scenarios. From then, we introduce the ReVeRIES modeling language.

# 2 What Needs for Authoring Situated Learning Games?

## 2.1 Designing Mobile Learning Games for a Natural Park

In this section, we try to identify why the existing authoring tools do not seem to be adapted to the needs of teachers and experts who wish to design MLGs. To that end, we organized a creativity session with botanists and learning games experts who wish to create MLGs *Echologia*[2] natural park. This creativity session, described in detail in [4], resulted the production of eight MLGs.

Because of the limited space of the article, we will not describe each of these scenarios, but rather their main characteristics. First of all, the designers described the MLGs as a sequence of **situated game units**. We deliberately chose the term "unit" because the content of each unit formed a coherent set of activities aimed at teaching the **characteristics of a plant**. These game units are also "situated" because they are composed of activities that can only be done in the vicinity of this real plant. As illustrated in Fig. 1, each game unit, represented by a post-it, is physically located on a **map** and linked to **objects of interest** (botanical or non-botanical) alongside the path. Most of the game units were composed of an activity that consist in finding the point of interest (i.e. the plant or a group of plants) and one or several activities that are done once they are on site. Several MLGs also integrated forms **collaboration** between the players of a same group.

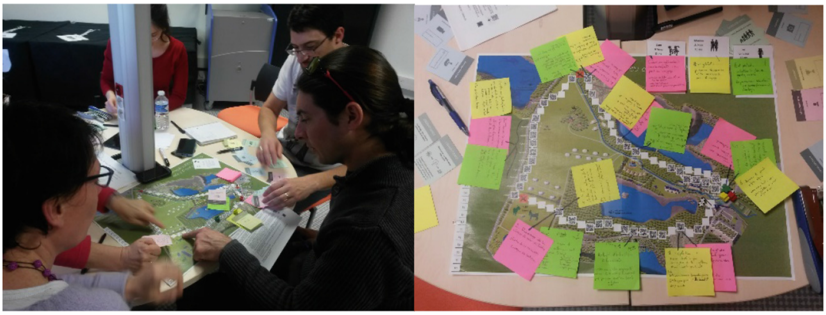


**Fig. 1.** Creativity session, with a map of the park, for designing MLGs

The goal of this creativity session was to determine the types of MLGs fields experts in botany would potentially like to create for their educational outings. It is therefore

[2] visited in April 2016, visited in April 2016.

important to take into consideration the fact that the *Echologia* park is quite particular in the fact that the visitors are forced to walk along a simple path for safety reasons. This physical limitation resulted in the fact that all the MLGs designed for the park have strictly ***linear scenarios***. In other words, the game units are done one after the other. If the MLGs were designed for an open natural park, some of them would probably have ***emergent scenarios*** in which the game units are triggers when the players are physically in the vicinity of the point of interest. This is often the case for MLGs that offer interactive educational walks through cities or archeological sites. Another possibility is to design the scenario as an ***activity hub*** for which all the game units are available from the beginning and the players choose in witch order they want to do them. This is for example the case for *Florex* [5], a paper-based game used in primary school, in with the players are given exercise sheets, in relation to six specific trees in a public park.

### 2.2 Limitations of Existing Authoring Tools

After collecting eight scenarios designed during the brainstorming session, we tried to create these MLGs by using the existing authoring tools. First of all, let us note that we could only find two authoring tool specifically dedicated to MLGs: *Aris* and *ARLearn*. We therefore extended our state of the art to authoring tools designed for mobile games (*FuretFactory*). In the following section, we describe the limitations we encountered with each of these tools when trying to implement the eight scenarios designed by the botanists.

***Aris***[3] [6] is a very rich tool but is extremely complicated to use. The main reason that makes it difficult to create game units is the fact that the model implemented by *Aris* is based on low level items. In other words, the user needs to create a multitude of low level items (text, quests, information plaques, buttons, items for scoring and resources) and link them together with locks and triggers.

***ARLearn***[4] [7] is as complex to use as *Aris*. The authors of this tool actually recommend a "scripting phase" during which the teachers' scenarios are formalized and entered into the authoring tool by computer scientists.

***FuretFactory***[5], on the contrary, is very easy to use. In just a few minutes, the ergonomic user interface allowed to create simple situated activities linked to a point on a map. An important limitation is the fact that the scenario is always linear (i.e. one game unit after another) and do not support collaboration between players.

We can conclude that the existing tools are either not adapted to the type of MLGs teachers want, or too complex to use. In the next section, we propose a high-level modeling language that matches the natural concepts used by teachers. We then explain the work in progress to develop a simple authoring tool, based on this modeling language, which will allow teachers to create their own MLGs.

[3] visited in April 2016, visited in April 2016.
[4] http://portal.ou.nl/web/arlearn/, visited in April 2016.
[5] http://www.furetfactory.com/, visited in April 2016.

## 3 ReVeRIES Model: High-Level Modeling Language for Mobile Learning Games

As we have seen during the analysis of the existing authoring tools, the difficulty is to create a modeling language that offers high-level concepts, similar to the natural concepts used by teachers, and that is rich enough to adapt to various types of MLGs.

In order to do so, we propose the ReVeRIES modeling language, partially depicted in Fig. 2. First of all, this modeling language integrates the concepts naturally used by designers during their MLG creativity session. Indeed, the central concept of the modeling language is the ***SituatedGameUnit*** that is linked to a Point Of Interest (***POI****)* on which the game unit takes place. We can point out that this POI is a circular zone that can contain zero, one or several objects of interest (e.g. plants). The game unit is composed of:

- a trigger that starts the game unit (end of another game unit, when the player asks for it or when the learners is in proximity with a POI*);*
- a clue to help the player find the POI. This clue can contain resources (text, pictures and multimedia), various guidance functionalities (showing POI marker on the map or GPS, beeper or vibration guidance). The teachers can also add extra clues that can be "bought" by the players in exchange for points. It also has a form of validation to determine that the player has arrived at the POI and can therefore start the on-site activities. This validation can be done with GPS (the learner must be in the geographical zone of the POI*),* by scanning a QRcode on the POI or by using an external specific system, such as *Folia*[6], to prove that the learner has found the right tree species;
- a reward for finding the *POI* (items for the inventory or points);
- zero, one or several ***OnSiteActivities*** that are meant to be done at the POI. These activities are composed of resources (text, video, augmented reality, situated chat), tasks (take a photo, collect items, compare objects, describe an object or answer a question). They also contain rewards for succeeding in the activity (item for the inventory or points);
- a pedagogical conclusion that appear when the learner signals the end of the game unit. This is the perfect place for the teacher to underline the knowledge freshly acquired by the learner.

A *MobileLearningGame* is composed of several *SituatedGameUnits.* It also has:

- a type that determines the way these game units follow each other. For a ***linear scenario*** (e.g. treasure hunt), the game units are chained up, one after the other, meaning that the end of a game unit (completion status changed to "finished"), triggers and the *clue* of the next unit. For an ***emergent scenario*** (e.g. interactive walk), the clue for a unit is triggers when the player is physically the vicinity of the point of interest on which the unit takes place. Finally, for an ***activity Hub*** (e.g. geocaching

[6] https://itunes.apple.com/fr/app/folia/, visited in April 2016.

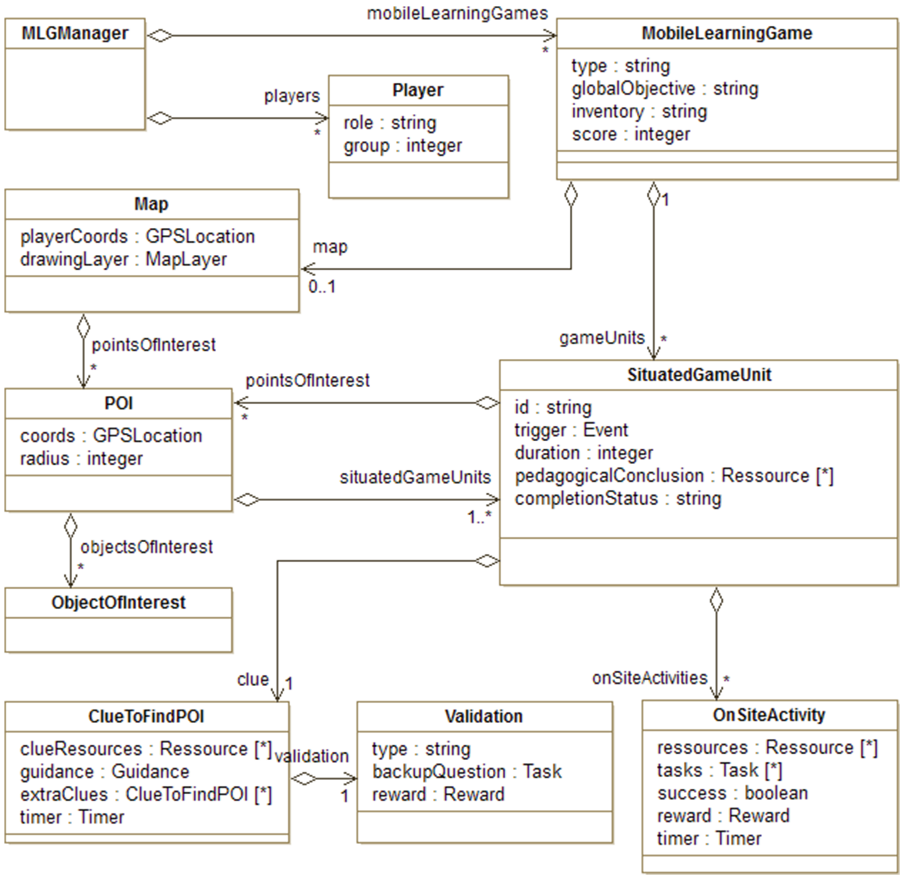



**Fig. 2.** ReVeRIES modeling language for Mobile Learning Games

or *Florex* [5]), all the clues of the units are given at the beginning of the MLG and the players can therefore decide the order in which they want to do them;

- an inventory that contains all the virtual items earned as rewards (e.g. virtual objects, information sheets);
- a score that keeps track of the points earned by the player;
- a *map* of the geographical zone where the MLG will take place and can be viewed by the players at any time. The teacher can also choose to show the player's position and add another map layer in order to have a personalized map (e.g. map of *Echologia* park). The map also contains several points of interests *(POIs)* that can be shown or hidden depending on the time of MLG.

In order to encourage collaborative behavior, the ReVeRIES modeling language also allows teachers to send rewards, earned during the game, to the inventory of another persona of the same group.

## 4 Conclusion

In this paper, we present the ReVeRIES modeling language for creating Mobile Learning Games (MLGs). This language was designed with the help of field experts in botany, but is generic enough to be used in other fields. This language fulfills a gap in the existing MLG authoring tools that are either too complex or too specific to allow teachers to use them.

We transformed each of the classes in the ReVeRIES model into a web component. For the time being, it is possible to create instances of each class by using an html tag. For instance, one can create an *OnSiteActivity* by using the <on-site-activity> tag. Each of these components takes parameters defined by the user, for instance in the case of a MCQ activity, the component takes the questions, possible answers and correct answers as parameters. These parameters are then used to instantiate the components on the web page. We are currently developing an authoring tool prototype that will allow non-computer scientist to create these instances without having to manipulate html. We are now in a phase of internal testing of the prototyping tools, and we plan to test them with field specialists soon. In future work, we will focus on automating the trace generation and processing to obtain feedback on the user activity. We will also work on defining MLG patterns that could provide a basic high-level succession of activity in a learning situation.

**Acknowledgments.** This work is supported by the French National Agency for Research with the reference ANR-15-CE38-0004-01 (ReVeRIES project).

## References

1. Ardito, C., Costabile, M.F., De Angeli, A., Lanzilotti, R.: Enriching archaeological parks with contextual sounds and mobile technology. Comput. Hum. Interact. **19**(4), 1–30 (2012)
2. Huizenga, J., Admiraal, W., Akkerman, S., ten Dam, G.: Mobile game-based learning in secondary education: engagement, motivation and learning in a mobile city game. J. Comput. Assist. Learn. **25**(4), 332–344 (2009)
3. Spikol, D.: Exploring novel learning practices through co-designing mobile games. In: Researching Mobile Learning: Frameworks, Tools, and Research Designs, Peter Lang (2009)
4. Marfisi-Schottman, I., Gicquel, P.-Y., George, S.: Meta serious game: supporting creativity sessions for mobile serious games. In: Proceedings of European Conference on Game Based Learning, ECGBL, Peisly, Scotland (2016, in press)
5. Marzin, P., Triquet, E., Combaz, B.: Recognized trees in primary school: the game situation «Florex»: report of innovation. Concepts Misconceptions (22), 117–136 (2003)
6. Gagnon, D.J.: ARIS. Master's Thesis, University of Wisconsin-Madison (2010)
7. Ternier, S., Klemke, R., Kalz, M., Van Ulzen, P., Specht, M.: ARLearn: augmented reality meets augmented virtuality. J. Univ. Comput. Sci. **18**(15), 2143–2164 (2012)

# Combining Adaptive Learning with Learning Analytics: Precedents and Directions

Anna Mavroudi(✉), Michail Giannakos, and John Krogstie

Norwegian University of Science and Technology, Trondheim, Norway
{anna.mavroudi,michailg,john.krogstie}@idi.ntnu.no

**Abstract.** Adaptive learning and learning analytics are powerful learning tools on their own. Scholars have reported outcomes on combining them, but the lack of a summary from these studies prevents stakeholders from having a clear view of this combination. In this paper, we consider the key dimensions of learning analytics applications in adaptive learning, in order to suggest a proper reference framework that serves as the basis to systematically review the literature. The findings suggest that interesting research work has been carried out during the last years on the topic. Yet, there is a clear lack of studies (a) on school education and in topics outside STEM and (b) that do not focus solely on the (self-)reflection of students or tutors. Finally, the majority of the studies merely concentrates on narrow measures of learning like student performance. A niche area taking into account more complex student behaviors, like collaboration, is emerging.

**Keywords:** Adaptive learning · Learning analytics · Review

## 1 Introduction

A great potential can be seen from the synergy of adaptive learning and Learning Analytics (LA) in Adaptive Learning Analytics (ALA) by illuminating aspects of learning that were previously difficult to observe and, in turn, empowering students to participate in lessons that can be personally adapted. Still, there is a lack of comprehensive reviews on the topic in order to provide a summary of the findings, recommendations and interesting future directions. Another systematic review from 2013 was found in the literature [19]. Yet, the results of the current review validate our initial assumption that work of high quality has been conducted from 2013 onwards on the topic.

ALA can be defined as a subset of LA that focuses on the features and the processes of learning in an adaptive online learning environment, where LA can help to track the progress of the students over time and empower the stakeholders to make well-informed and evidence-based decisions. Although there is an overlap between LA and Educational Data Mining (EDM) a critical analysis of the literature revealed that EDM mostly follows a “bottom-up” approach, whereas LA adopt a “top-down” approach [38].

Adaptive learning can be defined by the triplet: adaptation strategy, adaptation method(s), and adaptation parameter(s). Adaptation method(s) and adaptation parameter(s) pertain to the questions “what will be adapted?” and “to what will it be adapted?”, respectively [28]. The adaptation strategy is built on a set of rules which combine the


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 434–439, 2016.
DOI: 10.1007/978-3-319-45153-4_39

adaptation method(s) with the adaptation parameter(s) by means of a meaningful rationale that caters for the objective(s). Also, adaptation might entail adaptivity and/or adaptability, depending on who has the control of or who takes the initiative to the adaptation, the learner or the system [14]: adaptation is automatically performed by the system, whereas in the latter case it is performed by a human agent, usually the tutor or the student.

## 2 A Reference Framework for Learning Analytics in Adaptive Learning

The suggested framework is a modified version of two previous generic models in LA [5, 10] and caters for the application of LA in the field of adaptive learning. In particular, it is inspired by [10] which considers four dimensions (what, who, why and how), each of them containing several components which are included in one or both of these frameworks. For example, to categorize the objectives, [5] proposes recommendation and reflection whereas [10] proposes reflection and prediction. Consequently, the final model considers all three objectives. The new additions in the proposed framework extend the "how" question to also include the adaptation aspect.

The dimensions of the proposed framework are the following: (a) "What?" (in what context are ALA managed or used?) which pertains to subject, educational technology tools used, institutional context, and constraints, (b) "Who?" (who is targeted by the use of ALA?) which pertains to the stakeholders, (c) "Why?" (why exploiting ALA?) which pertains to objective(s) and (d) "How?" (how are learning analytics and adaptation performed?) which pertains to the adaptation type, the adaptation strategy, and LA aspects. Finally, to categorize the adaptation method, we distinguish among support-related, content-related, presentation-related adaptation, as suggested in [28].

## 3 Methodology

We adopt a review protocol based on existing well-established methodology for systematic literature reviews [15]. It has five phases: (a) create a search strategy by identifying search terms and databases, (b) identify selection criteria and perform an initial paper selection, (c) identify quality criteria and filter the results, (d) reach consensus about the final selection with other reviewers, (e) report findings. Phase (a) also benefited from preliminary searchers aimed identifying existing systematic reviews. The selection criteria were decided during the protocol definition in order to reduce the likelihood of bias, as suggested by the adopted methodology. Data from the two reviewers were compared and disagreements were resolved by consensus between reviewers (phase d). All papers were reviewed by both researchers and inter-researcher agreement was assessed. The review methodology adopted herein does not require the use of an ICC (Intra-Class Correlation) or Cronbach's alpha score for the agreement between reviewers, provided that inter-researcher consistency is reached in the data extraction phase (via a consensus meeting between the reviewers). Finally, duplicate reports were avoided as they would seriously bias the results.

Search terms: a search of peer-reviewed articles was conducted during Nov.–Dec. 2015, based on combinations of the terms "adaptive learning", "personalised learning", "Intelligent Tutoring System", "adaptive instruction", "adaptive hypermedia", "adaptive system", and "adaptive educational hypermedia" with the term "learning analytics". Also, the terms "adaptive learning analytics", and "personalised learning analytics", as well as the logical expressions "teaching analytics" AND personalised, "teaching analytics" AND adaptive.

Databases and other resources: the databases were selected as representative of the core literature in the areas of education and technology; ACM Digital Library, IEEE Xplore, ERIC, Science Direct, EdITLib, Springer, Elsevier, Wiley online library, JSTOR, Routledge, Sage and Cambridge University Press; and the journals: the International Review of Research in Open and Distributed Learning, Language Learning & Technology, RECALL, and Journal of Online Learning and Teaching.

Resources to be searched and selection criteria: the study included peer-reviewed journals and (full or short) conference papers, published from 2009 onwards include empirical data and follow a "top-down" approach or a combination of a "bottom-up" and "top-down" approach. They are all relevant to the topic at stake.

Quality assessment criteria: they filter the papers based on whether (a) the impact of the study (the actual impact, not the envisaged impact) for the learners or the practitioners was justified [6], (b) these participants could act upon the evidence discovered through LA and (c) the aims and the objectives were clearly reported [6].

## 4 Results

The search strategy revealed 485 papers. The selection criteria were satisfied by forty-one papers. The quality assessment process has filtered the papers down to twenty-one, after resolving one initial disagreement between reviewers on whether a particular study also satisfied the quality criteria. There was an agreement for the remaining nineteen studies which were not included in the review. All twenty-one studies [1–4, 7–9, 11–13, 16–18, 20–26, 29] were analysed according to the coding scheme suggested by the proposed framework for ALA.

Regarding the "What?" question: in more than half (11) of the studies the subject matter was STEM–related and in their vast majority the institutional context was higher education (20). Adaptive learning platforms (4), Learning Management Systems (3), and social media (3) were used in half of the studies. Finally, regarding constraints, like privacy or security, most of the studies (18) did not reference any.

Regarding the "Who?" question: the main participants were students and tutors. In two cases other experts also participated. Regarding stakeholders, that is, for whom are the results of the study interesting, the situation is similar; students are the main beneficiaries (15) followed by tutors (9). An example of a study that was mostly targeted to tutors aimed at lowering their cognitive load by exploiting LA tools to inform tutors whether a group of students in discussion forum is on-task or not and provide information about concepts coverage in the online discussion thread [29]. A few studies are addressed to designers (2) and developers of educational software (1).

Regarding the “Why?” question: the objective was mostly to promote (self-)reflection for students or tutors (18) and at a much lesser extent to propose recommendations to them or to provide predictions of students’ progress (3). The purpose of reflection was typically to help tutors make informed decisions by recognizing performance gaps and misconceptions, providing proper feedback [4], monitoring online student discussions [29] etc. Similarly, recommendations for the tutors were based on student performance and other student behaviour metrics, like collaboration indicators [16]. A related example involves dynamic modelling of roles in a collaborative online learning environment and subsequent suggestions presented by the collaboration analysis system to the tutor about emerging student roles in a given scenario [16].

Regarding the “How?” question: regarding adaptation, more than half of the studies (11) involved adaptability, whereas six studies involved adaptivity and one study involved both. In the remaining studies, the adaptation type was not clearly mentioned and could not be inferred by the reviewers. This was also the case for the adaptation strategy in some studies. Adaptive support, as an adaptation method, entailed feedback or student grouping or other type of instructional support. Adaptive content involved adaptive content release in one case [17] and the content of LA in another case [9]. Finally, adaptive presentation involved the different types of LA [25] and the presentation of the Open Learner Model [1]. Regarding the adaptation parameter(s) exploited, it was mostly the student’s performance or group performance. In two studies it was based on their role [9, 16], and in one study it was based on their informational needs [24]. With regards to the measurement of LA, a diversity of metrics emerged, including collaboration data [22], time spent on learning materials [17, 24], number of peer endorsements [11], and so on. With respect to the collection of LA, log files were frequently used as data sources and tracking systems or plug-ins were used as collection mechanisms.

## 5 Conclusions

Regarding the “what” question, more empirical research in diverse domains and institutional contexts would help assessing the reproducibility of results and the generalizability of models. Concerning constraints, and taking into account that the importance of issues like data privacy and security has been frequently stressed by researchers of both constituent fields, a recommendation for the ALA researchers would be to raise visibility and explicitly mention in their work the measures taken that ensured that no violations occurred. Regarding the “who” question, there is a need to distinguish participants from beneficiaries. Regarding the “why?” question, possible future research directions include integrating reflection with recommendation or prediction in order to maximize learning. From this review emerges that within the reflection strand, collaborative learning was an aspect that has just started to attract the interest of some researchers. This follows the developments of adaptive learning systems where a new interesting approach is flourishing, which is the constructivist-collaborative approach. Recommendations in this direction include the identification of specific theoretical frameworks that guided the ALA endeavors from a pedagogical point of view.

Regarding the "how" question, a future direction is related to a shift towards adaptability and the uptake of the locus of control by the student. A notable weakness that the review has identified concerning the "how" question is that the adaptation parameter in the majority of the studies was the student performance. Although performance is an important indicator of how the student is progressing, it is a quite narrow indicator. Thus, an ensuing recommendation for future ALA endeavors includes the acquisition of students' skills, or, the exploitation of ALA for learning at the student attitudes level.

**Acknowledgments.** This research work is supported by the European Research Consortium for Informatics and Mathematics (ERCIM).

## References

1. Ahmad, N.: Self-directed learning: student's interest in viewing the learner model. In: IEEE Research and Innovation in Information Systems (ICRIIS), pp. 493–498 (2013)
2. Ammari, A., Lau, L., Dimitrova, V.: Deriving group profiles from social media to facilitate the design of simulated environments for learning. In: ACM 2nd International Conference on Learning Analytics and Knowledge, pp. 198–207 (2012)
3. Anaya, A.R., Luque, M., Peinado, M.: A visual recommender tool in a collaborative learning experience. Expert Syst. Appl. **45**, 248–259 (2016)
4. Baneres, D.: Towards an analytical framework to enhance teaching support in digital systems design course. In: IEEE 9th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), pp. 148–155 (2015)
5. Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Thüs, H.: A reference model for learning analytics. Int. J. Technol. Enhanced Learn. **4**(5–6), 318–331 (2012)
6. Dybå, T., Dingsøyr, T.: Empirical studies of agile software development: a systematic review. Inf. Softw. Technol. **50**(9), 833–859 (2008)
7. Ezen-Can, A., Grafsgaard, J.F., Lester, J.C., Boyer, K.E.: Classifying student dialogue acts with multimodal learning analytics. In: ACM 5th International Conference on Learning Analytics and Knowledge, pp. 280–289 (2015)
8. Falakmasir, M.H., Hsiao, I.H., Mazzola, L., Grant, N., Brusilovsky, P.: The impact of social performance visualization on students. In: IEEE 12th International Conference on Advanced Learning Technologies, pp. 565–569 (2006)
9. Florian-Gaviria, B., Glahn, C., Gesa, F.R.: A software suite for efficient use of the european qualifications framework in online and blended courses. IEEE Trans. Learn. Technol. **6**(3), 283–296 (2013)
10. Greller, W., Drachsler, H.: Translating learning into numbers: a generic framework for learning analytics. Educ. Technol. Soc. **15**(3), 42–57 (2012)
11. Hickey, D.T., Quick, J.D., Shen, X.: Formative and summative analyses of disciplinary engagement and learning in a big open online course. In: ACM 5th International Conference on Learning Analytics and Knowledge, pp. 310–314 (2015)
12. Hijon-Neira, R., Velazquez-Iturbide, A., Pizarro-Romero, C., Carrico, L.: Merlin-know, an interactive virtual teacher for improving learning in Moodle. In: IEEE Frontiers in Education Conference (FIE), pp. 1–8 (2014)
13. Kamardeen, I.: Adaptive e-tutorial for enhancing student learning in construction education. Int. J. Constr. Educ. Res. **10**(2), 79–95 (2014)
14. Kay, J.: Learner control. User Model. User Adap. Inter. **11**(1–2), 111–127 (2001)

15. Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic literature reviews in software engineering–a systematic literature review. Inf. Softw. Technol. **51**(1), 7–15 (2009)
16. Marcos-García, J.A., Martínez-Monés, A., Dimitriadis, Y.: DESPRO: a method based on roles to provide collaboration analysis support adapted to the participants in CSCL situations. Comput. Educ. **82**, 335–353 (2015)
17. Martin, F., Whitmer, J.C.: Applying learning analytics to investigate timed release in online learning. Technol. Knowl. Learn. **21**, 1–16 (2015)
18. Mora, N., Caballé, S., Daradoumis, T., Gañán, D., Barolli, L.: Providing cognitive and social networking assessment to virtualized collaborative learning in engineering courses. In: IEEE International Conference on Intelligent Networking and Collaborative Systems (INCoS), pp. 463–468 (2014)
19. Papamitsiou, Z., Economides, A.A.: Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. J. Educ. Technol. Soc. **17**(4), 49–64 (2014)
20. Groba, A.R., Barreiros, B.V., Lama, M., Gewerc, A., Mucientes, M.: Using a learning analytics tool for evaluation in self-regulated learning. In: IEEE Frontiers in Education Conference (FIE), pp. 1–8 (2014)
21. Ruipérez-Valiente, J.A., Muñoz-Merino, P.J., Leony, D., Kloos, C.D.: ALAS-KA: a learning analytics extension for better understanding the learning process in the khan academy platform. Comput. Hum. Behav. **47**, 139–148 (2015)
22. Rodríguez-Triana, M.J., Martínez-Monés, A., Asensio-Pérez, J.I., Dimitriadis, Y.: Scripting and monitoring meet each other: aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. Br. J. Educ. Technol. **46**(2), 330–343 (2015)
23. Santos, J.L., Verbert, K., Govaerts, S., Duval, E.: Addressing learner issues with StepUp! an evaluation. In: ACM 3rd International Conference on Learning Analytics and Knowledge, pp. 14–22 (2013)
24. Santos, J. L., Govaerts, S., Verbert, K., Duval, E.: Goal-oriented visualizations of activity tracking: a case study with engineering students. In: ACM 2nd International Conference on Learning Analytics and Knowledge, pp. 143–152 (2012)
25. Tabuenca, B., Kalz, M., Drachsler, H., Specht, M.: Time will tell: The role of mobile learning analytics in self-regulated learning. Comput. Educ. **89**, 53–74 (2015)
26. Timmers, C.F., Walraven, A., Veldkamp, B.P.: The effect of regulation feedback in a computer-based formative assessment on information problem solving. Comput. Educ. **87**, 1–9 (2015)
27. Vahdat, M., Ghio, A., Oneto, L., Anguita, D., Funk, M., Rauterberg, M.: Advances in learning analytics and educational data mining. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (2015)
28. Vandewaetere, M., Desmet, P., Clarebout, G.: The contribution of learner characteristics in the development of computer-based adaptive learning environments. Comput. Hum. Behav. **27**(1), 118–130 (2011)
29. van Leeuwen, A., Janssen, J., Erkens, G., Brekelmans, M.: Teacher regulation of cognitive activities during student collaboration: effects of learning analytics. Comput. Educ. **90**, 80–94 (2015)

# An Adaptive E-Learning Strategy to Overcome the Inherent Difficulties of the Learning Content

Anna Mavroudi[1(✉)], Thanasis Hadzilacos[2], and Charoula Angeli[3]

[1] Norwegian University of Science and Technology, Trondheim, Norway
anna.mavroudi@idi.ntnu.no
[2] Open University of Cyprus, Nicosia, Cyprus
thh@ouc.ac.cy
[3] University of Cyprus, Nicosia, Cyprus
cangeli@ucy.ac.cy

**Abstract.** In this paper we propose a strategy of adaptive e-learning that aims to help students overcoming the inherent difficulties of STEM-related subject matters for which they have known misconceptions. The paper reports on empirical findings derived from classroom interventions which were undertaken to investigate the impact of the proposed strategy. For each intervention, an adaptive e-course was developed and tested with encouraging results. Since the proposed strategy is descriptive in nature the paper can be used as the basis for future studies that validate it with other subject matters than those mentioned herein.

**Keywords:** Adaptive learning · E-learning strategy · STEM · Feedback · Content presentation · Inherent difficulties · Common student errors

## 1 Introduction

This study presents an adaptive learning strategy which assisted the participant students to overcome the inherent difficulties of five STEM-related topics. In the proposed adaptive learning strategy, learning style and feedback were considered as an adaptation parameter. Regarding the former, a content analysis of 70 publications from 2000 to 2011 on adaptive educational hypermedia accommodating learning styles [1] revealed that the direct and positive influence on learning outcomes of adaptation based on learning styles was still unclear. The most preferred learning style model for research work was the Felder–Silverman Learning Style model [3], which was utilized in 35 studies (50 %), followed by the Model of Kolb's Cycle of Learning and learning styles [8], the VARK model [4], the Honey and Mumford model [6] and other individual models. Regarding adaptive feedback, it is a form of adaptive scaffolding mostly related to the student's help and support researchers converge that effective feedback (feedback that facilitates the greatest gains in learning) should provide the student with two types of information incorporated into the item response: verification and elaboration.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 440–445, 2016.
DOI: 10.1007/978-3-319-45153-4_40

## 2 Course Design

### 2.1 Phase 1: Preparatory Phase-Learning Style Typology Selection

The four popular learning style models mentioned above were explained to 40 teachers, in a written form, avoiding domain-specific terminology, through an online questionnaire survey. In their vast majority, the teachers are working in secondary education schools in Cyprus. Thirty six of the teachers live in Cyprus, three live in Greece and one in the UK. There is no known bias on the teacher sample used for the survey; however it was not a scientifically random sample. Teachers approached were those who have active contact with a teachers' union (those that live in Cyprus). They were asked to select the model they thought was closer to their everyday teaching practices (i.e., more applicable). The question was purposefully formed so that teachers could specify one or several models as being close to their practice. There was no limit and no indication whether one or more were expected. The questionnaires were answered anonymously. The preferred model would be incorporated in the proposed adaptive e-learning strategy. According to the survey results, 20 % of the participant teachers selected the Honey-Mumford model (8 votes), 57.50 % selected the VARK model (23 votes), 15 % selected the Kolb model (6 votes) and 35 % selected the Felder & Silverman model (14 votes). As a result, the VARK model was used as the learning preference typology.

### 2.2 Phase 2: Course Design and Development

A series of adaptive e-courses were designed, developed, implemented, and tested in real classroom settings. Five learning topics were selected from the STEM (Science, Technology, Engineering, Mathematics) domain and in particular, Mathematics and Informatics. A literature review conducted by the authors revealed clear evidence, which converges on specific inherent difficulties, i.e., common student errors or topics that present difficulties for the students to understand or for the teachers to teach (Table 1).

Accordingly, five adaptive e-courses that focused on inherent topic difficulties were designed and developed. An open source learning design editor, named "ReCourse" was used for their development and an open source learning design player, named "Astro player" was used for their implementation in the classroom. The adaptive e-courses incorporated a rule-base logic that supported the adaptation strategy described below. The table below outlines: the topics of the e-courses, the respective domains and the difficulties that each subject matter presents. The design was participatory and the protocol used to enable teachers to act as co-designers of the adaptive e-courses is described in detail in [12].

Three types of e-courses were developed for each of the above topics: a non-adaptive e-course, an adaptive e-course that incorporated one adaptation parameter (e.g., adaptive learning flow) and an adaptive e-course that incorporated two adaptation parameters (e.g., adaptive learning flow and content presentation). The three types shared common characteristics: (a) focus on the inherent difficulties of the topic to be

**Table 1.** Inherent difficulties in each topic

| Topic/domain | Difficulties |
|---|---|
| Inequalities/Mathematics | Students reject solutions that do not fit the general pattern, i.e., an interval for inequalities, a unique value for equations [15]; students multiply or divide the two sides of an inequality by the same number without checking whether the number is positive, negative, or zero [5]. |
| Ratios and analogies/Mathematics | Students have the tendency to treat pseudo-proportionality problems as if they were actual proportionality problems, and, consequently, they apply linear models to them [13]; students use additive reasoning instead of multiplicative in proportionality problems [9]. |
| System of linear equations/Mathematics | Students are facing difficulties in understanding that different representations of a system (graph, algebraic solution, ordered values table) are equivalent, and move back and forth between them [14]; students are not sure what to do when all variables are eliminated, or when the system doesn't have a solution [14]. |
| WWW, Internet and communication protocols/Informatics | It is difficult to teach how two different digital devices communicate [7]; students do not differentiate between Internet and WWW [7]. |
| Main and auxiliary memory/Informatics | There is a complexity in explaining the differences between the two types of memory [7]. |

taught, (b) inclusion of a preparatory phase at the beginning of the e-course aiming to help students recall prior knowledge, and (c) incorporation of a variety of different content representations. The non-adaptive e-courses had a linear/sequential learning flow in tandem with knowledge-of-response type of feedback and supplementary, elaborative feedback. Knowledge-of-correct-response feedback provided learners with the correct answer, while elaborative feedback explained why the specific answer was the correct one. The adaptive e-courses had one or both of the following design attributes:

- Adaptive learning flow; non-linear/networked learning flow in tandem with response-contingent adaptive feedback. Response-contingent feedback provided knowledge of the correct response along with an explanation of why the incorrect answer was wrong and why the correct answer is correct. Then, in case of an incorrect answer, the student was presented with a similar problem. This second problem was treated with knowledge-of-response type of feedback in tandem with elaborative feedback, as it was the case with the control group.
- Adaptive content presentation; media in accordance to students' diagnosed learning styles (see [11]). To this end, the VARK model [4] was used.

## 3 Participants

Five teachers and 149 students from six schools participated in the classroom interventions. Seventy students were assigned to the control group and 79 to the experimental group. All participant teachers and students live in Cyprus. As shown in Table 2, the numbers and ages of participant students involved in each intervention are varying.

**Table 2.** Student participation in each intervention

| Topic/domain | Age of students | Student number in the focus group | Student number in the control group |
|---|---|---|---|
| Inequalities | 14 | 16 | 16 |
| Ratios and analogies | 12–13 | 17 | 11 |
| System of linear equations | 15 | 10 | 8 |
| WWW, internet communication protocols | 14 | 26 | 25 |
| Main and auxiliary memory | 16 | 10 | 10 |

## 4 Procedures and Instruments

The five interventions followed the randomized control pre-test-post-test experimental design paradigm. In each intervention, an adaptive learning e-course was implemented with the experimental group, and a non-adaptive e-course with the control group. In each classroom intervention, the control group and experimental group were tested simultaneously. All student participants in each group were individually tested in a computer lab. A short session of 15 min took place before each intervention that aimed to familiarize students with the digital environment. The role of the teacher during each intervention was to support students with the use of the digital environment whenever needed. The developer had a short training session with each of the participant teachers about the digital environment, prior to the beginning of each intervention. The developer was also present in the classroom during each intervention in case extra technical support would be needed. Regarding the duration of the adaptive e-courses, due to the embedded learning strategy, it was varying (60 min in average). The participant students completed pre-tests and post-tests that (a) focused on the topics mentioned in Table 1 above, (b) were identical for each intervention and (c) contained problems and questions related to the inherent difficulties of the topics. The Mathematics tests mostly contained open-ended problems, and the Informatics tests contained multiple choice questions. A week before each intervention, all students completed the diagnostic learning style questionnaire in a printed format. The researcher analysed the answers in order to diagnose the learning style of each student. In the Mathematics courses, the students completed the performance tests in a printed format, and they were graded by two participant Mathematics teachers after being anonymized. In the Informatics courses computerised tests were used which were automatically graded.

## 5 Results

The gain scores were used for the assessment of students' performance improvement which can provide a means for assessing the impact of the interventions [2]. The gain scores D were calculated using the formula D = Y2 −Y1, where Y1 = pre-test scores and Y2 = post-test scores. The mean value of the gain scores (which indicates performance improvement) in the control group was equal to 1.737 (S.D. = 2.46), whereas in the experimental group the mean value of the gain scores is 2.79 (S.D. = 2.81). Consequently, the mean difference in the performance improvement between the experimental group and the control group is equal to 1.053 (out of 10 grades). The mean performance of the students in the pre-test was equal to 3.481 (out of 10) and in the post-test was equal to 5.766 (out of 10). Gain scores were normally distributed for the focus group students but not for the control group students, as assessed by Shapiro-Wilk's test ($p < .05$). Consequently, a Mann-Whitney U test was run to determine if there were differences in gain scores between control group students and experimental group students. Distributions of the gain scores for control group students and experimental group students were approximately similar, as assessed by visual inspection. Gain scores for experimental group students (mean rank = 83.46) were statistically significantly higher than for control group students (mean rank = 65.45), U = 3433.500, z = 2.552, p = .05. Finally, it should be noted that there were no statistically significant differences in gain scores between the different age groups, $F(4,74) = 1.482$, p = .216.

## 6 Discussion and Conclusions

The students that followed the adaptive e-courses outperformed those that followed non-adaptive e-courses, while, in their majority, the students' grades before and after the interventions were relatively low. This is justified by the nature of the pre-post tests: they did not have increasing difficulty level, but, instead, they were comprised by problems or questions focused almost exclusively on the inherent difficulties or known misconceptions of the learning content. Added value to the work discussed herein offers the fact that, despite adapting to individual differences is considered essential, there is a lack of empirical studies on computer-supported adaptive instruction interventions [10]. The adaptive learning strategy which was embedded in the e-courses aimed at helping students overcome the inherent difficulties of the learning content, exploited a cognitive constructivist point of view towards instruction, which considers knowledge as something that is actively constructed by students, based on their existing cognitive structures. Indeed, students' prior knowledge and learning style were taken into account and, consequently, corrective actions were automatically provided to the students that followed the adaptive e-courses in the forms of personalized scaffolds to targeted problems directly related to the inherent difficulties. These scaffolds pertained to targeted, elaborative type of feedback and the engagement of the students with extra problems, whenever needed.

## References


1. Akbulut, Y., Cardak, C.S.: Adaptive educational hypermedia accommodating learning styles: a content analysis of publications from 2000 to 2011. Comput. Educ. **58**(2), 835–842 (2012)
2. Dimitrov, D.M., Rumrill Jr., P.D.: Pretest-posttest designs and measurement of change. Work J. Prev. Assess. Rehabil. **20**(2), 159–165 (2003)
3. Felder, R.M., Silverman, L.K.: Learning and teaching styles in engineering education. Eng. Educ. **78**(7), 674–681 (1988)
4. Fleming, N.D.: Teaching and Learning styles: VARK strategies. N.D. Fleming, Christchurch (2001)
5. Halmaghi, E.: Undergraduate students conceptions of inequalities: sanding the lens. In: The International Conference of Psychology of Mathematics Education Proceedings, pp. 41–46 (2010)
6. Honey, P., Mumford, A.: Using your learning styles, 2nd edn. Peter Honey, Maidenhead (1986)
7. Ioannou, I., Angeli, C.: Teaching computer science in secondary education: a technological pedagogical content knowledge perspective. In: Caspersen, M.; Romeike, R.; Knobelsdorf, M. (Eds.) Proceedings of the 8th Workshop in Primary and Secondary Computing Education - WiPSCE 2013. ACM (2013)
8. Kolb, D.: Experiential learning as the science of learning and development. Englewood Cliffs NPH (ed.) (1984)
9. Lamon, S.J.: Ratio and proportion: connecting content and children's thinking. J. Res. Math. Educ. **24**(1), 41–61 (1993)
10. Lee, J., Park, O.: Adaptive instructional systems. In: Spector, J.M., Merill, M.D., van Merrienboer, J., Driscoll, M.P. (eds.) Handbook of Research for Educational Communications and Technology, pp. 469–484. Routledge, Taylor & Francis Group, New York (2008)
11. Mavroudi, A., Hadzilacos, T.: Broadening the use of e-learning standards for adaptive learning. In: Popescu, E., Li, Q., Klamma, R., Leung, H., Specht, M. (eds.) ICWL 2012. LNCS, vol. 7558, pp. 215–221. Springer, Heidelberg (2012)
12. Mavroudi, A., Hadzilacos, T., Panteli, P., Aristodemou, A.: Integration of theory, ICT tooling and practical wisdom of teacher: a case of adaptive learning. In: Popescu, E., Lau, R.W., Pata, K., Leung, H., Laanpere, M. (eds.) ICWL 2014. LNCS, vol. 8613, pp. 152–158. Springer, Heidelberg (2014)
13. Modestou, M., Gagatsis, A.: Students' improper proportional reasoning: a result of the epistemological obstacle of "linearity". Educ. Psychol. **27**(1), 75–92 (2007)
14. Proulx, J., Beisiegel, M., Miranda, H., Simmt, E.: Rethinking the teaching of systems of equations. Math. Teach. **102**(7), 526–533 (2009)
15. Tsamir, P., Bazzini, L.: Can x = 3 be the solution of an inequality? A study of Italian and Israeli students. In: Proceedings of PME25, Utrecht, The Netherlands, vol. IV, pp. 303–310 (2001)

# Evaluating the Effectiveness of an Affective Tutoring Agent in Specialized Education

Aydée Liza Mondragon(✉), Roger Nkambou, and Pierre Poirier

Université de Québec à Montréal (UQAM), Montreal, Canada
aydeelizamondragon@gmail.com, nkambou@gmail.com,
pierre.g.poirier@gmail.com

**Abstract.** Autism spectrum disorder (ASD) is a neurological disorder affecting the way in which the brain processes information. Autism is characterized by impairments in learning and communication, in social interaction, imaginative ability as well as in repetitive and restricted patterns of behavior [9]. This research contributes to the advancement of intelligent tutoring systems by proposing an affective intelligent tutoring system in the field of specialized education. We have conducted an experiment in mathematical learning with one controlled group of six participants who interacted without the support of the pedagogical agent Jessie, while the test group of six participants interacted with Jessie. The purpose of this study was to validate the support provided by the pedagogical agent Jessie based on our accompaniment model. The results showed significant improvement in learning by the test group.

**Keywords:** Autism · Affective intelligent tutoring systems · Specialized education · Personalized education · Model of accompaniment

## 1 Introduction

Studies reveal that individuals with learning disabilities pose a 'complex multi-factor' problem in the educational system [21]. The problem stems from the fact that in most educational institutions, one-on-one intervention is difficult to implement due to budgetary and human constraints. Individuals with learning disabilities (LD) who require extra resources comprise 13 % of all students in the USA [14]. The prevalence of ASD reveals an increasing trend in the occurrence of autism. In Canada, the recent estimates (March 2014) by the [6] and the Developmental Disabilities Monitoring (DDM) suggest that 1 in every 68 children were born with an ASD. Why an affective Intelligent tutoring system (AITS) in specialized education? The Integrated Specialized Learning Application (ISLA) provides individualized support to help autistic children manage their emotions by analyzing the learning trace and considering the learner's current performance to respond accordingly to it during a mathematical learning situation. This paper is divided into six sections. The first section is the introduction. The second section presents a brief literature review on autism, emotions, learning and intelligent tutoring systems (ITS). Section three describes ISLA's components. In section four, the main experiment of the prototype and the results are presented. Finally, the conclusion and the limitations are discussed outlining the contribution of this research.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 446–452, 2016.
DOI: 10.1007/978-3-319-45153-4_41

## 2 Autism, Emotions and Learning

Emotions and learning have been broadly recognized as challenging among individuals diagnosed with autism [15]. The socio-cognitive and behavioral problems experienced by individuals with ASD are considered to stem from the difficulty of understanding others' mental states [3, 10]. During intervention, one important challenge is due to the difficulty of anticipating and recognizing negative behaviors, consequently calibrating the child's affective state for effective intervention and learning. These challenges vary from child to child as these individuals may have profound cognitive deficiencies while others may have IQ scores that are equal to or higher than the typical person [9]. This diversity of profiles causes multiple challenges in terms of methodologies and teaching programs directed towards the education of children with autism. This is the reason why we believe that modeling affect is the proper approach for ISLA to teach mathematics to children in the spectrum of autism.

An Intelligent Tutoring System (ITS) is a computer system designed with the objective of providing instant and customized instruction or feedback to students as effective as one-to-one tutoring [4]. This is generally done without intervention from a human teacher, and with the intention of enabling learning in a meaningful effective way by using a variety of computing technologies [18]. Many ITSs have been developed and are being used in different domains in education (i.e. Auto Tutor [12], ANDES [20], among others), in corporate and industry training (i.e. Sherlock [11]). Within the domain of intelligent tutoring systems, [2] points out that the companion agent has the potential of providing students of all ages with information that will help the student to become self-regulated, consequently become an independent learner. In [2], they examined the effectiveness of pedagogical agents (PAs') with MetaTutor for the purpose of training students in self-regulated learning (SRL) processes to cognitive diagnosis through prompting and providing feedback that facilitated learning about the human circulatory system. Researchers claim that if computers are to interact naturally with humans, they must also be able to recognize, affect and express social competencies [17]. The 'affective' approach within ITS has been validated as having a positive impact on the learner's intuition, and on his/her self-esteem, as it relates to problem solving tasks [16]. Thus 'cognition', 'motivation' and 'emotional affect' are three components of learning [7]. In [1], they have demonstrated different sensors that can be used to detect emotions. Other examples of physiological sensors that have a direct correlation according to two dimensions valence (positive or negative emotion), and arousal (intensity of emotion) are the GSR sensor (Galvanic Skin Response), the RSP sensor (respiration), the BVP sensor (Blood Volume Pressure), and TEMP sensor. One example of AITS is the Wayang Tutor intended for middle school and high school level mathematical learning composed of four physiological sensors [1, 21].

## 3 System Overview and Pedagogical Model

ISLA is unique and its contribution entails the model of accompaniment to help autistic children manage their emotions. ISLA is an adaptive application that provides individualized intervention which evolves along with the autistic learner's needs. This is done by analyzing the learning trace and the student's current performance. Through the incorporation of aspects of the accompaniment model, ISLA both supports and integrates the domain and learner models. For example, ISLA makes use of an individualized intervention plan (IIP) [13], which provides guidance and key elements about the curriculum, the pedagogy, and the behavior required from the autistic learner to best meet the individual's learning needs. In ISLA, the pedagogical agent called Jessie is capable of detecting the affective state of an autistic child in mathematical learning. This is displayed in the user's interface and related to the accompaniment model. The interface provides a three-dimensional view that allows personalizing the interaction of the three core models of ISLA. This is from the domain model point of view by providing tools to manipulate domain objects. The accompaniment model point of view through Jessie (pedagogical agent), and the learner model point of view using an open-learner modeling approach [5]. The accompaniment model of ISLA implements rules that should be followed by Jessie to help an autistic learner manage his/her emotions based on the learning trace and his/her current performance. This component is drawn from the self-regulated learning theory highlighting the essential role that metacognition plays in self-regulation and learning [19]. The ASD learner must finish a task before moving to the next phase in order to increase the chance to master the prerequisites of the activity at hand. When a right answer is provided, positive reinforcement is used by Jessie, with social rewards and feedback in order to encourage and motivate the learner, such as 'Y*es, you did it!*', or '*Good Job!*' By contrast, when a wrong answer is given, Jessie can say something like this: '*That was close, nice try!*' and it invites the ASD learner with prompting to try again. Furthermore, if the learner needed help, hints were provided based on pedagogical scenarios. The learner model is made of the cognitive profile and the affective profile of the learner. Both profiles are maintained by the system and the specialized educator during learning activity. The affective profile selected in this study includes the affects of: disengagement, encouragement, frustration, interest, anxiety, happiness, guidance, and anger because they are considered relevant in autism intervention practices [8].

## 4 The Methodology

We would like to mention that a preliminary study experiment was previously carried out. The results of the preliminary experiment revealed that the performance of the ASD learners, in mathematical learning with the use of a pedagogical agent providing real-time support, had a positive impact on these participants' performance. For the main experiment dealing with the prototype, we have developed an interactive quiz in mathematical learning for the two groups interacting with ISLA. The quiz was validated by professionals in the field of specialized education related to autism and consisted of thirty questions. The first version of the interactive quiz was intended for the six participants who interacted with ISLA without the support of the pedagogical agent Jessie, while the other version was used for the test group who interacted with the pedagogical agent Jessie. A study protocol and intervention protocol were created for each of the main experiment providing the guidelines on how the intervention was conducted. The research population consisted of twelve participants diagnosed with high functioning autism spectrum disorders (ASDs), i.e. boys and girls aged from 6 to 12 years old, with the consent of their parents and under the supervision of a specialized educator. Each learning session lasted one hour, in which, a one-on-one structured intervention in mathematical learning was provided to the ASD participant. The participants recruited in this study came from private clinics, specializing in autism, as well as from centers for rehabilitation and specialized education related to autism, all located in Montreal, Canada.

## 5 The Results

### 5.1 Methods

Descriptive statistics summarize all study variables of interest. For categorical variables, we reported counts and percentages whereas for continuous variables we reported medians and inter-quartile range (IQR[1]), because the values did not follow an approximate normal distribution. We compared scores between the group with and without Jessie. Due to small sample size and the difficulty to verify the assumption that the scores in the population follows an approximate normal distribution, we performed the exact version of Wilcoxon Rank Sum (WRS) test for independent samples, a non-parametric equivalent of the t-test. The null hypothesis for this test is that the distribution of values of scores for the two groups do not differ. All statistical tests of hypothesis were two-sided and performed at the pre-specified level of significance of 5 %. The *p*-values reported are not adjusted for multiple testing. We used SAS, version 9.3 (SAS Institute Inc., Cary, NC, USA) for all statistical analyses.

### 5.2 Results Analysis

Participation of each child in each group was allocated randomly. In the group without Jessie, the age of the children ranged from 7 years old to 12 years old.

The participants' profile for the group without Jessie is presented in Table 1. In the group with Jessie, the age of the children ranged from 6 years old to 12 years old. The participants' profile for the group with Jessie is illustrated in Table 2.

### 5.3 Comparison of Performance Scores

We present the results related to the relationship between support and performance dealing with the score of each participant for both groups with and without Jessie during the mathematical activity. Raw scores were corrected to give the ASD participant his/her level of success according to his/her level of competency in addition. Except for a participant whose score was 100 %, no participant was able to complete the quiz according to their level of competency. The raw scores fluctuated from 7 % as

[1] IQR = Inter-Quartile Range = 25 % percentile – 75 % percentile.

**Table 1.** Participants' profile group without Jessie

| Participant # | Diagnosis | Age | Gender |
|---|---|---|---|
| #7 | Autism disorder | 9 | Female |
| #8 | Autism disorder | 6 | Male |
| #9 | Autism disorder | 7 | Male |
| #10 | Autism disorder | 8 | Male |
| #11 | Asperger Syndrome | 10 | Female |
| #12 | Autism disorder | 12 | Male |

**Table 2.** Participants' profile group with Jessie

| Participant # | Diagnosis | Age | Gender |
|---|---|---|---|
| #1 | Autism disorder | 12 | Male |
| #2 | Autism disorder | 9 | Female |
| #3 | Autism disorder | 9 | Male |
| #4 | Autism disorder | 8 | Male |
| #5 | Autism disorder | 11 | Male |
| #6 | Autism disorder | 7 | Male |

being the lowest score to 100 % as being the maximum score with a median of 41.7 (IQR 23.3–63.3). On the other hand, in the group with Jessie, where participants benefited from its support, all six children were able to complete the quiz according to their level of competency. The raw scores differed from 10 % as being the lowest score to 67 % as being the maximum score. The results indicated a median of 50.0 (IQR 33.3–63.3). For the competency scores, in the group without Jessie, the scores fluctuated from 40 % to 100 %. In this group, the median for the competency scores was 72.0 (IQR 58.3–86.4). In the group with Jessie, the competency scores differed from 60 % to 92 %. The results indicated a median of 86.4 (IQR 83.3–90.9). The exact WRS test on raw scores reveals no difference in the distribution of scores between the groups (S = 24.0, $p = 0.33$). Similarly, the exact WRS test displays no difference in the distribution of competency scores between the groups (S = 20.0, $p = 0.08$). We noted that when the possible outlier was removed from the group without Jessie, the competency scores revealed a statistical difference between the groups in a two-sided statistical test ($p = 0.08$). Beside, a one-sided WRS test on competency scores revealed a significant difference between the groups (WRS test, S = 20.0, $p = 0.04$), with a distribution with higher values for the group with Jessie.

### 5.4 Inter-Group Variation of Affective States

The results of the statistical analysis to compare the affective states between the groups without Jessie and with Jessie (N = 12, N-11) reveals that the support of Jessie to help the autistic child to calibrate his/her emotions during the mathematical activity had a significant difference for the affects of encouragement between the groups (WRS test, S = 57.0, p = 0.002), frustration (WRS test, S = 19.5, p = 0.05), and guidance (WRS test, S = 51.5, p = 0.04). A one-sided WRS test on the effect of anxiety revealed a significant difference between the groups (WRS test, S = 27.0, p = 0.03), with a distribution with higher values for the group with Jessie. Similarly, for the effect of anger, a one- sided WRS test revealed a significant difference between the groups (WRS test, S = 29.0, p = 0.05). The results showed that when the possible outlier was removed from the group without Jessie (N = 11), it had a significant difference for the effects of disengagement (WRS test, S = 42.0, p = 0.03), encouragement (WRS test, S = 15.0, p = 0.004), and anger with (WRS test, S = 51.0, p = 0.04). A one-sided WRS test on the effect of frustration revealed a significant difference between the groups (WRS test, S = 19.5, p = 0.05), with a distribution with higher values for the group with Jessie. Similarly, for anxiety, a one-sided WRS test showed a significant difference between the groups (WRS test, S = 40.0, p = 0.04), and for guidance, a one-sided WRS test showed a significant difference between the groups (WRS test, S = 20.0, p = 0.04).

## 6 Conclusion, Limitations and Future Work

In this research, we have conducted a study using a prototype of ISLA that implemented Jessie as a pedagogical agent based on our accompaniment model. The results revealed that the majority of participants in the test group benefited from the personalization and support provided by the pedagogical agent Jessie, which aimed at helping the autistic student be- come self-regulated by calibrating his/her emotions and encouraging motivation during the mathematical activity. One limitation is that the groups were heterogeneous for the two experiment with and without Jessie in terms of age. Also, the level of competency had a limitation, especially in the group without Jessie, one participant scored 100 % on the quiz. Future research will be dealing with a full implementation of ISLA by reproducing what has been done according to the prototype experiment. A larger group of participants with autism will be interacting with the pedagogical agent Jessie, in which, the behavior of the pedagogical agent Jessie will be programmed by providing real-time support to help calibrate the affective state of the ASD learner. Children will be grouped according to different criteria like age and competency level. They will be interacting with ISLA until the mastery level is achieved.

## References

1. Arroyo, I., Cooper, D.G. Burleson, W., Woolf, B.P., Mulder, K., Christopherson, R.: Emotion sensors go to school. In: Proceeding of AIED 2009. Brighton, pp. 17–24. IOS Press, Amsterdam (2009)
2. Azevedo, R.: Theoretical, methodological, and analytical challenges in the research of metacognition and self-regulation: a commentary. Metacognition and Learning (2009)
3. Baron-Cohen, S., Leslie, A.M., Frith, U.: Does the autistic child have a theory of mind? (1985)
4. Bloom, B.S.: The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring (1984)

5. Bull, S., Kay, J.: Open learner models. In: Nkambou, R., Bourdeau, J., Mizoguchi, R., Mizoguchi, R. (eds.) Advances in Intelligent Tutoring Systems, pp. 318–338. Springer, Heidelberg (2010)
6. Centre for Disease Control (CDC): cdc.gov
7. D'Mello, S.K., Craig, S.D., Gholson, B.F.S., Picard, R.W.: Interacting effects sensors in an intelligent tutoring systems. In: Affective Interactions, The Computer in the Affective Loop Workshop (2005)
8. Dautenhahn, K., Werry, I.: Towards interactive robots in autism therapy: back-ground, motivation and challenges. Pragmatics Cogn. **12**(1), 1–35 (2004)
9. Diagnostic and statistical manual of mental disorders: DSM-IV. American Psychiatric Association (2000)
10. Frith, U.: Interacting minds. Biol. Basis Rev. (1999)
11. Lajoie, S.P., Lesgold, A.: Apprenticeship training in the workplace: computer-coached practice environment as a new form of apprenticeship. Mach. Mediated Learn. **3**(1), 7–28 (1989)
12. Graesser, A.C., Moreno, K., Marineau, J., Adcock, A., Olney, A., Person, N.: AutoTutor improves deep learning of computer literacy: is it the dialogue or the talking head? In: Proceedings of Artificial Intelligence in Education, pp. 47–54. IOS Press, Amsterdam (2003)
13. Ministère de l'Éducation, de l'Enseignement supérieur et de la Recherche. www.education.gouv.qc.ca
14. National Center for Education Statistics (NCES)-2009-nces.ed.gov
15. National Research Council: Educating Children with Autism. National Academy Press, Washington, D.C. (2001)
16. Nkambou, R.: Towards affective intelligent tutoring system. In: ITS 2006 Workshop on Motivational and Affective Issues in ITS, Taiwan (2006)
17. Picard, R.W.: Future affective technology for autism and emotion communication. Philos. Trans. Royal Soc. B Biol. Sci. **364**(1535), 3575–3584 (2009)
18. Massey, J., Postka, L.D., Mutter, S.A.: Intelligent Tutoring Systems, Laurence Erlbaum. Lessons Learned, Hillsdale (1988)
19. Schraw, G.Z.: Promoting general metacognitive awareness. Instruct. Sci. **26**, 113–125 (1998)
20. VanLehn, K., Lynch, C., Schulze, K., Shapiro, S.L., et al.: The andes physics tutoring system: lessons learned. Int. J. Artif. Intell. Educ. **15**(3), 678–685 (2005)
21. Woolf, B., Arroyo, L., Muldner, K., Burleson, W., Cooper, D., Dolan, R., Christopherson, R.: The effect of motivational learning companions on low high achieving students and students with learning disabilities. In: International Conference on Intelligent Tutoring Systems (2010)

# MOOC Design Workshop: Educational Innovation with Empathy and Intent

Yishay Mor[1(✉)], Steven Warburton[2], Rikke Toft Nørgård[3], and Pierre-Antoine Ullmo[1]

[1] PAU Education, Barcelona, Spain
{yishay.mor,pa.ullmo}@paueducation.com
[2] University of Surrey, Guildford, UK
steven.warburton@surrey.ac.uk
[3] Aarhus University, Aarhus, Denmark
rtoft@tdm.au.dk

**Abstract.** For the last two years we have been running a series of successful MOOC design workshops. These workshops build on previous work in learning design and MOOC design patterns. The aim of these workshops is to aid practitioners in defining and conceptualising educational innovations (predominantly, but not exclusively MOOCs) which are based on an empathic user-centered view of the target learners and teachers. In this paper, we share the main principles, patterns and resources of our workshops and present some initial results for their effectiveness.

**Keywords:** MOOCs · Learning design · Learning experience design · Professional development · User-centered design · Learner-centered design

## 1 Introduction

The MOOC phenomena has opened up the field of online and blended education to institutions and individuals who had never before considered a depart from traditional modes and methods of instruction. Most major universities are either offering MOOCs or in the process of developing MOOCs, while many budget-constrained educational institutions are using MOOCs from high-ranked universities as (open) educational resources, thus developing a new type of hybrid education. We are witnessing institutions and individuals with literally no experience in online teaching (sometimes, with little experience in teaching at all) facing classes of tens of thousands of students, spread across the globe. The challenge that MOOCs present is not just in understanding and addressing the needs of these masses of learners: before that, we need to recognise the needs, desires, and dilemmas of the new breed of online educators, and find effective and principled ways to address them.

Littlejohn and Milligan [9] reviewed the design quality of 76 randomly selected MOOCs. Their results indicate that although most MOOCs are well


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 453–459, 2016.
DOI: 10.1007/978-3-319-45153-4_42

organised, their instructional design quality is low. Indeed, it seems that most educators that attempt to design and develop a MOOC begin by asking themselves 'what do I need to teach?', or, in other words 'what is the content I need to cover?'. We call this a content-centric approach. The problem with such an approach is that you can produce the most carefully selected content, in the most professionally produced manner, but if learners do not engage with it and make it their own - your efforts will have little lasting effect. In order to provide an effective and meaningful learning experience, we need to focus on the learners - who they are, where are they now (A), and where do we want them to be (B), and how do we guide them in their path from A to B.

## 2 Background

Our work is situated in the Learning Design (LD) tradition. LD is 'the act of devising new practices, plans of activity, resources and tools aimed at achieving particular educational aims in a given situation' [10]. This is a creative process; the designer is bringing new objects into existence. Yet it is also a process of inquiry: the designer needs to understand the situation and establish the efficacy of the objects she creates in bringing about the desired effects. This duality of LD, and the challenges that it poses, has been discussed in depth elsewhere [11].

Engaging educational practitioners in LD has benefits beyond the immediate task [15]. However establishing a design mindset is not trivial [11]. In recent years, there have several attempts to address this issue [2–4,14]. The Learning Design Studio (LDS) draws on these and other frameworks, to offer a process that explicitly interleaves the creative elements of design into a cycle of *Design Inquiry of Learning* [12,17]. In this cycle, participants identify an educational challenge they wish to address, investigate the context of this challenge and the forces that shape it, review relevant theory and practical examples, conceptualise a solution, implement a prototype of that solution, evaluate it and reflect on the process.

The purpose of education, as Dewey eloquently phrased it [5], is to provide learners with the experiences that promote growth. To serve such a cause educational design needs to adopt a clear user-centered position of empathy [1]. This call for empathy is inline with a growing acknowledgement of the role of empathy in design [6,7,13]. Postma et al. [13] define empathic design as '*a design research approach that is directed towards building creative understanding of users and their everyday lives for new product development*'. They describe creative understanding as a rich combination of cognitive (knowledge) and affective (feeling) perception of the user, which the designer can translate into new products that will meet the user's values, aspirations and constraints. They propose four principles of empathic design: balancing rationality and emotions in building understanding of users' experiences, making empathic inferences about users and their possible futures, involving users as partners, and engaging design team members as multi-disciplinary experts in performing user research. Despite the importance of empathy in education, most LD methodologies do not address the issues of empathy directly.

## 3 The Empathic MOOC Design Workshops

Following the success of the MOOC design pattern project [16], we turned our attention to the effective support of practitioners wishing to design and produce a new MOOC. Building on the LDS methodology, we designed a workshop format that leads participants through a rapid cycle of design inquiry of learning, with a clear empathic mindset, rooted in a vision of the learners, their values, needs and constraints. This cycle flows through the following phases:

1. **Imagine**: identify an educational challenge which your MOOC / educational innovation will address.
2. **Investigate**: Characterise your learners, and describe the transition they will achieve as a result of the educational innovation.
3. **Inspire**: Review evidence of effective, valuable and meaningful designs, and consider its implications for your educational innovation.
4. **Ideate**: Use the analysis of effective and valuable designs to conceptualise your educational innovation.
5. **Evaluate**: Scrutinise your solution to assess its efficacy and value for future learners.
6. **Reflect**: Take stock of the process you have completed, your achievements and lessons learnt.

These phases are realised through a series of group activities: My Dream MOOC, Personas, Transition Matrix, Force Mapping, Brief, Features and intentions, Educational Instruments, Pattern mapping, Storyboarding, Evaluation rubrics, Presentations, Reflective discussion. Some of these are present in all our workshops, others are selectively used when appropriate. The MOOC design workshops put a strong emphasis on empathy. For this reason, even in a limited time format, we start by considering personas and their expected learning journeys (encoded as a transition matrix). Traditionally, empathic design demands extensive fieldwork [8]. Obviously, this is not possible in a one-off workshop. Instead, we focus on nurturing an empathic mindset. Thus, for example, when participants do not have the capacity to construct persoanas based on observations, we ask them to choose personas from a set we provide. Even in such a seemingly superficial setup, having a persona card before their eyes prompts participants to think, and feel, their design from a learners' perspective.

A detailed description of the activities, with links to supporting resources, is available under a creative commons licence at: https://www.academia.edu/26528408/Educational_Innovation_design_kit

## 4 Results

In 2015 and 2016 we ran 8 workshops, 3 of them small, private workshops (up to 10 participants), 5 open workshops at conferences (up to 50 participants). 2 of the private workshops led to successful MOOC/online course projects. One of these was the Amnesty Rights1X course, which had over 30,000 participants. The

**Table 1.** Participant feedback from MOOC design workshops (n=18)

| Question | Median | Average | SD |
|---|---|---|---|
| I am planning a MOOC, the workshop was valuable for structuring my thoughts | 4 | 3.39 | 1.42 |
| The workshop raised my awareness to the challenges of MOOC design | 4 | 3.83 | 0.79 |
| I will use some of the techniques and resources in my work | 4 | 4.11 | 0.9 |
| I would like to engage my team in a similar, but more detailed, design process | 4 | 3.44 | 1.1 |
| It was fun! | 5 | 4.5 | 0.79 |
| I liked .. Introduction | 4 | 3.67 | 1.19 |
| I liked .. Dream MOOC | 4 | 4 | 0.69 |
| I liked .. Challenge | 4 | 4.25 | 0.62 |
| I liked .. Personas | 5 | 4.39 | 0.78 |
| I liked .. Transition Matrix | 4 | 4.17 | 0.86 |
| I liked .. Feature Cards | 4 | 4.17 | 0.79 |
| I liked .. Design Patterns | 4 | 4.11 | 0.9 |
| I liked .. Storyboarding | 5 | 4.42 | 0.79 |
| I liked .. Evaluate | 4.5 | 3.92 | 1.38 |
| I liked .. Discussion | 4 | 4.17 | 1.04 |

third private workshop was held quite recently, and we are hoping to see follow-up work. Several additional workshops are scheduled for the spring/summer. Most workshops ran for either half a day or a full day, with exceptional cases being significantly shorter. One workshop was conducted online, all others were face to face. We surveyed the participants at 3 of the open workshops, and collected 18 responses. The median, average and standard deviation of the responses (on a likert scale of 0–5) are shown in Table 1 and Fig. 1.

To the question ‘Did you get what you came for?’, we received 10 strong positive responses, 3 positive or mildly positive responses, and 2 neutral responses.

Some of the specific comments we received highlighted issues related to empathy: “I especially liked the design patterns and the concept of personas”, “(My biggest takeaway is ...) Do take the client and his/her context as the starting point”, “(My biggest takeaway is ...) The viewpoint that you start with personas and the transition matrix”.

Interestingly, several participants noted: “I think everything that we discussed can be applied to ‘normal’ online courses, too”.

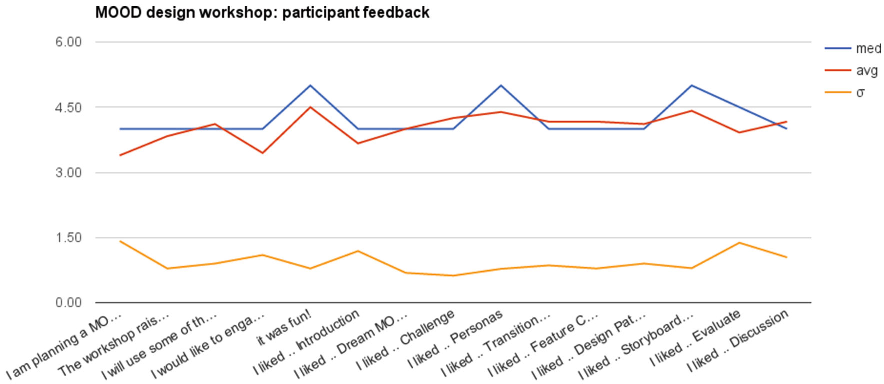



**Fig. 1.** Participant feedback from MOOC design workshops (n = 18)

## 5 Conclusions

The MOOC design workshops are designed to introduce participants to a learner-centered empathic approach to designing MOOCs. This process is rooted in a deep cognitive and emotional understanding of the target learners in the MOOC as holistic learners, their current intentional, physical and social state, the desired effect of the MOOC, and the assets and constraints that shape their zone of possibilities. Analysis of the feedback from the workshops we had surveyed suggests that participants recognise the main messages of the workshop, and acknowledge their value. This analysis is confirmed by the observed outcomes in the MOOCs that have emerged from the workshop and follow-up design consultancy.

The workshops draw on the outputs of the MOOC design patterns project, and are based on the Learning Design Studio framework. They extend this framework by adding a stronger emphasis on empathy, through the use of personas, transition matricis, and force maps.

The workshop design has shown excellent adaptability it is flexible enough to run in as little as 75 min to a whole day. We are planning to expand this to a *MOOC design and development sprint* where by the prototyping step (mentioned earlier) could be brought into the process over an intensive 3-day session which incorporated digital content developers and media specialists to realise the projects on a designated platform.

The resources we use in our workshops are available under a creative commons licence at: http://moocsandco.com/kit.

## References

1. Aaen, J.H., Nørgård, R.T.: Participatory academic communities: a transdisciplinary perspective on participation in education beyond the institution. Conjunctions. Transdisciplinary J. Cult. Participation **2**, 67–98 (2016)
2. Asensio-Péerez, J.I., Dimitriadis, Y., Prieto, L.P., Hernáandez-Leo, D., Mor, Y.: From idea to vle in half a day: metis approach and tools for learning co-design. In: Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality, TEEM 2014, pp. 741–745. ACM, New York (2014)
3. Asensio-Prez, J.I., Dimitriadis, Y., Hernndez-Leo, D., Pozzi, F.: Teacher continuous professional development and fulllifecycle learning design: first reflections. In: Garreta-Domingo, M., Sloep, P., Stoyanov, S., Hernndez-Leo, D., Mor, Y. (eds.) Proceedings of the Workshop Design for Learning in Practice, EC-TEL, Toledo, 18 September, 2015
4. Conole, G.: The 7Cs of learning design a new approach to rethinking design practice. In: Proceedings of the 9th International Conference on Networked Learning, pp. 502–509 (2014)
5. Dewey, J.: Experience and Education. Simon and Schuster, New York (1938)
6. Gagnon, C., Ct, V.: Learning from others: a five years experience on teaching empathic design. In: Lim, Y.K., Niedderer, K., Redstrm, J., Stolterman, E., Valtonen, A. (eds.) Proceedings of Design Research Society Biennial International Conference DRS 2014: Designs Big Debates, pp. 16–19. Ume Institute of Design, Ume University, Ume, Sweden, June 2014
7. Köppen, E., Meinel, C.: Empathy via design thinking: creation of sense and knowledge. In: Plattner, H., Meinel, C., Leifer, L. (eds.) Design Thinking Research. Understanding Innovation, pp. 15–28. Springer, Switzerland (2015)
8. Leonard, D., Rayport, J.F.: Spark innovation through empathic design. Harvard Bus. Rev. **75**, 102–115 (1997)
9. Littlejohn, A., Milligan, C.: Designing MOOCs for professional learners: tools and patterns to encourage self-regulated learning. e-Learning Papers. **42** (2015)
10. Mor, Y., Craft, B.: Learning design: reflections on a snapshot of the current landscape. Res. Learn. Technol. **20**, 85–94 (2012)
11. Mor, Y., Craft, B., Hernndez-Leo, D.: Editorial: the art and science of learning design. Res. Learn. Technol. **21**, 1–8 (2013)
12. Mor, Y., Mogilevsky, O.: Learning design studio: educational practice as design inquiry of learning. In: Hernández-Leo, D., Ley, T., Klamma, R., Harrer, A. (eds.) EC-TEL 2013. LNCS, vol. 8095, pp. 233–245. Springer, Heidelberg (2013)
13. Postma, C.E., Zwartkruis-Pelgrim, E., Daemen, E., Du, J.: Challenges of doing emphatic design: experiences from industry. Int. J. Des. **6**(1), 59–70 (2012)
14. Salmon, G., Gregory, J., Lokuge Dona, K., Ross, B.: Experiential online development for educators: the example of the carpe diem MOOC. Br. J. Educ. Technol. **46**(3), 542–556 (2015)

15. Voogt, J., Westbroek, H., Handelzalts, A., Walraven, A., McKenney, S., Pieters, J., de Vries, B.: Teacher learning in collaborative curriculum design. Teach. Teach. Educ. **8**, 1235–1244 (2011)
16. Warburton, S., Mor, Y.: A set of patterns for the structured design of MOOCs. Open Learn. J. Open Distance e-Learning **30**(3), 1–15 (2015)
17. Warburton, S., Mor, Y.: Double loop design: configuring narratives, patterns and scenarios in the design of technology enhanced learning. In: Mor, Y., Maina, M., Craft, B. (eds.) The Art and Science of Learning Design. Sense publishers, Boston (2015)

# OERauthors: Requirements for Collaborative OER Authoring Tools in Global Settings

Irawan Nurhas(✉), Jan M. Pawlowski, Marc Jansen, and Julia Stoffregen

Computer Science Institute,
Ruhr West University of Applied Sciences, Bottrop, Germany
{irawan.nurhas,jan.pawlowski,marc.jansen,
julia.stoffregen}@hs-ruhrwest.de

**Abstract.** Open Educational Resources (OER) intend to support access to education for everyone. However, this potential is not fully exploited due to various barriers in the production, distribution and the use of OER. In this paper, we present requirements and recommendations for systems for global OER authoring. These requirements as well as the system itself aim at helping creators of OER to overcome typical obstacles such as lack of technical skills, different types of devices and systems as well as the cultural differences in cross-border-collaboration. The system can be used collaboratively to create OER and supports multi-languages for localization. Our paper contributes to facilitate global, collaborative e-Learning and design of authoring platforms by identifying key requirements for OER authoring in a global context.

**Keywords:** Authoring tool · Collaborative authoring · Global collaboration · Open educational resources · OER

## 1 Introduction

The production of E-learning courses is in many cases rather laborious and costly. The Open Educational Resources (OER) – learning resources with an open license – are an alternative for developers and users. In particular, in the context of global settings, Open Educational Resources (OER) are a solution providing free access to digital educational materials for everyone in different regions or countries [17]. Despite the potentials of OER, several barriers have been identified which prevent the use of OER [20]. One important barrier is the difficulty to produce OER by using Authoring tools (AT) collaboratively in global settings. It is a key challenge to solve in order to improve the quality and the success of OER [22]. However, this key aspect is rarely studied and there are no clear requirements for AT to create Learning Objects (LO). Apart from the global aspect, the growing numbers and variety of mobile devices brings up further challenges to develop widely adopted OER. This study attempts to close this research gap and answer the question: what are the requirements for OER authoring in a cross-border, collaborative environment? To answer this question, this study presents a literature review as starting point as well as the research methodology. Afterwards, we


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 460–465, 2016.
DOI: 10.1007/978-3-319-45153-4_43

present the main results: requirements towards this class of systems including the evaluation of the prototype.

## 2 Background: Open Education in the Global Context

In this chapter, we provide a brief literature review as the background for our research focusing on the key concept of Open Educational Resources (OER) and the corresponding barriers. Many definitions of OER have emerged with different focuses since UNESCO had coined this term. A broad definition by [22] defines OER as "…any digital resource for educational purposes that can be used, distributed and redistributed freely". Three components – learning contents, tools/software, and attached licenses [10] – need to be considered as they significantly influence the adoption of OER.

Barriers of OER have been studied in-depth. Among the most important ones are challenges to apply OER which are culturally distant (unfamiliar values, symbols, beliefs etc.), impact of geographical distance and lack of trust towards authors of LO. Part of the problem is that OER do not give enough information on the context where it was created and used, and availability of native language to encourage online collaboration [20, 21].

Apart from socio-technical barriers, the huge market of diversified mobile devices delivers many challenges for OER as well. According to [31], they can be addressed by using HTML5 which can be used to support collaborative environments. Cloud-based applications could provide accessibility and interoperability [12]. In this paper, barriers and challenges will be used as a reference to create initial requirements for the system design and development. The following chapter will elaborate on the methodology for this aim.

## 3 Methodology

To elaborate requirements of an authoring system for collaborative, international OER processes and systems, we followed a case study-design [30] including two countries, Germany and Indonesia. In this combination, a variety of cultural and social barriers can be exemplary identified. The case also allows comparison and embedding of results with previous studies in the field [23]. To identify the main artefacts (requirements for AT), we oriented on principles of Design Science Research [9]. As an initial step to find barriers, a literature review was performed. An ideal set of literature based on [28] was used to classify and categorize the literature. The main points of literatures were summarized by preparing a matrix of concepts [27]. In a second step of the creation of the software artifact, we chose a qualitative data gathering method which allowed obtaining and refining more in-depth information about barriers and requirements of AT in a global collaboration. Interviews were held with one participant (male, 12 years-teaching-experience(YTE)) from Germany and two participants from Indonesia (male, 12 YTE; female, 3 YTE). In a third step, the evaluation was performed by asking five users of the prototype 'OERauthors'. The questions were grouped into two types: one addresses the importance of requirements (IoR-scale) and the other represents the System Usability Scale (SUS) [3]. Respondents were all male (3 lecturers from Indonesia and 2 from Germany). The sample size of 5 persons is the minimum number of users to test the usability of a system [15]. A 5-point Likert scale was used to quantify statements. Results of this ranking were assessed with Cronbach's Alpha as commonly conducted in studies in the field [24].

## 4 Requirements for Global OER Creation

Requirements for a Global OER Authoring Tool (AT) were derived by analyzing the results from literature review and interviews. Additionally, observations of existing authoring systems for Learning Objects (LO) respectively OER provide a general process of an AT that can be used to derive functional requirements of the system. We distinguish functional requirements (which are to a certain extent similar to all authoring systems), non-functional requirements as well as specific requirements necessary for using OER in a global context. Table 1 shows the requirements of the system (including at least one source of reference).

**Table 1.** System requirements

| **Functional requirements**[b] | | |
|---|---|---|
| 1. Create a new LO<br>2. Delete page of LO<br>3. Add new metadata<br>4. Edit metadata<br>5. Delete metadata<br>6. Reuse a revision<br>7. Edit LO project<br>8. Add page to LO | 9. Send message<br>10. Add to-do list<br>11. Edit to-do list<br>12. Define the license of LO<br>13. Download project | 14. Delete a revision of LO<br>15. Add a block of template to page<br>16. Delete the block from the page<br>17. Arrange position of the block<br>18. Add a revision or version of LO<br>19. Change application setting (adaptation of symbol, colors, layout and language) |
| **Nonfunctional requirements** | | |
| 1. The system should be mobile friendly [18][a]<br>2. The system should integrate with social media/networks [4][a]<br>3. The system should provide consistent size and style of LO [26][b]<br>**4. The system should offer levels of usability that author with no familiarity or little knowledge with computer and editing of content [2]**[a]<br>5. Authors should be able to produce LO after completing the short training [2][a]<br>6. The user interface of LO should be easy to navigate [13]<br>7. The user interface of LO should be load quickly [16]<br>8. The system should be able to identify mistakes and error prevention [2]<br>9. The system should have system status [2][b]<br>10. LO should have uniform/consistent editorial tone across object [13, 14]<br>11. The system should provide easy to use direct manipulation interface [7][a]<br>12. The system should have active community [22]<br>13. LO should be free from technical problems [8][b]<br>14. The system should provide share ownership [6][a]<br>**15. LO should be delivered in standalone possibility [19]**[a]<br>**16. The capability of a system to handle high concurrent editing (i.e. any number of users should be able to edit concurrently a shared LO [11]**[b] | | |

(*Continued*)

**Table 1.** (*Continued*)

| Nonfunctional requirements for global aspect |
|---|
| 17. The system should be free to use [29][a] |
| 18. The system, output or LO should be platform independent [14] |
| 19. The system, output or LO should have open formats and standard [18][b] |
| 20. The system should support localization, internationalization by paying attention on user interface such as symbol, colors, language, layouts [14][c] |
| 21. The system should provide cloud based solution for hosting project [7][b] |
| 22. The system should provide real-time typed conversation [5][c] |

[a]interview; [b]observation; [c]interview and observation; **bold**: high priority requirements

The symbols, colors, language, and layouts of the interface are an essential element of the application. They contribute to make users from different cultures feel comfortable when using the system and to maximize a positive user experience that improves the usability of the system. The requirements were implemented into a prototype, called OERauthors. The corresponding system, which uses HTML5 and Operational Trans-formation [25] to fulfil nonfunctional requirements number 1 and 16 (Table 1) shows an implementation of those requirements as an example how future AT should be created. To illustrate the results, the user interfaces of the main page editor of the systems for Indonesia and Germany as well as the responsive design can be seen in Fig. 1:

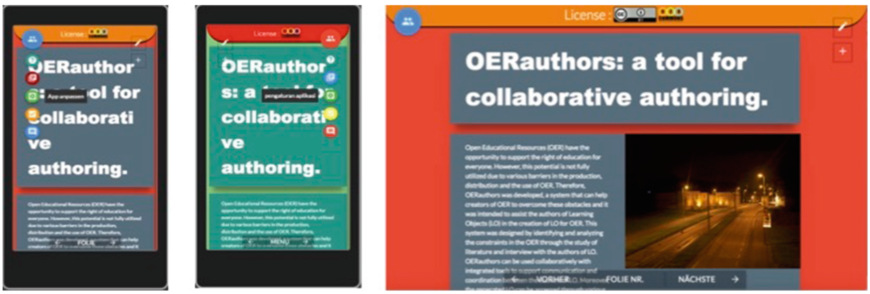



**Fig. 1.** User interface design of the main page editor. Right-image: for German users; Center-image: for Indonesian users; Left-image: display on widescreen/laptop (Color figure online)

Figure 1 shows the main page editor. The editable license area is displayed on top of the interface. The button with symbol of two people is the collaboration button (red-located on the right side for Indonesia; blue-located on the left side for Germany). This example illustrates a subset of how those requirements were implemented.

Subsequently, requirements were evaluated. Based on the user rating, the most important requirements belong to nonfunctional requirements: see nr. 15, 18, 17, 4, 5, 16 and 21 in Table 1. Concerning usability (SUS = 56, $\alpha = 0.61$) OERauthors was rated as good [1]. There is space for improvement to be rated as excellent (SUS score = 73).

Yet, the SUS-score is adequate given the low number of respondents. Even though the threshold ($\alpha$ = 0.70) is not passed, this score belongs to a category that can be used for analysis [1]. This is supported by the positive reliability of IoR-scale ($\alpha$ = 0.92). In the following, all results and implications will be discussed.

Summarizing the findings, main system requirements for global, mobile OER authoring were elaborated from the literature (Table 1). In contrast to previous studies on barriers [20–22], we have transferred barriers to concrete system requirements. These requirements can serve as a base for further research regarding generic requirements for OER and AT in global settings. Based on this iterative analysis, our identified requirements should be applied beyond the system OERauthors; they can be generalized and implemented in global OER systems. For instance, AT-developer should pay attention to icons, symbols and layouts for different countries. Independent and multi-platform LO should be supplied to areas without internet connection, low level or bad internet connectivity.

## 5 Conclusion

Summarizing the key results of our research, we have provided a collection of requirements for OER authoring systems on a fine-grained level enriched by requirements for the global context. Further research is needed regarding the evaluation in mobile authoring environments as well as the level of importance of each requirement in different contexts. Yet, our results offer guidance for developers who aim at generating a global-friendly collaborative authoring system for OER.

## References

1. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective rating scale. J. Usability Stud. **4**(3), 114–123 (2009)
2. Battistella, P.E., von Wangenheim, A., von Wangenheim, C.G.: Evaluation of free authoring tools for producing SCORM-conform learning objects. IEEE Technol. Eng. Educ. **5**(4), 15–26 (2011)
3. Brooke, J.: SUS-A quick and dirty usability scale. In: Jordan, P.W., Thomas, B., McClelland, I.L., Weerdmeester, B. (eds.) Usability Evaluation in Industry, pp. 4–7. CRC Press, New York (1996)
4. Chiappe, A., Lee, L.L.: Understanding open teaching: difficulties and key issues. In: Recent Advances in Education and Educational Technology, WSEAS (2015)
5. Dabbagh, N.H., Schmitt, J.: Redesigning instruction through web-based course authoring tools. Educ. Media Int. **35**(2), 106–110 (1998)
6. Davidson, J., Deus, L.: A Case Study in Technology Transfer of Collaboration Tools. The Edge, Singapore (1998)
7. Fidas, C., Sintoris, C., Yiannoutsou, N., Avouris, N.: A survey on tools for end user authoring of mobile applications for cultural heritage. In: Proceedings of 6th International Conference on Information, Intelligence, Systems and Application, Corfu (2015)
8. Georgieva, E.S., Smrikarov, A.S., Georgiev, T.S.: Evaluation of mobile learning system. Procedia Comput. Sci. **3**, 632–637 (2011)

9. Hevner, A., Chatterjee, S.: Design Science Research in Information Systems, pp. 9–22. Springer, Heidelberg (2010)
10. Hylén, J.: Open educational resources: Opportunities and challenges. Proc. Open Educ. **2006**, 49–63 (2006)
11. Ignat, C., Norrie, M.C.: Tree-based model algorithm for maintaining consistency in real-time collaborative editing systems. In: CSCW 2002. IEEE (2002)
12. Jansen, M., Baloian, N., Bollen, L.: Cloud services for learning scenarios: widening the perspective. In: WCLOUD 2012, Antigua, Guatemala, pp. 33–37 (2012)
13. Koohang, A., Floyd, K., Stewart, C.: Design of an open source learning objects authoring tool–the LO creator. Interdisc. J. E-Learn. Learn. Ob-j. **7**(1), 111–123 (2011)
14. Longmire, W.: A primer on learning objects. Learn. Circ. **1**(3) (2000)
15. Nielsen, J., Landauer, T.K.: A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT 1993 and CHI 1993 Conference, pp. 206–213 (1993)
16. Ozdamli, F., Cavus, N.: Basic elements and characteristics of mobile learning. Procedia Soc. Behav. Sci. **28**, 937–942 (2011)
17. Pawlowski, J.M.: Global Open Education: A Roadmap for Internationalization (2013)
18. Pawlowski, J.M., McGreal, R., Hoel, T., Treviranus, J.: Open educational resources and practices for educational cross-border collaboration. In: UNESCO Workshop at the World Summit on the Information Society, Geneva (2012)
19. Penman, C.: Collaborative production of learning objects on French literary works using the LOC software. In: Borthwick, K. et al. (eds.) 10 Years of the LLAS eLearning Symposium: Case Studies in Good Practice, p. 117 (2015)
20. Pirkkalainen, H., Jokinen, J.P., Pawlowski, J.M., Richter, T.H.: Removing the barriers to adoption of social OER environments. In: Zvacek, S., Restivo, M.T., Uhomoibhi, J., Helfert, M. (eds.) Computer Supported Education. Communications in Computer and Information Science, vol. 510, pp. 19–34. Springer, Heidelberg (2015)
21. Pirkkalainen, H., Pawlowski, J.: Global social knowledge management: from barriers to the selection of social tools. Electron. J. Knowl. Manage. **11**(1), 3–17 (2013)
22. Pirkkalainen, H., Pawlowski, J.M.: Open educational resources and social soft-ware in global e-learning settings. Sosiaalinen Verkko-oppiminen. IMDL, pp. 23–40 (2010)
23. Richter, T., McPherson, M.: Open educational resources: education for the world? Distance Educ. **33**(2), 201–219 (2012)
24. Santos, J.R.A.: Cronbach's alpha: A tool for assessing the reliability of scales. J. Extension **37**(2), 1–5 (1999)
25. Sun, C., Ellis, C.: Operational transformation in real-time group editors: issues, algorithms, and achievements. In: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, pp. 59–68 (1998)
26. Watson, J., Dickens, A., Gilchrist, G.: The LOC tool: creating a learning object authoring tool for teachers. In: EDMEDIA, AACE, pp. 4626–4632 (2008)
27. Watson, R.T.: Introducing MISQ review-A new department in MIS Quarterly. MIS Q. **25**(1), 103–106 (2001)
28. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q. **26**, 13–23 (2002)
29. Wiley, D.: A Modest History of OpenCourseWare. Autounfocus blog, Chicago (2003)
30. Yin, R.K.: Case Study Research: Design and Methods. Sage publications, Thousand Oaks (2013)
31. Zbick, J., Jansen, M., Milrad, M.: Towards a web-based framework to support end-user programming of mobile learning activities. In: Proceedings of ICALT 2014, pp. 204–208 (2014)

# Virtual Reality for Training Doctors to Break Bad News

Magalie Ochs[1(✉)] and Philippe Blache[2]

[1] Laboratoire des Sciences de l'Information et des Systèmes, LSIS, UMR7296, Aix Marseille Université, CNRS, ENSAM, 13397 Marseille, France
magalie.ochs@lsis.org
[2] Laboratoire Parole et Langage, LPL, UMR7309, CNRS, ENSAM, Université de Toulon, 13397 Marseille, France
philippe.blache@lpl.fr

**Abstract.** The way doctors deliver bad news has a significant impact on the therapeutic process. In this paper, we present our overall project to develop an embodied conversational agent simulating a patient to train doctors to break bad news. The embodied conversational agent is incorporated in an immersive virtual reality environment (a CAVE) integrating several sensors to detect and recognize in real time the verbal and non-verbal behavior of the doctors interacting with the virtual patient. The virtual patient will adapt its behavior depending on the doctor's verbal and non-verbal behavior. The methodology used to construct the virtual patient behavior model is based on a quantitative and qualitative analysis of corpus of doctors training sessions.

**Keywords:** Embodied conversational agent · Virtual reality · Virtual patient · Training platform

## 1 Introduction

The way doctors deliver bad news has a significant impact on the therapeutic process: disease evolution, adherence with treatment recommendations, litigation possibilities (Andrade et al. 2010). However, both experienced clinicians and medical trainees consider this task as difficult, daunting, and stressful. Nowadays, training health care professional to break bad news, recommended by the French Haute Autorité de la Santé (HAS), is organized as workshops during which doctors disclose bad news to actors playing the role of patient. This training solution requires a huge amount of human resources as well as high level of preparation (each 30 mn session requires an hour of preparation), not to speak about funding.

In this project, we aim at developing an embodied conversational agent (ECA) simulating a patient. Such a platform would play a decisive role for institutions involved in training (hospitals, universities): the needs concern potentially thousands of doctors/students. Organizing such training at this scale is not realistic with human actors. A virtual solution would be then an adequate answer.

Our objective is to develop an *immersive* platform that enables doctors to train to break bad news with a virtual patient. For this purpose, we adopt a multidisciplinary approach in the project gathering computer scientists, linguists, psychologists and doctors. Moreover,


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 466–471, 2016.
DOI: 10.1007/978-3-319-45153-4_44

we adopt a *corpus-based methodology* to model the virtual agent. One goal in this project with this multidisciplinary and corpus-based approach is to try to simulate as realistic as possible the environment of breaking bad news and the virtual patient behavior.

The objective of the paper is to present the overall project and more particularly the global methodology to develop such a training platform. In the following, after a presentation of a state of art in this domain (Sect. 2), we present the corpus-based approach used to model the virtual patient (Sect. 3) and we introduce the training platform and its different components (Sect. 4).

## 2 State of Art

For several years, there has been a growing interest in Embodied Conversational Agents (ECAs) to be used as a new type of human-machine interface. ECAs are virtual entities, able to communicate verbally and nonverbally. They can attract and maintain the attention of users in an interaction, to make the interaction more expressive and more socially adapted. Indeed, research has shown that embodied conversational agents are perceived as social entities leading users to show behaviors that would be expected in human-human interactions (Krämer 2005).

Moreover, recent research showed that virtual agents could help human beings improve their social skills. For instance, in (Finkelstein et al. 2013), a virtual agent is used to train kids to adapt their language register to the situation. In the European project TARDIS (Anderson et al. 2013), an ECA endowed the role of a virtual recruiter is used to train young adults to job interview. This research shows that embodied conversational agent can be used for social training since users will react to the ECA in a similar way that to another person and the socio-emotional responses of the agents will help them practice and improve their social skills.

Several ECAs embodied the role of virtual patients have already been proposed for use in clinical assessments, interviewing and diagnosis training (Andrade et al. 2010; Kenny et al. 2008; Lok et al. 2006). Indeed, previous research has shown that doctors demonstrate non-verbal behaviors and respond empathetically to a virtual patient (Deladisma et al. 2006). In this domain, the research has mainly focused on the anatomical and physiological models of the virtual patient to simulate the effects of medical interventions or on models to simulate particular disorder. For instance, *Justina* is a virtual patient simulating Post Traumatic Stress Disorder (PTSD) to train medical students' interview skills and diagnostic acumen for patient with such disorder (Kenny et al. 2008). DIANA (DIgital ANimated Avatar) is a female virtual character playing the role of a patient with appendicitis (Lok et al. 2006). In the eViP European project (http://www.virtualpatients.eu), the objective is specifically to develop a large amount of virtual patients simulating different pathologies. In our project, we focus on a virtual patient to train doctors to deliver bad news.

A first study (Andrade et al. 2010) has analyzed the benefits of using a virtual patient to train doctors to deliver bad news. The results show significant improvements of the self-efficacy of the medical trainees. The participants consider the virtual patient as "excellent instructional method for learning how to deliver bad news". The major limit of the proposed system, highlighted by the participants, is the lack of non-verbal behaviors of the patients simulated in the limited environment Second Life (Linden Labs, San Francisco, CA). Our objective in this project is to simulate the non-verbal expression of the virtual patient to improve the believability of the virtual character and the immersive experience of the doctor. Indeed, as shown in (Witmer and Singer 1998), the realism of the environment as well as the social behavior of the virtual characters improve the user experience in the virtual environment. Consequently, we suppose that in the context of a simulation of breaking bad news in a virtual environment, a particular attention on the modeling of the virtual character's behavior (verbal and non-verbal) could lead to a better performance of the trainee.

Most of the embodied conversational agents used for health applications have been integrated in 3D virtual environment on PC. Virtual reality in health domain is particularly used for virtual reality exposure therapy (VRET) for the treatment for anxiety and specific phobias. For instance, people with a fear of public speaking may speak to an audience of virtual characters in virtual reality environment to reduce their anxiety in reality (Parsons and Rizzo 2008). In our project, in order to offer an immersive experience to the doctor, we have integrated the virtual patient in a virtual reality environment.

Moreover, a virtual patient should be able to display verbal and non-verbal reactions appropriately according to the doctor's behavior. During an interaction, the interlocutors in fact coordinate or align their verbal and non-verbal behavior (e.g. feedback, mimicry). According to the Communication Accommodation theory (CAT), the interlocutors adapt the coordination of their behavior to express different social attitudes (Giles et al. 1991). Recent research in human-machine interaction confirms this hypothesis: the coordination of the virtual agent's behavior on the one of the user or the divergence of its behavior reflects the agent's social attitude (e.g. appreciation, cold, mutual understanding, etc.) (e.g. Bailenson et al. 2005). In the project described in this paper, sensors will be used to automatically detect in real-time the verbal and non-verbal behavior of the doctors during their interaction with the virtual patient (Sect. 4). These inputs will then be used by the virtual patient to coordinate its behavior on the doctor's one depending on the social attitude to express. The methodology used to define the virtual agent's behavior is based on the analysis of a corpus of doctor-patient interaction to extract rules on its verbal and non-verbal reactions.

## 3 A Corpus-Based Approach to Simulate a Virtual Patient

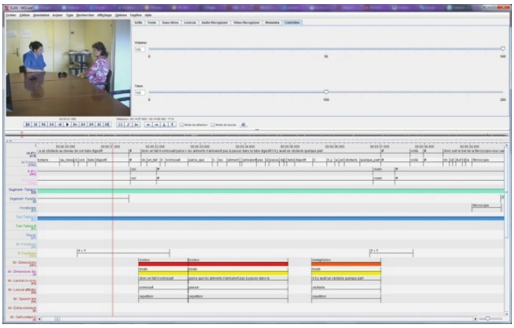


**Fig. 1.** Corpus annotation with Elan

In order to model the behavior of the virtual patient (verbal and non-verbal), we propose a corpus-based approach to identify precisely the reactions of the patient (when and what) to replicate it accordingly on the virtual one. For ethical reasons, it is not possible to videotape real breaking bad news situations. Instead, simulations are organized with actors playing the role of the patient. A corpus of such interactions has been collected in different medical institutions (the Institut Paoli Calmette and the hospital of Angers). Simulated patients are actors trained to play the most frequently observed patients reactions (denial, shock…). The actor follows a *pre-determined scenario*. The doctor (*i.e.* the trainee) receives details of a medical case before the simulated interaction starts (patient medical history, family background, surgery, diagnosis, etc.). On average, a simulated consultation lasts 30 min. The collected corpus is currently composed of 23 videos of patient-doctor interaction with different scenario (e.g. patient aggressive or accommodating).

These simulated interactions of the collected corpus are transcribed and annotated (with Elan software – Fig. 1) on *several levels*: at the discourse level (*e.g.* dialog phases, turn-taking) and at non-verbal levels (e.g. feedback, gaze, gestures, etc.). The coding scheme has been defined based on a preliminary analysis of the corpus (Saubesty and Tellier 2015). Both the doctor and the patient verbal and non-verbal behavior are annotated and transcribed. By this way, we can analyze precisely the coordination of their behavior and identify when the virtual agent should trigger which behavior during the interaction.

The method used to extract information from the annotated corpus is based on a *multidisciplinary approach* combining (1) a manually analysis of the data by linguists for a qualitative analysis and (2) automatic requests (using the SPPAS software (Bigi 2015)) and datamining algorithms (e.g. Rabatel et al. 2010) on the data for a more quantitative study performed by computer scientists. Our objective by this ongoing analysis of the corpus is to extract probabilistic rules on the patient behaviors in order to trigger the appropriate virtual patient's behavior during the interaction with some variability.

## 4 Virtual Reality Platform of Training

The tool we use to animate the virtual patient is the Greta System (Pelachaud 2009). Greta offers several modules, each dedicated to particular functionality to both design new facial expressions and gestures and to animate in real-time 3D virtual agents in virtual environments. The lexicon of gestures of Greta is enriched by specific gestures and facial expressions of patients identified in the corpus. The virtual patient has been integrated in the CRVM (Centre de Réalité Virtuelle de Marseille, platform of the ISM partner). The visualization

**Fig. 2.** Virtual reality environment for training

system consists of a high-end platform called “CAVE™”. CAVE™ is composed of four projection screens: frontal, ground and lateral projections. Each frontal and lateral screen has a projection surface of 3 meters wide by 4 meters high (Fig. 2). The speech of the doctor is recognized by an automatic speech recognition system (Nocera *et al.* 2002). We are improving the system by learning the lexicon and language model of the speech recognition system on the transcript corpus of patient-doctor interactions. For this project, we have defined a specific use case that reflects a real situation in which doctors may train their social competences in delivering bad news. It is based on real scenario of training including some possible variations (e.g. patient aggressive or accommodating). The dialog model of the patient is based on this scenario. We are defining the dialog rules that depend on the profile of the patient (aggressive or accommodating) and the discourse of the trainee (e.g. level of details, use of medical terms, etc.). The dialog rules and more generally the behavior of the virtual patient are based on the corpus analysis (as described in the previous section).

To consider the non-verbal behavior of the trainee, on the table in the CAVE, we will install a Kinect[1] to automatically detect some non-verbal signals of the doctors (e.g. the gaze direction, head movements, posture, and facial expressions). The objective is to coordinate (e.g. the virtual patient smile back to the doctor) or in contrary discoordinate (e.g. the virtual patient is nodding its head in agreement whereas the doctor shakes her head) the non-verbal behavior of the virtual patient on those of the trainee depending on the attitude that the virtual patient should convey.

## 5 Conclusion

In this paper, we have presented a project as a whole that aims at developing a virtual reality platform to train doctors to break bad news with virtual patient. One challenge to obtain an efficient training it to simulate as realistic as possible the behavior of the virtual patient: both its verbal behavior but also non-verbal (gaze, facial expressions, gestures, etc.). The methodology presented in the paper to achieve this goal is based on a multidisciplinary analysis of audio-visual corpus of doctor-patient interaction in the context of delivering bad news. Moreover, to replicate multimodal interaction, the immersive platform will be endowed with several sensors to detect the verbal and non-verbal behavior of the doctor and to coordinate the virtual patient behavior accordingly.

Currently, we are working more particularly on the development of a stochastic model of the virtual patient behavior to automatically determine the appropriate verbal and non-verbal behavior of the virtual patient given the doctor’s behavior.

[1] Note that in order to simulate the alignment of the behaviors, we are interested in the detection of social signals (such as gaze, smile, or head nodes) and not interpreted state such as emotions.

## References


Anderson, K., et al.: The TARDIS Framework: Intelligent Virtual Agents for Social Coaching in Job Interviews. In: Reidsma, D., Katayose, H., Nijholt, A. (eds.) ACE 2013. LNCS, vol. 8253, pp. 476–491. Springer, Heidelberg (2013)

Andrade, A.D., Bagri, A., Zaw, K., Roos, B.A., Ruiz, J.G.: Avatar-mediated training in the delivery of bad news in a virtual world. J. Palliat Med. **13**, 1415–1419 (2010)

Bailenson, J.N., Swinth, K.R., Hoyt, C.L., Persky, S., Dimov, A., Blascovich, J.: The independent and interactive effects of embodied agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. Presence Teleoper. Virtual Environ. **14**, 379–393 (2005)

Bigi, B.: SPPAS - Multi-lingual approaches to the automatic annotation of speech the phonetician. International Society of Phonetic Sciences, ISSN 0741-6164, Number 111–112/2015-I-II, pp. 54–69 (2015)

Deladisma, A.M., Cohen, M., Stevens, A., et al.: Do medical students respond empathetically to a virtual patient? In: Association for Surgical Education Meeting (2006)

Finkelstein, S., Yarzebinski, E., Vaughn, C., Ogan, A., Cassell, J.: The effects of culturally congruent educational technologies on student achievement. In: Lane, H., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS, vol. 7926, pp. 493–502. Springer, Heidelberg (2013)

Giles, H., Coupland, N., Coupland, J.: Accommodation theory: communication, context and consequence. J. Context Accommodation Dev. Appl. Sociolinguistics, 27 (1991). Gratch *et al.* 2006

Kenny, P., Parsons, T.D., Gratch, J., Rizzo, A.A.: Evaluation of Justina: a virtual patient with PTSD. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 394–408. Springer, Heidelberg (2008)

Krämer N., Iurgel I., Bente G.: Emotion and motivation in embodied conversational agents. In: Canamero, L. (ed.) Proceedings of the Symposium 'Agents that Want and Like', AISB 2005, pp. 55–61, Hatfield, SSAISB (2005)

Lok, B., Ferdig, R.E., Raij, A., Johnsen, K., Dickerson, R., Coutts, J., Stevens, A., Lind, D.S.: Applying virtual reality in medical communication education: current findings and potential teaching and learning benefits of immersive virtual patients. Virtual Real. **10**, 185–195 (2006)

Nocera, P., Linares, G., Massonié, D., Lefort, L.: Phoneme lattice based A* search algorithm for speech recognition. In: Sojka, P., Kopeček, I., Pala, K. (eds.) TSD 2002. LNCS (LNAI), vol. 2448, pp. 301–308. Springer, Heidelberg (2002)

Parsons, T.D., Rizzo, A.A.: Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J. Behav. Ther. Exp. Psychiatry **39**(3), 250–261 (2008)

Pelachaud, C.: Studies on gesture expressivity for a virtual agent. Speech Commun. **51**, 630–639 (2009). special issue in honor of Björn Granstrom and Rolf Carlson

Rabatel, J., Bringay, S., Poncelet, P.: Contextual sequential pattern mining. In: IEEE International Conference on Data Mining Workshops, pp. 981–988, December 2010

Saubesty, J., Tellier, M.: Multimodal analysis of hand gesture back-channel feedback. In: Gesture and Speech in Interaction, Nantes, France (2015)

Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. Presence Teleoper. Virtual Environ. **7**(3), 225–240 (1998)

# User Motivation and Technology Acceptance in Online Learning Environments

Maxime Pedrotti[1(✉)] and Nicolae Nistor[1,2]

[1] Ludwig-Maximilians-Universität München, Munich, Germany
{maxime.pedrotti,nic.nistor}@lmu.de
[2] Walden University, Minneapolis, USA

**Abstract.** Research on technology acceptance in educational contexts often shows little or no influence of user acceptance on use intention or use behavior. While recent attempts to factor in learners' motivation appear promising, the problem of limited explanatory value of technology acceptance models remains. This paper further explores the relationship between motivational and acceptance factors in different learning contexts. Data (N = 673) from four studies conducted among users of two online learning environments at a major university in Germany are analyzed using a combined data set containing items relating to user motivation (according to Self-Determination Theory) and technology acceptance (according to the Unified Theory of Acceptance and Use of Technology). The data show significant differences in acceptance and motivational levels between these four groups, as well as a connection between acceptance and motivation. Implications of these results include a recommendation to revisit UTAUT assumptions and variables in future research.

**Keywords:** Motivation · Technology acceptance · Learning environment · UTAUT · SDT

## 1 Introduction

While learning technologies such as online lecture videos (OLV) or online learning environments such as Moodle have received much attention in recent years, particularly with recent developments of Massive Open Online Courses (MOOCs) and with universities trying to provide a modern learning environment, the question of how (potential) users view any new learning technology, what their attitudes towards new technologies are, and how these factors influence their use behavior, remains a difficult subject. A very popular acceptance approach is taken from Information Systems (IS) research. However, unlike in IS, technology acceptance models regularly fail to be reproduced in educational contexts. This paper aims to address this problem by refining the view on motivational variables, which have been shown to have a strong influence on learning activities. By analyzing data from four different studies conducted at a major university in Germany, we hope to show the importance of including additional variables measuring a person's motivation, so as to better understand a person's attitudes towards learning technologies, and how learning behavior can be supported by such technologies.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 472–477, 2016.
DOI: 10.1007/978-3-319-45153-4_45

## 2 Technology Acceptance in Educational Contexts

One of the most prominent models to explain the use of technological solutions in professional contexts was proposed by Venkatesh et al. in 2003 with the Unified Theory of Acceptance and Use of Technology (UTAUT) [1]. According to their research, which they base on various already popular and established technology acceptance models, four main factors influence a person's intention to actually make use of a proposed technological tool: performance expectancy (PE), effort expectancy (EE), facilitating conditions (FC) and social influence (SI). The first two variables are best described as expectations a person may have towards the benefits gained from using the technology at hand. The more someone hopes to achieve through the use of the tool, the higher the value of PE. The less effort someone expects to have to put to a certain task using a certain technology, the higher the value of EE. In terms of cost and benefit: PE describes how much benefit a person hopes to gain; EE describes how much the same person hopes to reduce costs. FC and SI describe contextual concepts concerning the institutional and social surroundings of a person. While facilitating conditions are the conditions set by the institutional surrounding (e.g. employer of a person, university someone is enrolled in, etc.), social influence is derived from people within the direct social environment of a person and their attitudes towards the technology in question.

According to the UTAUT model, these four factors directly influence a person's intention to use a certain technological solution to achieve work related goals. Drawing from the Theory of Planned Behavior (TPB) [2], the UTAUT model then proposes a following influence from use intention to actual use behavior.

In summary, the Unified Theory proposed by Venkatesh et al. aims to include individual factors (PE and EE, i.e. cost and benefit) as well as social (SI) and institutional ones (FC). The inclusion of the TPB model allows for a final differentiation between use intention and actual use behavior, since not all intention necessarily lead to execution of said intent.

The UTAUT model has been applied in various studies since its first formulation, and empirical evidence shows strong support in workplace environments when analyzing attitudes towards work tools. However, when applied in educational context, specifically in higher education, studies have difficulties reproducing the theorized influences proposed in UTAUT [3, 4].

One major difference between workplace and educational settings is the motivational aspect of people's behavior. Typically, behavior in the workplace is driven by extrinsic motivators, e.g. salary, hierarchical position within an organization, social status, etc. Whereas in educational settings intrinsic motivators play a much more important role determining a person's positive learning behavior than in typical workplace settings [5, 6].

Self-Determination Theory (SDT) [7] proposes a view on a person's motivational attitude as a spectrum determined by the level of autonomy they feel they have over their decision making process. The more freedom someone feels, i.e. the more they feel self-determined in making a decision, the more likely they will be intrinsically motivated in their behavior. The motivational spectrum ranges from amotivation, where decisions are made without any amount of self-determination, through four stages of (semi-)extrinsic motivations – defined by the level and type of so-called "regulation", i.e. external control factors – up to a state of absolute autonomy in making a decision, described as intrinsic motivation.

UTAUT already includes motivational aspects, albeit primarily extrinsic motivators (i.e. cost and benefit from using a certain technology). Following the premise of SDT, however, a comprehensive model should also include the rest of the spectrum (i.e. intrinsic motivation and amotivation). Another interesting point could be made in analyzing the relationship between technology acceptance and motivation in general: Is motivation a part of technology acceptance? Is it one of the factors contributing to use intention or use behavior? Are motivation and technology acceptance actually two separate concepts which should be distinguished from one another? These questions suggest revisiting the original UTAUT model to analyze the relationship between motivational and acceptance factors influencing use behavior, especially – but not only so – in educational settings.

## 3 Comparative Study of Different Learning Environments

Data from four previously conducted survey studies were combined to create one combined data set for this analysis. Two studies (A & C) were conducted amongst users of a faculty-wide online learning management system in 2013 and in 2014. The other two studies (B & D) were conducted amongst users of an OLV system in 2013 and 2015. Studies A, B, and C administered online questionnaires within the respective learning environment, in study D pen-and-paper questionnaires were distributed during a lecture which was being recorded and made available online through the OLV system in question. All four studies were conducted at the same major German university, focused on aspects of technology acceptance and user motivation, and in all questionnaires participants were confronted with the same question items concerning their attitudes towards the respective online learning environment (variation only in the name of the respective online system); therefore, a joint analysis in a combined data set was possible. This combination of four different measurements in two different online learning environments with different educational settings was chosen to compare the differences in motivation and corresponding differences in technology acceptance. Study A yielded 251 valid cases, study B 210 cases, study C 100 cases, and study D 112 cases – the complete data set therefore consists of 673 responses from all four studies. 79.3 % participants were female, 16.9 % were male (3.6 % with missing values), the average age was 24 ($N = 638$, $M = 24.28$, $SD = 6.43$). While the gender distribution may seem unnaturally skewed towards female participants, however, registration numbers at this particular university show a general majority (about 60 %) of female students, the percentage being even higher in courses for pedagogy, psychology, and teacher education, where most users of the learning systems in this analysis are located. Thus, we do not expect much of an impact on the following results.

The variables from UTAUT and SDT were measured using four questionnaire items each. The questions for UTAUT constructs were adapted from the original study by Venkatesh et al. [1], while the questions for motivational concepts were adapted from a study by Standage et al. [8], which in turn are based on the Academic

Motivation Scale [9]. Due to constraints of the combined data set, the following analysis will focus on three of the theorized aspects of motivation: intrinsic motivation (IM), identified regulation (IR), and amotivation (AM). The exclusion of further aspects of extrinsic motivation was necessary, since not all four studies measured all these sub-concepts, or used different questionnaire items to determine the motivation of participants. Therefore, only variables present in all four studies and measured with the same questionnaire items were included in this analysis.

The following analysis consists of two main steps: First, a confirmatory factor analysis using a principal component analysis (PCA) was performed to assess the validity of the scales proposed by UTAUT and SDT. Second, a variance analysis (one-way ANOVA) was performed to compare the mean values between the four studies. All statistical calculations were made with IBM SPSS 23 for Windows.

The PCA confirmed six factors with a total of 72.68 % explained variance. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy is well within the accepted range (KMO = .918), together with the results from Bartletts's Test of Sphericity (Approx. Chi-Square = 11125,354, df = 378, $p < .001$) we can interpret the results as a valid factor analysis. Results from the rotated component matrix (Varimax rotation) indicate a few items have to be omitted due to weak or cross-loading. Most notably, items for facilitating conditions do not form a single coherent factor, but show factor loadings towards one or more of the other identified constructs. Social influence, on the other hand, appears to form two distinct factors with strong loadings from their respective items. Looking at the corresponding questions in the questionnaires, the two factors can be interpreted as Social influence coming from the institutional surrounding (i.e. university and professors), and social influence coming from personal surroundings (i.e. friends, fellow students, etc.). The eight items indicating measures for intrinsic motivation and identified regulation show strong loading values towards one factor, which will be considered as "Motivation" in the following analysis. To summarize, the six factors identified and confirmed by way of PCA are: Motivation (F1MO), Effort Expectancy (F2EE), Performance Expectancy (F3PE), Amotivation (F4AM), Social Influence by the Institution (F5IS), and Social Influence by Peers (F6PS). A reliability analysis shows high values of Cronbach's Alpha throughout the identified scales: Alpha values range from .919 through .944 for the first five constructs, while F6PS yields only .780, though it is still within the acceptable range of $> .7$. The scales can assume values ranging from 1 through 7, where high values represent a strong foundation of the concept in a person's attitudes. Over all four studies, participants average high expectations towards performance gain and effort minimization. They feel moderately motivated (M = 3.41, SD = 1.53), moderately supported by their institution in using the respective technology (M = 3.90, SD = 1.99), but a little stronger by their peers (M = 4.31, SD = 1.58). The ANOVA results show statistically significant differences between groups for all UTAUT and SDT variables. Since the data did not meet the requirement of variance homogeneity within the studies (as determined by Levene test for variance homogeneity), a Welch F-Test was computed, as presented in Table 1.

**Table 1.** ANOVA results (Welch F-Test)

| Robust tests of equality of means | | | | |
|---|---|---|---|---|
| | Statistic[a] | df1 | df2 | Sig. |
| F1MO Motivation | 18,060 | 3 | 280,432 | ,000 |
| F2EE Effort Expectancy | 404,819 | 3 | 269,498 | ,000 |
| F3PE Performance Expectancy | 202,390 | 3 | 274,425 | ,000 |
| F4AM Amotivation | 544,554 | 3 | 243,642 | ,000 |
| F5IS Institutional Support | 40,953 | 3 | 264,728 | ,000 |
| F6PS Peer Support | 113,892 | 3 | 294,574 | ,000 |

[a]Asymptotically F distributed.

A post-hoc Scheffé test was performed to assess differences between the four studies concerning the different variables. Participants from study A exhibited the lowest average value for motivation of all four studies (M = 2.91, SD = 1.57) and the lowest institutional support (M = 3.04, SD = 1.71). Peer support and Effort Expectancy were relatively high, though not the highest of all four groups. Participants from study B showed very high values for EE and PE (M = 6.30, SD = .74; M = 6.16, SD = 1.03), while also showing high values of institutional support (M = 4.90, SD = 1.82) and very low values of amotivation (M = 1.17, SD = .47). Study C showed the lowest values for EE and PE (M = 2.47, SD = .97; M = 2.82, SD = 1.32), while yielding the highest values for amotivation (M = 6.00, SD = 1.16).

## 4 Discussion and Conclusion

The purpose of this paper was to illustrate the need for a more inclusive approach to technology acceptance research in educational contexts. We propose revisiting the UTAUT and including an autonomy-based view on motivation following the concept of Self-determination Theory, and the inclusion of intrinsic motivation as well as amotivation into traditional acceptance models, to better understand the attitudes of people using learning technologies such as online learning management systems.

Results from an ANOVA show statistically significant differences between different learning contexts. Institutional support appears to coincide with decreased amotivation amongst participants as well as their expectations of reduced effort and increased learning performance. On the other hand, moderate to low social support appears to be linked to high amotivation as well as low gain expectations from using the system. These results – while not yet an in-depth analysis of the statistical connections – are indicative of a possible link between autonomy-based constructs of motivation and the acceptance of technological solutions to assist learning. Future research should therefore include such measures and further investigate the connections between users' motivation and their acceptance of technology, as well as the combined influence on their use intentions and their use behavior. With such detailed insight, (online) learning environment as well as the corresponding learning scripts coming from educators could be adapted to increase the success of technology enhanced learning.

## References

1. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. MIS Q. **27**, 425–478 (2003)
2. Ajzen, I.: The theory of planned behavior. In: van Lange, P.A.M., Kruglanski, A.W., Higgins, E.T. (eds.) Handbook of Theories of Social Psychology, pp. 438–459. SAGE Publications Ltd., London (2012)
3. Pedrotti, M., Nistor, N.: Online lecture videos in higher education: acceptance and motivation effects on students' system use. In: Sampson, D.G. (ed.) IEEE 14th International Conference on Advanced Learning Technologies (ICALT), Athens, Greece, pp. 477–479. IEEE, Piscataway, NJ, 7–10 July 2014
4. Nistor, N.: When technology acceptance models won't work. non-significant intention-behavior effects. Comput. Hum. Behav. **34**, 299–300 (2014)
5. Beachboard, M.R., Beachboard, J.C., Li, W., Adkison, S.R.: Cohorts and relatedness. Self-determination theory as an explanation of how learning communities affect educational outcomes. Res. High. Educ. **52**, 853–874 (2011)
6. Bailey, T.H., Phillips, L.J.: The influence of motivation and adaptation on students' subjective well-being, meaning in life and academic performance. High. Educ. Res. Dev. **35**, 201–216 (2015)
7. Deci, E.L., Ryan, R.M.: Self-determination theory. In: van Lange, P.A.M., Kruglanski, A.W., Higgins, E.T. (eds.) Handbook of Theories of Social Psychology, pp. 416–437. SAGE Publications Ltd., London (2012)
8. Standage, M., Duda, J.L., Ntoumanis, N.: A test of self-determination theory in school physical education. Br. J. Educ. Psychol. **75**, 411–433 (2005)
9. Vallerand, R.J., Pelletier, L.G., Blais, M.R., Briere, N.M., Senecal, C., Vallieres, E.F.: The academic motivation scale. a measure of intrinsic, extrinsic, and amotivation in education. Educ. Psychol. Measur. **52**, 1003–1017 (1992)

# Reflective Learning at the Workplace - The MIRROR Design Toolbox

Sobah Abbas Petersen[1(✉)], Ilaria Canova-Calori[1,3], Birgit R. Krogstie[2], and Monica Divitini[1]

[1] Department of Information and Computer Science, NTNU, Trondheim, Norway
{sap,divitini}@idi.ntnu.no

[2] Department of Informatics and E-Learning, NTNU, Trondheim, Norway
birgit.r.krogstie@ntnu.no

[3] House of Knowledge AS, Trondheim, Norway
ilaria.canova@hakvag.no

**Abstract.** Using the theoretical understanding of a domain to inform design is a well-known challenge. In this paper we address this challenge in relation to reflective learning at the workplace. The work is informed by theoretical and empirical work that led to the Computer Supported Reflective Model (CSRL model), a model describing the different phases of reflective learning and how it can be supported by different tools. Starting from this model, we have conceived a set of conceptual tools that can help designers without prior knowledge of reflective learning to design for it, taking into account the complex interleaving of work practices and reflection.

**Keywords:** Reflective learning · Design · Design Toolbox · Workplace learning

## 1 Introduction

Using the theoretical understanding of a domain to inform design is a well-known challenge. This has recently been addressed successfully by transforming the knowledge captured in an abstract framework into a set of cards to be used in design, (e.g. [1] and [2]). Cards are also widely used in creativity workshops [3] and to support creative designs such as gamification in social applications [4]. In this paper, we propose the use of cards for the design of technology to support workplace reflection.

Organizations often lack a culture of supporting learning, sharing and reflecting, as well as tools to support these processes [5]. The MIRROR Design Toolbox, presented in this paper, was developed to support the systematic design of reflection apps, based on actual learning practices and learning needs in the organization. The toolbox provides cards to support designers in early phases of design with the added goal of developing insight on how reflection unfolds, or might unfold, in the organization. The toolbox is grounded in theoretical and empirical work done in the MIRROR project [5]. The MIRROR Design Toolbox is available under the Creative Commons License.

The paper is structured as follows: Sect. 2 briefly introduces the theoretical framework that guided the design of the toolbox; Sect. 3 presents the rationale and overview


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 478–483, 2016.
DOI: 10.1007/978-3-319-45153-4_46

of the toolbox; Sect. 4 describes how the toolbox can be used in different situations in an organization and Sect. 5 provides a summary.

## 2 Reflective Learning

In work life, reflection is ubiquitous to everyday sense making and problem solving, triggered by discrepancies between existing knowledge and new experience. Reflection may happen spontaneously and "in action" or with more distance to the experience reflected upon [6]. Furthermore, reflection may be undertaken individually or in a group, and these are often intertwined [7]. Dewey linked reflection and thinking, focusing on the reflective attitude and skills of the learner [8]. Boud et al. describe reflective learning as *"a return to experience in which the experience – behaviour, ideas and/or feelings - is re-evaluated and an outcome is produced*" [9]. Reflection is thus part of a *learning cycle* [10, 11]. The reflective learning process can be scaffolded, e.g. through *helping learners ask the right questions*. For instance Driscoll [12] suggests that the questions "What?", "So What" and "Now what?" guide reflection through the learning cycle.

The model of Computer Supported Reflective learning (CSRL) was developed in the MIRROR project [13] with the aim of framing new insight on reflection at work and supporting the design of technology for this reflection. The core of the model is a reflection cycle of four main stages (see Fig. 1). The rectangles show the stages, the arrows with broken lines show triggers of reflection, and the arrows in complete lines show the inputs from one stage of the cycle to the other.

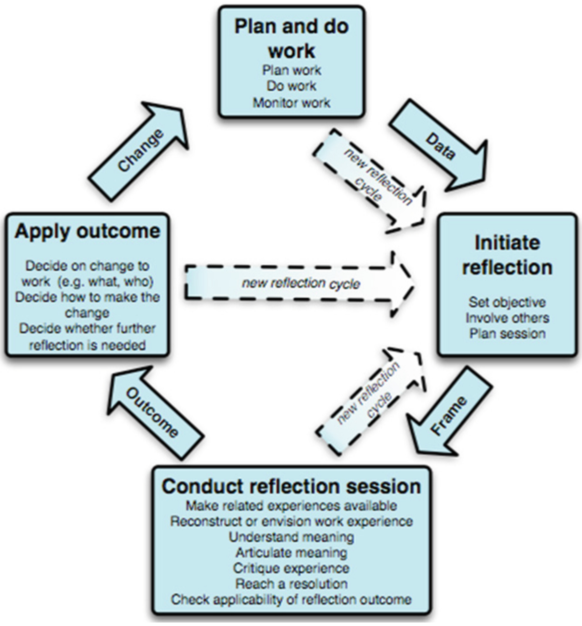



**Fig. 1.** CSRL model - Reflection cycle diagram

The Plan and do work stage corresponds to engaging in work activity in which experiences are made. Data about experiences is created in this process (possibly only represented in human memory). The Initiate reflection stage may involve the setting of objectives for reflection, including other people (for collaborative reflection), and planning a reflection session. A more or less explicit frame for the reflection results from this. The Conduct reflection session stage comprises activities in which the return to, and re-evaluation of, experience happens, thereby creating an outcome. Possible activities in a reflection session include the reconstruction and sharing of experiences and work to reach a solution and consider its applicability. The Apply outcome stage comprises of deciding what will be the changes to work and how to bring them about. Application of an outcome to work, back in the first stage of the model, can amount to a visible, measurable change, but may also be more subtle, e.g. in the form of a changed readiness for certain action. Tool support can be provided for each of the four stages and the transitions between them. MIRROR apps provide some of these tools.

## 3 The MIRROR Design Toolbox

The MIRROR Design Toolbox supports the design of apps that build on the understanding of reflective learning captured in the CSRL model. The toolbox has been designed to provide this understanding to designers of ICT applications with limited or no prior knowledge of the model, or more in general of reflective learning. The toolbox comprises a set of conceptual tools in the form of cards and templates. These tools are designed in a way that the designer can work with the users in a collaborative manner while gathering the requirements and obtaining an understanding of the work context. Ideas from co-design (e.g. [14]) and participatory design are used to build trust among the stakeholders and to ensure that the requirements stem from the users. Examples of some tools are shown in Fig. 2.


**Fig. 2.** Different sets of cards in the MIRROR design toolbox

The toolbox supports a designer to consider all the stages in the MIRROR CSRL model (see Fig. 1) and promotes considering reflection at the individual, team and organizational levels. Wherever possible, tools are provided to support a creative design process (e.g. creativity cards) where the designer is prompted to think in an alternative way, either through keywords or examples from the MIRROR apps. The tools in the toolbox are categorized into three parts, each supporting a different part of the design process. A description of the categories and some example tools for each category are provided in Table 1.

**Table 1.** Overview of the tools in the MIRROR design toolbox

| Toolbox category | Category description | Specific tools |
|---|---|---|
| Landscape | To capture a good understanding and describe the work and organisational context where reflection takes place. | Stakeholder map |
| | | Personas in practice |
| | | Organization description and core practices |
| | | Technology infrastructure |
| | | Privacy concerns |
| Storyline | To obtain a deeper insight into the practices and use of reflection tools at the workplace, based on the CSRL model (Fig. 1). | A set of Storyline word cards |
| | | A set of storyline question cards to capture reflection in the specific workplace |
| Specification | To specify tools for reflection, which may be new tools or enhancements and adaptations of existing ones. | A set of questions to obtain overview of existing tools |
| | | A set of cards to prompt functionality specification |
| | | A set of high-level guidelines |
| | | A list of Barriers to reflection |

## 4 The MIRROR Design Toolbox in Practice

The need for the MIRROR Design Toolbox arise due to several situations and causes. A designer may choose the tools depending on the maturity of the reflection process at the workplace, the tools that are available for supporting reflection and her knowledge about the organisation's context and needs. We present the Toolbox through three examples, to illustrate different situations in an organization.

**Example 1**: a change in the existing tools and practices in a workplace. This may be the case where the current tools and practices could be improved or enhanced to support reflection and learning. The organization may already have a culture of supporting reflection and learning and the designer may already understand the context of work. In such a situation, the designer may have been involved in the design of tools that already exist and have an overview of existing tools to support reflection. The designer may start by updating the overview of tools and landscape or enhancing the existing design of the tool(s). In this case, the main set of tools from the toolbox that the designer may use are the ones for Specification; see example shown in Fig. 3. Here, we can assume that the designer is knowledgeable about the content of the workplace and may only use the tools from the Landscape and Storyline that complement existing knowledge.

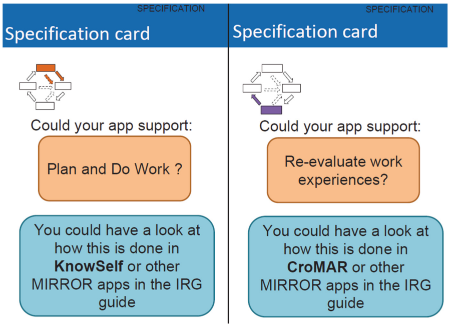



**Fig. 3.** Example of tools from specification

**Example 2**: Reflective practices and supporting tools are to be introduced to a new team within the organization. The designer may not be familiar with the needs to support reflection and learning within the new team. The designer may have to start with the Landscape tools to understand the differences in the context of work between the new team and teams that currently use tools for learning and reflection, and then update the overview of tools and landscape or enhance the existing designs of tools to support reflection. The tools from Storyline may be useful here to understand the reflection process, particularly if the two work contexts are very different. The tools from Specification could be used to consider different functionalities.

**Example 3**: An organization wants to introduce reflective learning at work. The chances that the designer may be an external person is high or even if the designer is from within the organization, there may not be many tools that support learning and reflection. In this case, the designer may have to start with the set of tools for capturing the Landscape to understand the work context and the need to support reflection and learning. These tools enable a designer to obtain an overview of the current practices, which are relevant as they may give insights into how the new practices and tools could leverage on the current practices. It is equally important to include the users in the design to promote a culture for learning and reflection and to gain their trust in the ICT tools that are designed. The designer may then use the Storyline tools to obtain a complete overview of the workplace before the specification of functionalities for tools to support reflection and learning.

## 5 Summary

The MIRROR Design Toolbox provides a set of conceptual tools that may be used to design new reflection apps or enhance existing tools and practices for better support for reflection in the workplace. This paper provides an overview of the toolbox and how it had been developed to meet the theoretical foundations of reflection and learning in organizations, described by the MIRROR CSRL model. The aim of the toolbox was to provide a comprehensive set of tools supporting the designers while giving them flexibility to choose when to use which tool. The next stage in our work is to evaluate it by using the tools in different organizational contexts as illustrated in the three cases.

**Acknowledgements.** This work was conducted as a part of the EU TEL MIRROR project and partly funded by the TELL GEMINI Center (http://www.tell-gemini.org/). We thank the participants of the project for feedback and fruitful discussions.

## References

1. Hornecker, E.: Creative idea exploration within the structure of a guiding framework: the card brainstorming game. In: Fourth International Conference on Tangible, Embedded, and Embodied Interaction, TEI2010. ACM Press (2010)
2. Mueller, F., et al.: Supporting the creative game design process with exertion cards. In: CHI 2014. ACM Publications, Toronto (2014)
3. Michalko, M.: Thinkertoys: A Handbook of Creative-Thinking Techniques, 2nd edn. Ten Speed Press, Berkeley (2006)
4. Oliveira, M., Petersen, S.A.: Co-design of neighbourhood services using gamification cards. In: HCI International, Crete, Greece (2014)
5. MIRROR. Project Homepage. http://www.mirror-project.eu/. Accessed 4 Apr 2016
6. Schön, D.: The Reflective Practitioner. Basic Books, Inc., New York (1983)
7. Prilla, M., Pammer, V., Krogstie, B.R.: Fostering collaborative redesign of work practice: challenges for tools supporting reflection at work. In: Bertelsen, O.W., et al. (eds.) Proceedings of the 13th European Conference on Computer Supported Cooperative Work, ECSCW 2013. Springer, London (2013)
8. Dewey, J.: How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process. D.C. Heath, Boston (1933). Revised edition
9. Boud, D., Keogh, R., Walker, D.: Reflection: Turning Experience into Learning. RoutledgeFalmer, London (1985)
10. Kolb, D.A.: Experiential Learning: Experience as a Source of Learning and Development. Prentice Hall, New Jersey (1984)
11. Kolb, D.A., Fry, R.: Towards an Applied Theory of Experiential Learning. In: Cooper, C.L. (ed.) Theories of Group Processes, pp. 33–58. Wiley, London (1975)
12. Driscoll, J., Teh, B.: The potential of reflective practice to develop individual orthopaedic nurse practitioners and their practice. J. Orthop. Nurs. **5**, 95–103 (2001)
13. Krogstie, B.R., Prilla, M., Pammer, V.: Understanding and supporting reflective learning processes in the workplace: the CSRL model. In: Hernández-Leo, D., Ley, T., Klamma, R., Harrer, A. (eds.) EC-TEL 2013. LNCS, vol. 8095, pp. 151–164. Springer, Heidelberg (2013)
14. Morelli, N., Würtz, P.: My Neighbourhood Deliverable 2.2 - Handbook of co-design activities for co-designing services. Department of AD:MT, Aalborg University, Aalborg (2013)

# Toward a Play Management System for Play-Based Learning

Eric Sanchez[1,2(✉)], Claudine Piau-Toffolon[3], Lahcen Oubahssi[3], Audrey Serna[4], Iza Marfisi-Schottman[3], Guillaume Loup[3], and Sébastien George[3]

[1] Ecole Normale Supérieure de Lyon, Lyon, France
eric.sanchez@unifr.ch
[2] University of Fribourg, Fribourg, Switzerland
[3] Université Bretagne Loire, Université du Maine, LIUM, 72085 Le Mans, France
{claudine.piau-toffolon, lahcen.oubahssi,iza.marfisi,guillaume.loup, sebastien.george}@univ-lemans.fr
[4] INSA de Lyon, LIRIS, UMR5205, 69622 Lyon, France
audrey.serna@insa-lyon.fr

**Abstract.** This position paper is dedicated to describing a preliminary model of an integrated system, called *Play Management System* (PMS). PMS is designed to support both players and teachers to deliver, use, manage and track play situations. This PMS model results from a design-based research methodology. Our approach focuses on (1) the learners and the situation that emerges when they play the game, rather than the system dedicated to play and (2) the teachers who want to manage a game-based learning situation. Thus, we argue for a shift from a game-based to a play-based perspective.

**Keywords:** Game-based learning · Play management system · Classroom orchestration · Design-based research · Teachers' requirement analysis

## 1 Introduction

Within a context marked by the development of alternative pedagogies, this position paper aims to describe a model of an integrated system, called *Play Management System* (PMS), dedicated to support players and teachers to deliver, use, manage and track play situations. The purpose of this article is to propose an innovative approach for implementing a play-based learning approach by (1) focusing on the learners and taking into consideration the situation that emerges when they play rather than the artifact dedicated to play (play *vs* game) and (2) focusing on the teachers who want to implement and manage a play-based learning situation in their classroom (play management *vs* game design). Thus, we address the issue of teachers' requirements for the orchestration of a play situation within an educational context. In the first section of this paper, we advocate for a player-centered approach for game-based learning. The second section presents a game developed during the project and the design-based research methodology adopted for designing this game. The third


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 484–489, 2016.
DOI: 10.1007/978-3-319-45153-4_47

section describes the model of a *Play Management System,* based on the results that emerged implementing and testing a game in real school contexts.

## 2 Switching from Game to Play

Digital Epistemic Games (referend to as JENs[1] in this paper) are playful and authentic *learning situations* that lead the learners to solve complex, interdisciplinary and non-determinist problems [1]. JENs allow students to develop their own ways of thinking and acting by designing and trying out their own solutions [2, 3]. JENs also rely on mixed reality technologies [4] to create contextual and situated activities and to support knowledge co-construction among learners. "*Learning situation"* are key words of this definition and Henriot emphasized the importance of distinguishing the game, as an artifact, and play, its usage [5]. For Henriot, play emerges from the interactions between a player and a game. In other words, play depends on the *lusory attitude* [6] of the players, i.e. their willingness to take on the rules of the game and to participate. Usually, studies on game-based learning are focused on the characteristics of a given game. We consider that it is the *learning situation* in which the game is used that is paramount, and that a shift from a game-based to a play-based perspective is needed.

Implementing a game-based pedagogy implies that the teachers have to manage the classroom orchestration [7]. Their role entails the introduction of the game to the students. A teacher may also act as a game master and be involved in the assignment of rewards, if it is not automatic. Following a gamification trend, for some *Learning Management Systems* such as *Moodle*[2], badge functionalities enable students to represent their achievements and skills. Such an approach has been implemented for the game *Classcraft*[3], designed for classroom management. The success of the game, in terms of its adoption by teachers, demonstrates its relevance [8]. Another teachers' roles relates to the animation of debriefing sessions, dedicated to foster reflection and metacognition after the game, or in-between game sessions [9]. Regarding the importance and the complexity of the teachers' role for game-based learning, it becomes apparent that they need to be taken into account when designing a technical solution to implement their game and that such a solution should include support for dynamic classroom orchestration.

## 3 Play Management System Requirement Analysis

The methodology of this research work is based on the collaboration of practitioners (teachers) and researchers. Thereafter, the objectives are both pragmatic (producing innovative digital applications adapted to the teachers' expectations) and theoretical (developing new models for instruction and learning). As a result, the methodology applied is influenced by the *Design-Based Research* (DBR) approach [10]. This design process is combined with the analysis of these educational practices, carried out

---

1 JEN stands for *Jeu Epistémique Numérique* in French.
2 https://moodle.com/, visited on March 2016.
3 http://www.classcraft.com/fr/, visited on March 2016.

collaboratively by researchers and practitioners. In the field of *Technology Enhanced Learning*, DBR has close relationships with software design methodologies that aim to integrate end-users in the early stages of the design process, such as *Agile* or *user-centered methodologies* [12] and *participatory design* [11] from the Human Computer Interaction field. Considering our needs for play orchestration of JENs, we therefore choose to apply DBR for the analysis of the teachers' requirement. In order to provide an example of the JENs designed with this method, let us present *Insectophagia*, one of the three games we designed with teachers.

*Insectophagia* is a game that convers the principles of sustainable development. This game was designed for five classes (86 students) from 15 to 17 years old. The global objective, for each team (composed of 3 or 4 learners) is to create a start-up company, specialized in insect-based food production. First, the team needs to choose the type of insect they want to farm, based on ecological and dietary properties. Then, they have to find a proper location to build their factory and make the right investments in terms of sustainable energy sources. Finally, they need to come up with an innovative and appealing product for customers. The players use digital technologies and real-world settings depending on the performed mission. The game lasts approximatively 7 weeks (approx. 18 h) depending on the school. Rewards and points depend on how the players manage to deal with environmental, social and economic issues. The teacher is responsible for introducing the different missions, rewarding the students, time keeping and also for chairing the debriefing session. We also designed two other very different JENs: *Rearth,* a science-fiction game dedicated to science and programming and *Generalisima*, an exploration game, used by a company to train their employees.

Six design sessions, involving researchers and teachers, were dedicated to define the learning objectives, the game universe and the gameplay of the JENs. Several students also participated and provided ideas. After the design phase, we developed prototypes that were partial paper-based and partial digital. The main structure of the game was paper-based (e.g. paper cards, tokens representing points, game booklet) while punctual activities were completed on computers (e.g. documentary research, smartphone game for exploring the potential locations for the farm). The prototype was experimented in naturalistic conditions by the teachers involved in the design phase. After the experimentation, two sessions were organized to discuss the lessons learned from the experimentation of the JENs. From the teachers' point of view, the game was valuable for students since they were immerged in a complex situation for which they had to collaborate and use various resources. However, the pedagogical situation emerging from this game was not easy to manage in real-time in the classroom. The teachers expressed the need for a tool dedicated to manage class orchestration (assigning specific goals to players, organizing teams) and play management (rewarding successful players with points and badges). These fruitful discussions helped us to define a preliminary generic model of PMS for JENs.

## 4 Towards a Play Management System

The first user studies described previously support the need for a generic system, dedicated to play management. We describe the global architecture of the system and the functionalities that we identified.

### 4.1 A System Dedicated to Play Management

We called this platform *Play Management System* (PMS) in reference to *Learning Management Systems* (LMS). Indeed, there are several analogies between PMS and LMS: the shift from game to play is an analogy of the shift from teaching to learning, in line with a learner-centered approach for education; the shift from a resource-centered to an activity-centered approach, in line with the *Educational Modelling Language* community approach [13]; and finally, management system refers to the complexity of the game master's roles: designing and implementing the learning/gaming situation, real-time tutoring and assessing/rewarding, and debriefing.

Figure 1 summarizes the main ideas behind the PMS as discussed during debriefing session. A PMS is an integrated system that supports players and teachers to carry out play-based learning. This system, separated from the game itself (JEN units), supports the different dimensions of an educational play situation (learning context, game documents, game characteristics, social interactions, and technological aspects). As a result, the PMS may be used to plan, implement, and assess specific learning processes based on play activities.

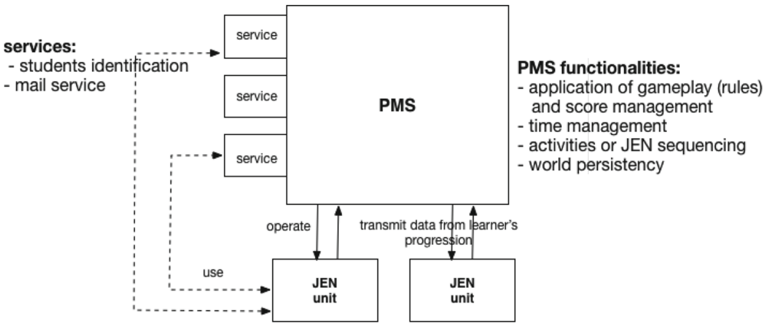



**Fig. 1.** A global scheme of the PMS emerging from debriefing sessions

Within a JEN, both individual play and collaborative play must be encouraged. Therefore, a PMS must offer the means for developing individual and collaborative activities that foster *production*, *communication* and *coordination* [14].

There is also a need, for the teacher to track play activities in order to get information about players' achievements. Indeed, the teacher's expressed the need for PMS to take into consideration data collection (traces) for the asynchronous analysis of players' interactions. As a result, learning analytics [15] services might be offered.

Since traditional *Learning Management Systems* (LMS) have been developed to manage learning situations such as online or hybrid courses, they do not offer all the functionalities and dedicated interactions needed to manage play situation and they are not well suited to the specific needs of JENs. We therefore propose a new architecture.

### 4.2 Toward a Play Management System Architecture

The ideas expressed by the teachers and researchers during the debriefing sessions, show that the PMS architecture should be composed of 4 modules with a set of functionalities. (1) The *Administrator Module* offers the functionalities to customize the PMS, manage user accounts, the resources needed to play (documents for students) and the technical aspects of the game. (2) The *Player/Learner Module* offers all the functionalities for each individual player to perform the missions of the game and also to customize her/his personal data and avatar. (3) The *Game Master Module* is dedicated to *dynamic orchestration*. Teachers have the possibility to define (or redefine) specific goals, to organize teams of learners and to reward the players with points and badges. The PMS also offers different tools to support interactions between players and between players and the teacher (as a game master). (4) *The Service Module* provides teachers and learners with generic functionalities that can be used at any time during the JEN. These functionalities offer numerous ways of enhancing the play experience with specific reward tools (leader board, points, badges…), supporting collaborative work (collaborative production tools, task management tools…) and communication (chat, forum, mail…), and also tracking actions performed by the players (learning analytics).

## 5 Conclusion

The goal of this study was to propose an innovative approach for implementing play-based learning into secondary education. Since we decided to implement a player-centered approach for game-based learning and to offer support for dynamic classroom orchestration, a new perspective emerged. A PMS is an integrated system that supports players and teachers to deliver, use, manage, and track play situations dedicated to educational objectives. This system may be used to plan, implement, and assess *Digital Epistemic Games*. The PMS model proposed in this paper is composed of four modules: *Administration Module, Player/Learner Module, Game-Master Module* and *Service Module.*

The contributive, collaborative and iterative methodology based on experimentation in real school settings enabled linking pragmatic issues (implementing a game in a classroom) and theoretical issues (designing a model of a system adapted to play management). The main results consist of the identification of the functionalities needed for play orchestration (*i.e.* teachers' requirements). New experimentations are now conducted. The preliminary results underline that the persistence of the game has been recognized to be important. Thus, the prototype now offers the players the opportunity to consult their logbook, refreshed in real time. It is expected that PMS, by taking on the issue of persistence, will sustain the players' motivation and enhance decision making. In sum, the lessons learned from the ongoing experimentations, lead us to better take into consideration players' requirements.

**Acknowledgments.** The JEN.lab project is a multi-disciplinary project funded by the French Research Agency (ANR-13-APPR-0001). The authors want to thank the students and teachers for their full support and key contributions.

## References

1. Sanchez, E.: Learning, Serious Games, and Gamification. Inmedia (2014)
2. Shaffer, D.: Epistemic frames for epistemic games. Comput. Educ. **3**(46), 223–234 (2006)
3. Ohlsson, S.: Learning to do and learning to understand: a lesson and a challenge for cognitive modeling. In: Reiman, P., Spade, H. (eds.) Learning in Humans and Machines: Towards an Interdisciplinary Learning Science, pp. 37–62. Elsevier Science, Oxford (1995)
4. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE (Institute of Electronics, Information and Communication Engineers) Trans. Inf. Syst. **77**(12), 1321–1329 (1994). Special Issue on Networked Reality
5. Henriot, J.: Le jeu. Presses Universitaires de France, Paris (1969)
6. Suits, B.: Grasshopper: Games Life and Utopia. David R. Godine, Boston (1990)
7. Dillenbourg, P.: Design for classroom orchestration. J. Comput. Educ. **69**, 485–492 (2013)
8. Sanchez, E., Young, S., Jouneau-Sion, C.: Classcraft: from gamification to ludicization of classroom management. Educ. Inf. Technol. **5**(20), 1–17 (2016)
9. Garris, R., Ahlers, R., Driskell, J.E.: Games, motivation, and learning: a research and practice model. Simul. Gaming **33**(4), 441–467 (2002)
10. Design-Based Research Collective: Design-based research: an emerging paradigm for educational inquiry. Educ. Researcher **1**(32), 5–8 (2003)
11. Schuler, D., Namioka, A. (eds.): Participatory Design: Principles and Practices. Lawrence Erlbaum Associates, Hillsdale (1993)
12. Highsmith, J.: Agile Software Development Ecosystems. Addison-Wesley Professional, Boston (2002)
13. Giesbers, B., van Bruggen, J., Joosten-ten Brinke, D., Burgers, J., Koper, R., Latour, I.: Towards a methodology for educational modelling: a case in educational assessment. Educ. Technol. Soc. **10**(1), 237–247 (2007)
14. Ellis, C.A., Gibbs, S.J., Rein, G.L.: Groupware: some issues and experiences. Commun. ACM **34**(1), 39–58 (1991)
15. Long, P., Siemens, G.: Penetrating the fog: analytics in learning and education. EDUCAUSE Rev. Online **5**(46), 31–40 (2011)

# The Blockchain and Kudos: A Distributed System for Educational Record, Reputation and Reward

Mike Sharples[1(✉)] and John Domingue[2]

[1] Institute of Educational Technology, The Open University, Milton Keynes, UK
mike.sharples@open.ac.uk
[2] Knowledge Media Institute, The Open University, Milton Keynes, UK
john.domingue@open.ac.uk

**Abstract.** The 'blockchain' is the core mechanism for the Bitcoin digital payment system. It embraces a set of inter-related technologies: the blockchain itself as a distributed record of digital events, the distributed consensus method to agree whether a new block is legitimate, automated smart contracts, and the data structure associated with each block. We propose a permanent distributed record of intellectual effort and associated reputational reward, based on the blockchain that instantiates and democratises educational reputation beyond the academic community. We are undertaking initial trials of a private blockchain or storing educational records, drawing also on our previous research into reputation management for educational systems.

**Keywords:** Blockchain · Reputation management · Self-determined learning · e-portfolios · Records of achievement

## 1 Introduction

The blockchain is being proposed as a disruptive technology that could transform the finance and commerce sectors (see e.g. [1, 2]). In this paper we explore the disruptive potential of the blockchain for education and its value in support of self-determined learning. To understand the relevance of the blockchain to education, it is important to understand its components, as any one or more may be adapted for educational use.

First, there is the blockchain itself, a distributed record of digital events. The blockchain is a long chain of linked data items stored on every participating computer, where the next item can only be added by consensus of a majority of those participating. There are public blockchains that anyone can access and potentially add to, and there are private blockchains used within an organization or consortium. The best known, but not the only, blockchain is the one at the heart of the Bitcoin system of digital money [3].

Second, there is the 'distributed consensus' method to agree whether a new block is legitimate and should be added to the chain. This is done by requiring a participant's computer to perform a significant amount of computational work ('proof of work' or 'mining') before it can try to add a new item to the shared blockchain. To create a false blockchain and get that accepted by consensus would be prohibitively difficult. An unfortunate consequence of the 'proof of work' requirement, is that the computer


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 490–496, 2016.
DOI: 10.1007/978-3-319-45153-4_48

performing the mining operation to produce a new block must spend a considerable amount of computational power and electricity, just to provide the proof of work. Alternatives are being developed for distributed validation of new blocks, including 'proof of stake' where, to add a new block, a participant must show a certain amount of currency or reputation, which is lost if that block is not accepted by consensus [4].

Third, each block in the blockchain can hold a small amount of data (typically up to 1 Mb) which could be any information that is required to be kept secure, yet distributed. These could be records of currency transactions (as in Bitcoin) or, for education, exam credentials or records of learning. That information is stored across all participating computers and can be viewed by anyone possessing the cryptographic 'public key' but cannot be modified, even by the original author. The data records are timestamped, providing a trusted and timed record of the added data.

Last, there are Smart Contracts, segments of computer code which enact blockchain transactions when certain conditions have been met. These enable business and legal agreements to be stored and executed online, for example to automate invoicing. In October, 2015 Visa and DocuSign demonstrated Smart Contracts for leasing cars without the need to fill in forms.[1]

To explore the value of the blockchain for education, we take each of these elements separately, then examine how they fit together.

## 2 The Blockchain as a Distributed Digital Record

The distinguishing elements of the blockchain are that it is a single linked record of digital events, stored on each participating computer. It has the properties that:

- The entire record is distributed over a wide network of participating computers and so is resilient to loss of infrastructure;
- it is possible to confirm the identity of any addition or modification to the record;
- once a block has been added by consensus among participants, it cannot be removed or altered, even by the original authors;
- the events are publically-accessible, but not publically readable without a digital key.

An obvious educational use is to store records of achievement and credit, such as degree certificates. The certificate data would be added to the blockchain by the awarding institution which the student can access, share with employers, or link from an online CV. It provides a persistent public record, safeguarded against changes to the institution or loss of its private records. This opens opportunities for direct awarding of certificates and badges by trusted experts and teachers. The University of Nicosia is the first higher education institution to issue academic certificates whose authenticity can be verified through the Bitcoin blockchain [5] and Sony Global Education has announced development of a new blockchain for storing academic records [6].

The blockchain provides public evidence that a student identity received an award from an institutional identity, but does not, of itself, verify the trustworthiness of either party. A university could still award a bogus certificate or a student could still cheat in

[1] https://www.docusign.com/blog/the-future-of-car-leasing-is-as-easy-as-click-sign-drive/.

an exam. The blockchain solves a problem of rapidly and reliably checking the occurrence of an event, such as the awarding of a degree, but not its validity. However, just as MOOCs make teaching widely visible, so the blockchain may expose awarding bodies and their products to public scrutiny.

## 3 The Blockchain as a Proof of Intellectual Work

Consider a system where any person could lodge a public record of a 'big idea', such as an invention, a contribution to knowledge, or a creative work such as a poem or artwork. That record links to an expression of the work (e.g. the text or artwork). Each big idea is identified with its author, and timestamped to indicate when it was first recorded. Once lodged it cannot be modified, but it could be replaced by a later version.

This can act as a permanent e-portfolio of intellectual achievement, for personal use as a logbook, or to present to an employer. It also serves as a crowd-sourced method of patenting. There is no need for a person to make and prove claims for invention – the record is there to see. The startup company Blockai has already implemented a blockchain system to help creative workers register their work to protect it from copyright infringement [7].

The blockchain as record of intellectual work has resonances with the Xanadu project of Ted Nelson [8]. Conceived in the early 1960s, Nelson's vision was for a "digital repository scheme for world-wide electronic publishing" [9, p. 3/2] with aspects that go beyond the worldwide web including unbreakable links, attribution to authors, and micropayments for re-use of content. Each item in the Xanadu repository would be linked back to its author and the record would be stored across many locations to maintain availability in the case of disaster. Most of Nelson's 17 rules for Xanadu could be mapped onto the blockchain as a record of learning, e.g.: every user is uniquely and securely identified; permission to link to a document is explicitly granted by the act of publication; every record is automatically stored redundantly to maintain availability even in case of didaster; the communication protocol is an openly published standard.

A problem with the blockchain as a record of learning or intellectual effort is similar to that for its use as a digital store for certificates: it is proof of existence[2], but does not guarantee that the data held in the record is valid, authentic or useful. A user's claim to be the originator of an idea, invention claim or creative work could be contested, nor is there guarantee that the item is valuable or even interesting to others. This is a serious issue, but it is addressed by the academic community through processes of peer review and reputation management. Nelson proposed a payment and royalty mechanism for Xanadu. For the blockchain as a record of learning, we indicate a mechanism for intellectual credit and reputation.

## 4 The Blockchain as Intellectual Currency

Currently, the main use of the blockchain is as a mechanism for recording transactions of the Bitcoin digital currency. This is a public ledger that records Bitcoin transactions

[2] https://www.proofofexistence.com/.

(though it can store other types of record). Bitcoins, like traditional currencies, can be used to pay for products and services from merchants who accept them. Thus, Bitcoin micro-payments could be used as reward for small educational services, such as a student who carries out a peer assessment task being automatically rewarded [10].

But other commodities can have tradeable value, notably reputation [11]. Reputation is a foundation of the new digital economy, with companies such as AirBnB and Uber building trust through ratings and reviews. Amongst academics, reputation is already a tradeable commodity, with promotion and recruitment being based in part on reputation measured through number of citations and the H-index metric of publication impact.

Imagine that trading of scholarly reputation could be extended beyond the academic world and made the basis of an educational economy. Consider the following proposition. A new public blockchain is initiated to manage educational records and rewards, perhaps by a consortium of educational institutions and companies. Each recognized educational institution, innovative organization, and intellectual worker is given an initial award of 'educational reputation currency', which we will call Kudos. The initial award might be based on some existing (albeit crude) metric: Times Higher Education World Reputation Rankings for Universities, H-index for academics, Amazon author rank for published authors etc. An institution could allocate some of its initial fund of Kudos to staff whose reputation it wishes to promote. Each person and institution stores its fund of reputation in a virtual 'wallet' on a universal educational blockchain.

Then, any institution or individual can make a reputational transaction. For an educational institution such as a university, that might be the award of a degree or certificate, which would involve posting the certificate on the blockchain and also transferring some Kudos from awarding institution to the awardee. For individual, it could support an economy of online tutoring, with students paying a tutor for online teaching in financial (e.g., Bitcoin) currency, who would then pay the student in reputation (Kudos) for passing a test or completing the course. The Smart Contracts mechanism could allow such peer-to-peer micropayments to be made in a variety of currencies.

Any individual (not necessarily someone who already has reputational credit) can also post an item of note to the educational blockchain. It might be a creative or scholarly production, a work of art, or a great idea, which is timestamped and archived. Thus, a simple posting is a permanent record of authorship as well as an item in a personal, but shareable, e-portfolio.

In addition, an individual with reputation can decide to associate Kudos with one or more postings to the blockchain, up to the amount the person holds in their wallet. The amount would not be spent, but is an indication of the value of the work or idea. Other people might then transfer some of their reputational credit to the author, to boost the reputation of that person's artefact or idea. They might do that to promote or be associated with the idea, in a similar way to investing in a Kickstarter project, but with a currency of reputation.

A consequence is that the educational blockchain would provide a single universal record of lodged creative works or ideas, each associated with reputational credit. The amount of Kudos associated with each item indicates its value to the author and thus, if needed, its real world monetary value (e.g. for purchasing a copy of the creative work).

Lastly, reputation could be 'mined' by institutions, which stake part of their reputation on adding valid blocks to the chain (through a proof-of-stake algorithm) for which they are rewarded with additional Kudos. There is no limit in theory to the items that could be added to an educational blockchain – assignments, blog postings, comments – but there is computational cost in storing and maintaining a distributed educational record. That record is public, so anyone can determine how a person gained the reputation, and the rules for associating value are agreed by a consensus of the volunteers mining the blocks.

Such a reputational management system for education is not fanciful. Something similar, though without the blockchain and tradeable reputation, is in operation for The Open University iSpot citizen science site [12], where acknowledged wildlife experts are initially given a high reputational score on the platform and new users can earn visible reputation (indicated by reputation points as well as virtual badges) through making wildlife observations and validating the observations of others. This process of enhancing reputation on iSpot happens automatically and most of the computational complexity of managing an educational blockchain and reputation system could be hidden from the user or institution.

We have been experimenting in adding OpenLearn badges[3] to a private blockchain. OpenLearn hosts over 800 free Open University courses and attracts over 5 million visitors per year. Our Open Blockchain platform is implemented on the open source Ethereum infrastructure[4] which supports the creation of Distributed Applications comprising sets of Smart Contracts. Our system currently allows students to register for courses and receive badges which can be viewed in a student Learning Passport. An administration interface enables awarding of badges to students. All transactions are timestamped and are cryptographically signed. The transactions are peer-to-peer: in principle no host institution is required for the awarding of accreditation. Future work will integrate badges from other institutions including FutureLearn[5] and optionally place badges onto the public Ethereum blockchain.

## 5 Implications

What might be the implications for education of trusted distributed educational records combined with a system of tradeable reputation? The first benefit is in providing a single secure record of educational attainment, accessible and distributed across many institutions. Once there is a recognised educational blockchain, then individuals as well as institutions could store secure public records of personal achievement. Second, a generalized system of reputation management associated with blockchain technology could help to open up the system of scholarly reputation currently associated with academics. This will require thought to develop accepted and trusted practices of acquiring public reputation, but there are already of examples of reputation management at work in

3 http://www.open.edu/openlearn/get-started/badges-come-openlearn.
4 https://www.ethereum.org/.
5 http://www.futurelearn.com.

companies such as AirBnB as well as in educational systems including iSpot. Third, and more controversially, reputation could be traded, by being associated with academic awards, as well as being put up as collateral for important ideas or to validate the adding of new block to the chain.

There are deep practical and ideological issues raised by trading educational reputation as a currency. One practical problem is how to create a conversion rate between reputation and money. What is the financial value of a novel idea or an A* dissertation? A fundamental ideological concern is that a system of trading reputation will further entrench the commodification of education – where students browse, buy and consume educational products, with no empathy for scholarship or intellectual value. Yet it could be argued that reputation as a commodity has long been a part of academia, though citation counts, impact factors, and national research assessment exercises. The blockchain and reputational currency might reduce education to a marketplace of knowledge, or they might extend the community of researchers and inventors to anyone with good ideas to share.



## References

1. Jones, H.: Broker ICAP says first to use blockchain for trading data. Reuters, London, 15 March 2016. http://uk.reuters.com/article/us-icap-markets-blockchain-idUKKCN0WH2J7
2. Valenzuela, J.: Arcade City: Ethereum's Big Test Drive to Kill Uber. The Cointelegraph, 15 March, 2016. http://cointelegraph.com/news/arcade-city-ethereums-big-test-drive-to-kill-uber
3. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System, October 2008. http://www.cryptovest.co.uk/resources/Bitcoin%20paper%20Original.pdf
4. Buterin, V.: Understanding Serenity, Part 2: Casper, 28 December 2015. https://blog.ethereum.org/2015/12/28/understanding-serenity-part-2-casper/
5. University of Nicosia. Academic Certificates on the Blockchain. http://digitalcurrency.unic.ac.cy/free-introductory-mooc/academic-certificates-on-the-blockchain/
6. Sony Global Education. Sony Global Education Develops Technology Using Blockchain for Open Sharing of Academic Proficiency and Progress Records, 22 February 2016. http://www.sony.net/SonyInfo/News/Press/201602/16-0222E/index.html
7. Ha, A.: Blockai uses the blockchain to help artists protect their intellectual property, TechCrunch, 15 March 2016. http://techcrunch.com/2016/03/14/blockai-launch/
8. Struppa, D.C., Douglas R. D.: Intertwingled: The Work and Influence of Ted Nelson. SpringerOpen (2015)
9. Nelson, T.H.: Literary machines. Mindful Press, Sausalito (1993)

10. Devine, P.: Blockchain learning: can crypto-currency methods be appropriated to enhance online learning? In: ALT Online Winter Conference, 7th–10th December (2015)
11. Schlegel, H.: Reputation Currencies. Institute of Customer Experience. http://ice.humanfactors.com/money.html
12. Clow, D., Makriyannis, E.: iSpot Analysed: Participatory Learning and Reputation. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, 28 Feburary – 01 March 2011, Banff, Alberta, pp. 34–43 (2011)

# Game-Based Training for Complex Multi-institutional Exercises of Joint Forces

Alexander Streicher[1(✉)], Daniel Szentes[1], and Alexander Gundermann[2]

[1] Fraunhofer IOSB, Karlsruhe, Germany
{alexander.streicher,daniel.szentes}@iosb.fraunhofer.de
[2] KIT, Karlsruhe, Germany
alexander.gundermann@student.kit.edu

**Abstract.** This paper presents a new concept for modular, Web-based training tools for complex, multi-institutional joint forces exercise scenarios, based on the motivational principles of digital game based learning (serious games). In multi-national, large-scale exercises for NATO Joint Intelligence, Surveillance, and Reconnaissance (JISR) various participants in different roles and backgrounds must understand the processes and information flow between the participating heterogeneous hardware systems and software appliances. The high variability of multi-dimensional requirements result in the need for pre-exercise preparation and training tools. Further, participants must be motivated to engage in the preceding training activities. This paper presents the modular concept for the game-based exercise training tool as well as its application for a real exercise scenario.

**Keywords:** Assistant systems · Serious games · Modularity · Joint training

## 1 Introduction

In multi-national, large-scale exercises for NATO *Joint Intelligence, Surveillance and Reconnaissance* (JISR) [9] various participants in different roles with different backgrounds must understand the processes and information flow between the participating heterogeneous hardware systems (e.g., air-borne drones) and software appliances (e.g., image processing tools). The high variability of vertical (different roles) and horizontal (varying complex interactions) requirements result in the need for pre-exercise preparation and training tools. The objective of these multi-national exercises (e.g., NATO interoperability projects CAESAR or MAJIIC) is to improve the overall interoperability between the technical systems of the partnering nations. For these complex exercises thorough preparation of the participants is needed to effectively conduct the exercise and reach all objectives. People with different professional backgrounds (e.g., civilian managers or army personnel) and roles (e.g., officers or trooper) must understand the exercise plans be able to answer the questions how, when, and why certain


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 497–502, 2016.
DOI: 10.1007/978-3-319-45153-4_49

systems are directly or indirectly interconnected. Complex data and information processes must be handled and understood. Although the exercises are normally thoroughly planned in advance, the often conducted last-minute changes and hot-fixes revealed the need for adapted preparation and training tools. Additionally the given pre-exercise documentation is too voluminous to be thoroughly studied by all participants. This raises the need for effective training tools which intrinsically motivate the exercise participants to conduct pre-exercise preparation and even on-site training (learning about the processes and information flow while the exercise is running).

The solution approach is to provide the participants with a game-based learning and training solution which is adapted to the exercise scenarios. This paper presents a new approach for a modular and adaptable game-based learning concept, called Exercise Trainer (EXTRA), which exploits the motivational aspects of serious gaming. Serious games introduce narratives and playful components to computer simulations or assistance systems to increase the users motivation to interact, e.g., by keeping the user in the *flow channel* and to increase their immersion [3,6]. The objective of EXTRA has been the development of technical and content-wise modular, game-based training concepts for interoperability exercises which can be easily adapted to changing requirements.

Our contribution is the concept for adaptable, modular concepts (technical and content-wise) for game-based training tools which target interoperability aspects, and the results of technical studies how to implement such training systems as Web-based applications using scenario description languages.

## 2 Related Work

Assistance and training applications for the handling of system-of-systems tasks have long been an active research topic. For example, the mobile scenario assistant SCENAS assists with the automatic configuration of complex systems for demonstration scenarios [8]. Furthermore, substantial results have been shown in the field of emergency training with game-based learning techniques [2,7].

Immersive training environments, i.e., digital game based learning systems (serious games), are increasingly being used by the military to provide training on a range of skills, team operations, navigation and route clearance, operationally relevant language skills, small unit tactical operations, and mission rehearsal [5].

So far, we could not find a combination of game-based technologies with scenario or exercise training applications for the military domain. In particular, the approach to combine scenario description languages and game-based learning as a training tool for multi-institutional exercises has not yet been presented.

## 3 Modular Game Design Concept

A modular (serious) game design concept allows authors to easily exchange parts or efficiently create new games. The goal is to achieve modularity on multiple levels to allow for an easy and effective transfer to other application scenarios.

Our game design [1] follows Prenskys proposal [6] to first define your audience and your learning objectives in order to find the ideal game genre which matches both, and therefore leads to a high rate of acceptance. The proposed modular game design concept is based on a structural design pattern approach to break down the concerns of the game design process for an exercise trainer, as proposed in this work, into three levels: technical, scenario and game level. Besides the breakdown of concerns at the game design process an additional modularity comes in by using standardized scenario representation formats.

At the technical level the general structure (business processes) of the exercises is modeled using standard modeling formats and tools from the *Modeling and Simulation* domain. We propose to use SysML [4] with the *XML Metadata Interchange* (XMI) format, which allows for effective interoperability of the UML2-based models. The interoperability is needed for the implementation of EXTRA as a generic training tool which must be flexible towards varying scenarios and game mechanics.

The scenario level entails the modeling of the actual scenario (e.g., roles or processes) which differ between different exercises. Whereas the actual scenario varies, the underlying technical model does not necessarily need to be modified, if the technical description of the processes and data flows remains the same (e.g., same business processes).

The game level imposes narratives, playful interactions, game mechanics, etc. on the scenario. In the case of EXTRA we propose to use a scalable game design based on logistics processes (details in the next sections).

## 4 Exercise Trainer (EXTRA)

The aforementioned modular game design concept for the Exercise Trainer (EXTRA) has been implemented to empirically verify its feasibility and flexibility towards changing requirements.

For our game design we used military terminology and metaphors which reflect the real JISR systems as game models, e.g., a factory paraphrases a processing unit, a market place paraphrases a data distribution facility. The semantic encoding of the metaphors ought to be obvious to the users, because non-intuitive object names could lead to an aggravated understanding which could negatively influence the playing experience. The EXTRA game concept is designed as an isometric, browser-based, turn-based simulation game (Fig. 2). The game objects can easily be exchanged to reflect other application scenarios. However, the general shape of the game is fixed to games for training of processes which can be reduced to logistics or business processes. The complex roles, activities, and processes, as well as the technical system-of-systems structure of complex and large scenarios are abstracted to a flexible game world, in which one has to build factories and logistics infrastructures (e.g., Fig. 2).

The definition of the learning objectives has been conducted in close cooperation with experts in NATO multi-national exercises, i.e., participants, operators, and planners. The main learning objective of EXTRA is to provide the users with information on the scenario and train them in preparation for the planned exercise. Of core importance is the mediation of knowledge on how the whole scenario is designed (macro-perspective), i.e., the involvements and interconnections of the (sub-)systems. Two main levels for the learning objectives in EXTRA have been identified: technical and procedural knowledge transfer. Whereas the technical view explains which systems are interconnected in which way, the procedural view looks at the different roles, processes, and activities. In the procedural mode EXTRA must mediate knowledge how a business process is modeled and executed. As an example, this could be the training of a process on imagery-based reconnaissance which includes the activities tasking, collection, processing, exploitation, and dissemination.

The goal of the game is to satisfy demanding "customers" (metaphor for essential users) with their changing product requests (metaphor for information requests) by constructing optimal logistic chains (metaphor for data or information interconnections) to optimally distribute the products to the markets. The narrative is called "Boston Harbor". It plays in a fictitious Boston, where at the famous harbor demanding international customers request certain products. A high-score contest motivates the players to repeatedly play the game as optimizing the logistics and optimally satisfying the customers demands increases the score (consisting of reputation and gold). When the demands are not satisfied in time, the score decreases; if the score drops to zero the game is lost. The learning objects are interweaved with the gameplay for not to impair the immersion. Hence, the terminology and characteristics of factories or connections reflect real world systems, and the gameplay transparently supports the training and the receptive knowledge transfer.

## 5 Application Example

We realized the EXTRA concept for a NATO multi-national joint exercise in 2015. Basis for the modeling of the scenario were exercise plans and general handbooks on NATO joint training exercises. The extracted example process for EXTRA is depicted in Fig. 1. It shows a high-level JISR example for an imagery acquisition process. The process starts with the collection requirement (i.e., what to collect intelligence for), to the asset mission planning (i.e., which sensor to task), to the actual acquisition and processing of data by an asset (e.g., a Recce Tornado), to the output handling with exploitation and dissemination.

In EXTRA this process has to be recreated in the game by the user using the available factories, market places or logistics centers and route types from the game inventory (Fig. 2). The user achieves the game's goal by collecting as many score points as possible. This can be achieved by optimally placing and interconnecting the available facilities, i.e., optimal according to the scenario description and technically verified by the game controller on basis of the underlying (technical) scenario model.

The concept proofed to be flexible to changing requirements. In a real application case, the learning objectives changed substantially. Whereas the original concept covered mostly technical aspects, the revised concept had to cover

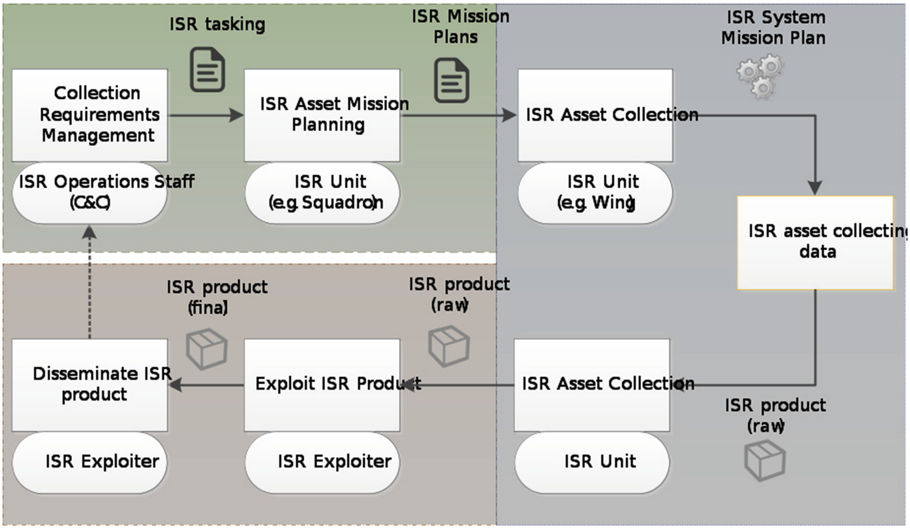



**Fig. 1.** Example process for a joint exercise for image acquisition with input handling (top left box), processing (right box), and output handling (bottom left box).

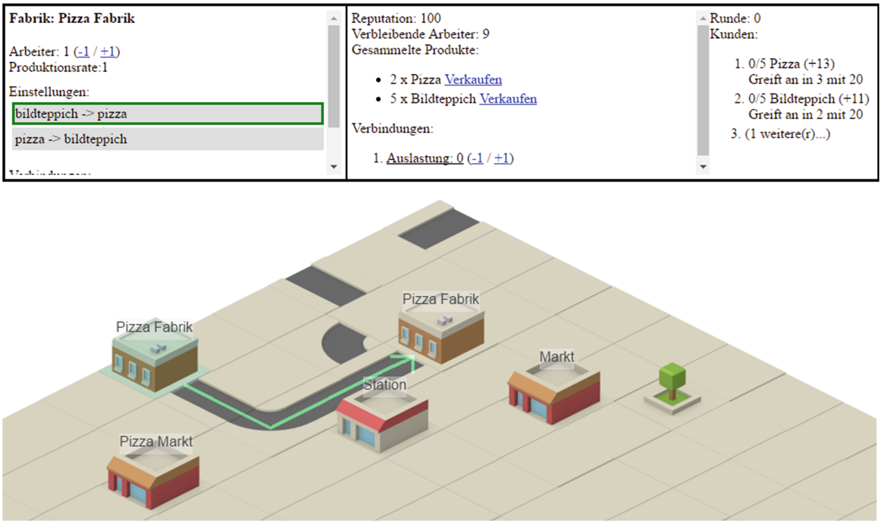



**Fig. 2.** Prototypical implementation of the EXTRA concept showing a cut-out of the running game (image sources from www.kenney.nl).

also procedural aspects. However, the changes in the game design concept could be kept minimal, since the game design is based on the modeling of logistics (business) processes. By adjusting only the metaphors for the game objects the game design could be easily adapted.

## 6 Conclusion

This paper presents a new concept for modular, Web-based exercise trainers for joint training scenarios based on the motivational principles of digital game based learning (serious games). In multi-national, large-scale exercises for NATO Joint Intelligence, Surveillance and, Reconnaissance (JISR) various participants in different roles with different backgrounds must understand the processes and information flow between the participating heterogeneous hardware systems and software appliances. The high variability of vertical and horizontal requirements results in the need for pre-exercise preparation and training tools. The application example of the implemented EXTRA concept for a NATO multi-national joint exercise according to a given exercise plan shows the feasibility of the presented concepts. Preliminary empirical application results show the flexibility of the concept towards changing requirements. The transfer of the EXTRA concept to other domains is subject of future work. Also, an evaluation is in preparation to verify the user acceptance and the learning effectiveness of EXTRA.

**Acknowledgments.** The underlying project to this article is funded by the Federal Office of Bundeswehr Equipment, Information Technology and In-Service Support under promotional references. The authors are responsible for the content of this article.

## References

1. Crawford, C., Peabody, S., Art, T., Web, W.W., Loper, D.: The Art of Computer Game Design. Computer, p. 81 (2003)
2. Crichton, M., Flin, R.: Training for emergency management: tactical decision games. J. Hazard. Mater. **88**(2), 255–266 (2001)
3. Csikszentmihalyi, M., Abuhamdeh, S., Nakamura, J.: Flow and the Foundations of Positive Psychology. Springer, Netherlands (2014)
4. Friedenthal, S., Moore, A., Steiner, R.: A Practical Guide to SysML: The Systems Modeling Language. Morgan Kaufmann, San Francisco (2008)
5. Hussain, T.S., Roberts, B., Menaker, E.S., Coleman, S.L.: Designing and developing effective training games for the US Navy. M&S J., p. 27 (2012)
6. Prensky, M.: Digital game-based learning. Comput. Entertainment (CIE) **1**(1), 21–21 (2003)
7. Stolk, D., Alexandrian, D., Gros, B., Paggio, R.: Gaming and multimedia applications for environmental crisis management training. Comput. Hum. Behav. **17**(5–6), 627–642 (2001)
8. Streicher, A., Szentes, D., Roller, W.: SCENAS - mobile scenario assistant for complex system configurations. In: International Conference on Theory and Practice in Modern Computing, MCCSIS 2014, pp. 157–165. IADIS, Lisbon, Portugal (2014)
9. US-Joint-Publication: Joint Intelligence (2007)

# Demo Papers

# DALITE: Asynchronous Peer Instruction for MOOCs

Sameer Bhatnagar[1], Nathaniel Lasry[2], Michel Desmarais[1(✉)], and Elizabeth Charles[3]

[1] Polytechnique Montréal, Montréal, Canada
{sameer.bhatnagar,michel.desmarais}@polymtl.ca
[2] John Abbott College, Montréal, Canada
[3] Dawson College, Montréal, Canada

**Abstract.** This demonstration will feature the Distributed Active Learning Integrated Technology Environment (DALITE), a novel LTI compliant application which allows Learning Management Systems to include an asynchronous peer instruction component as a part of their course. It has been successfully used in three different MOOCs on the edX platform (Harvardx, MITx, McGillx). This tool not only enables a novel type of formative assessment based on student self-explanations, but also provides a rich source of peer-assessed natural language data for educational research.

**Keywords:** Peer instruction · Massive open online classrooms

## 1 Introduction

One of the most widely accepted active learning pedagogical strategies is Peer Instruction (PI) [10]. The typical script followed by a teacher using PI:

1. teacher displays a multiple choice question item to their class, asking students to individually indicate their answer choice for what they think is the answer. This can be done using flash cards, signalling with fingers, or with wireless clickers. The intention is to give all students, no matter how introverted or confused, an opportunity to elicit their prior knowledge, anonymously
2. once all answer choices have been tallied, the teacher asks students to discuss with their neighbouring peers, and encourages them to convince one another of their own answer choice. After this discussion, teachers prompt students to once again, individually, indicate their answer choice (which may now be different than before).

The benefits of this as a classroom practice, especially in comparison to conventional, lecture-style content delivery, has been documented in different contexts [5,6,8,9]. It is with this success in mind, that our team of physics teachers and education researchers, working at colleges in Montreal, Canada, set out to develop a homework tool that would be centred on the same foundations of


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 505–508, 2016.
DOI: 10.1007/978-3-319-45153-4_50

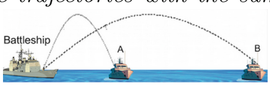
self-explanation, and intentional reflection surrounding a compare-and-contrast exercise. With the aim of delivering PI *asynchronously*, after several iterations [3,4] of Design Based Research [1], we present the most recent implementation of the Distributed Active Learning Integrated Technology Environment (DALITE).

## 2 DALITE

A DALITE question item proceeds as follows:

1. The question is displayed, and the student selects one of the multiple choice answers. They are then prompted to write a couple of sentences that explain why they selected their answer choice. These little paragraphs will from now on be referred to as "rationales" (Fig. 1).
2. Once a rationale is given, the system presents two sections of text: one for their answer choice, and one for another choice to the question (Fig. 2). Each section upto contains four rationales, written by previous students. The goal is to give students a chance to reflect on their thinking by providing them with an opportunity to compare and contrast other rationales, and maybe change their mind. The student is prompted to read the rationales from the two sections, and decide whether they would like to keep their answer choice, or switch. What's more, the student is asked to vote on one rationale out of the ones displayed, that they best like (They always have the option "I stick with my rationale").

*A battleship simultaneously fires two shells with different initial speeds at enemy ships. If the shells follow the parabolic trajectories with the same maximum height shown below, which ship gets hit first?*




*A - ship A*
*B - ship B*
*C - Both ships get hit simultaneously*
*D - Not enough information is given*
*Rationale:*

**Fig. 1.** DALITE: asynchronous peer instruction, part 1

The rationales displayed are anonymous, and can either be randomly selected from those in the database, or preferentially based on how many times they have been "upvoted" in the past. An important consideration is that any new question item requires a few "seed" rationales for each of the answer choice options, so as that the first students attempting it do not get an empty re-vote page.

*You answered **A**, and gave this rationale:*

*The closer the ship, the sooner it gets hit!*

*Consider the problem again, noting the rationales below that have been provided by other students. They may or may not, cause you to reconsider your answer. Read them, and select your final answer.*

– *A*
  - *"Battleship A must get hit first, since it is closer"*
  - *"they both have about the same maximum height, so since A is closer, it will get hit first"*
  - *I stick with my own rationale*

– *C*
  - *"the parabola of shell A has a different curvature than that of shell B, but the same x-intercepts. Hence mathematically they must land at the same time"*
  - *"The shells are fired at different speeds, but since they reach the same maximum height, the vertical component of their initial speed must be the same. Since "time in air" of any projectile depends only on initial vertical velocity, both shells spend the same amount of time in the air"*

**Fig. 2.** DALITE: asynchronous peer instruction, part 2

## 3 Scalable Asynchronous PI

In previous studies, we have shown that

– DALITE is as effective as in-class Peer Instruction for Quebec college level physics courses [4] (in terms of gain on the Force Concept Inventory [7])
– students appreciate the usefulness of the platform for formative assessment
– teachers are able easily integrate DALITE into "flipped-classroom" pedagogy
– weak students and strong students alike write rationales in DALITE that earn the votes of their peers [2]
– the tool provides a novel source of data for the Educational Data Mining, Learning Analytics, and Natural Language Processing research communities. Since students are constantly "up-/down-voting" their peers' rationales, there is a bootstrapping effect for the social annotation of constructed response data.

DALITE is now an open-source, Django-based web application, written to be compliant with the IMS Global Learning Consortium's *Learning Tools Interoperability (LTI)* standard, so that most major Learning Management Systems (LMS) can implement *asynchronous* PI, as an external resource. Over the past year, DALITE has been used on the edX platform as part of three different MOOCs (*Justice* from Harvardx, *Advanced Classical Mechanics* from MITx, and *Intro to Body* from McGillx). The tool is being successfully used in science items, but also contexts where there isn't necessarily a correct answer. In both *Justice* and *Intro to Body*, DALITE was used to elicit student opinions on ethical and scientific issues. The "up-voting" process allows instructors and students to easily determine which rationales are seen as most convincing by the participants of the course.

**Acknowledgements.** This work is funded by Entente Canada-Quebec, and the *Programme de Recherche sur l'Enseignement et l'Apprentissage* (PAREA) from the Government of Quebec. The development of the LTI compliant tool in the Django framework was financed by HarvardX. The user studies were made possible by the participating teacher researchers: Michael Dugdale (John Abbott College), Kevin Lenton (Vanier College), and Chris Whittaker (Dawson College).

## References

1. Anderson, T., Shattuck, J.: Design-based research a decade of progress in education research? Educ. Res. **41**(1), 16–25 (2012)
2. Bhatnagar, S., Desmarais, M., Whittaker, C., Lasry, N., Dugdale, M., Charles, E.S.: An analysis of peer-submitted and peer-reviewed answer rationales, in an asynchronous peer instruction based learning environment
3. Charles-Woods, E., Whittaker, C., Dugdale, M., Lasry, N., Lenton, K., Bhatnagar, S.: Designing of dalite: bringing peer instruction on-line. In: Rummel, N., Kapur, M., Nathan, M., Puntambekar, S. (eds.) Computer Supported Collaborative Learning (2013)
4. Charles-Woods, E., Whittaker, C., Dugdale, M., Lasry, N., Lenton, K., Bhatnagar, S.: Beyond and within classroom walls: designing principled pedagogical tools for students and faculty uptake. In: Computer Supported Collaborative Learning (2015) (in press)
5. Crouch, C.H., Mazur, E.: Peer instruction: ten years of experience and results. Am. J. Phys. **69**(9), 970–977 (2001)
6. Fagen, A.P., Crouch, C.H., Mazur, E.: Peer instruction: results from a range of classrooms. Phys. Teach. **40**(4), 206–209 (2002)
7. Hestenes, D., Wells, M., Swackhamer, G.: Force concept inventory. Phys. Teach. **30**(3), 141–158 (1992)
8. Kortemeyer, G.: The psychometric properties of classroom response system data: a case study. J. Sci. Educ. Technol. 1–14 (2016)
9. Lasry, N., Mazur, E., Watkins, J.: Peer instruction: from Harvard to the two-year college. Am. J. Phys. **76**(11), 1066–1069 (2008)
10. Mazur, E., Hilborn, R.C.: Peer instruction: a user's manual. Phys. Today **50**, 68 (1997)

# Digital and Multisensory Storytelling: Narration with Smell, Taste and Touch

Raffaele Di Fuccio[1], Michela Ponticorvo[1(✉)], Fabrizio Ferrara[2], and Orazio Miglino[1]

[1] Department of Humanistic Studies, University of Naples "Federico II", Naples, Italy
michela.ponticorvo@unina.it
[2] Department of Psychology, Second University of Naples, Caserta, Italy

**Abstract.** Storytelling is a methodology which exploits narration to give meaning and sense to reality. It is omnipresent in human culture and it finds relevant application in pedagogy. Telling children stories helps them to understand the world, to learn about their culture, to vehiculate specific concepts, to reflect upon experiences. In an educational context, storytelling facilitates literacy, building shared meanings between adults and children. In recent years, thanks to technological development, digital tools have been included in the storytelling process, giving life to digital storytelling. A relevant feature of storytelling, both traditional and digital, is the chance to put together the cognitive dimension with the emotional one. For this reason it has been employed to gather attention from people with profound intellectual and mental disabilities too. The emotional dimension can be indeed useful for everyone, to attract children, to increase their motivation and to improve learning.

For these reasons, we propose STTory, a hardware/software system for multisensory storytelling with smell, taste and touch that has been tested during a pilot study. Users feedback indicate that the use of more senses improved motivation.

**Keywords:** Digital storytelling · Multisensory storytelling · Technology enhanced learning · Motivation

## 1 Introduction

*"If you want your children to be intelligent, read them fairy tales. If you want them to be more intelligent, read them more fairy tales"*. This quote from Albert Einstein illustrates perfectly how stories are important to help children to develop their own abilities. Telling stories, an activity which is omnipresent in human culture, helps acquiring language, sharing meanings with the community, giving sense to reality.

Storytelling therefore finds relevant application in pedagogy [2], as it can become a powerful tool for communication, collaboration, and creativity between children and for teachers and children. A very relevent feature possessed by storytelling is the chance it offers to stimulate at the same time the cognitive and


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 509–512, 2016.
DOI: 10.1007/978-3-319-45153-4_51

the emotional dimension, which is, especially in life first years, very important to guarantee children harmonic growth [10]. In recent years, thanks to technologial development, digital tools have been applied to storytelling, thus giving birth to digital storytelling which has gained a respectable position between instructional tools. Moreover, many commercial products for storytelling have had a notable success; consider for example interactive books such as LivingBooks or authoring tools such as StoryMaker.

Even if these tools are indeed effective, there are some elements that are neglected in these applications, first of all the interaction with the physical world. Some solutions have been proposed to overcome this limit, consider for example the Interactive Storytelling Spaces for children proposed by Alborzi and colleagues [1] where room-sized immersive storytelling experiences for children is realized, or I-theatre [9], a collaborative storytelling system that allows children to draw their own characters and scenarios on paper and see them animated in a digital story.

These efforts constitute a relevant step forward, but, in our opinion, digital storytelling can be further enriched including multisensory elements. During the whole lifetime, it is important to stimulate all the senses, whereas some of them are neglected in modern societies were sight and hearing are undiscussed protagonists. Some psycho-pedagogical practices underline the important role of all senses proposing dedicated activities. Consider, for example, Montessori sensorial area [8] in the classroom with olfactory activities which aim at stimulating the sense of smell in children or tasting materials to foster this sense.

In the storytelling context, multisensory elements have been introduced in storytelling originating multisensory storytelling, which is widely employed to support children and adults with special needs [5]. For example, it has been employed to gather attention from people with profound intellectual and mental disabilities [11]. In this case, multisensory stories are personalized and thus stimulate the senses, adapting to the abilities, needs and desires of the individual with disabilities. Moreover, as it touches the emotional dimension of these individuals, it helps keeping the person focused on the story.

Given these premises, we propose STTory, a hardware/software system for digital and multisensory storytelling with smell, taste and touch. These senses which are usually neglected in digital applications, are strictly connected, also at neural level, to the emotional dimension. In next section the tool is described.

## 2 *STTory*: A Tool for Digital and Multisensory Storytelling

The tool we propose for digital and multisensory storytelling is an ICT device that blends the digital world with the real one. It is based on STELT platform, Smart Technologies to Enhance Learning and Teaching [6,7], that combines the management of hardware components (sensors and actuators) and software components (libraries for the storyboard and provision of feedback, authoring systems to be used by nonprogrammers). STELT implements augmented reality systems based on RFID (radio-frequency identification) and NFC (near field communication) technology, introduced below. The labels RFID/NFC (tags) are very thin transponders that can be applied to any type of object and are detected by small readers. The reader can be connected to a computer with either a wired or wireless connection or integrated into standard equipment on smartphones and tablets (NFC sensor). STELT combines communication protocols with the various hardware devices (readers and output devices), a storyboarding environment for creating various interaction scenarios, a database for tracking user behaviour and an adapting tutoring system that can build a user profile providing customised feedback. STELT platform has been used to implement different products, such as Block Magic, a hybrid physical/software tool that enhances traditional blocks and methods for teaching in kindergarten and primary schools [3,4]. *STTory*, represented in Fig. 1, is composed by:

1. the active table able to recognize real objects enhancing the physical materials using the RFID technology;
2. the software with the intelligent artificial tutor which manages the feedbacks;
3. the tangible objects;
4. the smelling jars;
5. the tasting jars;


**Fig. 1.** *STTory* hardware and software, see text for explanation

## 3 Pilot Test

A pilot test was run with this tool in a science fair at *Citta' della Scienza*, a cultural initiative to promote and popularize scientific knowldege in Naples, to test system usability and reception by users. About 40 users used the system with sight and hearing only or with smell, taste and touch. Users feedback indicate that the use of more senses improved motivation and engagement.

## 4 Conclusions and Future Directions

Digital and Multisensory storytelling has the potential to become a powerful instructional tool as it can have a strong impact on learners' understanding, motivation and recall. Next steps will be devoted to experimentally evaluate this claim comparing different sensory condition in a school context.

**Acknowledgments.** The project has been run under INF@NZIA DIGI.tales, funded by Italian Ministry for Education, University and Re-search under PON-Smart Cities for Social Inclusion programme.

## References

1. Alborzi, H., Druin, A., Montemayor, J., Platner, M., Porteous, J., Sherman, L., Kruskal, A.: Designing storyrooms: interactive storytelling spaces for children. In: Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 95–104. ACM, August 2000
2. Coulter, C., Michael, C., Poynor, L.: Storytelling as pedagogy: an unexpected outcome of narrative inquiry. Curriculum Inq. **37**(2), 103–122 (2007)
3. di Ferdinando, A., di Fuccio, R., Ponticorvo, M., Miglino, O.: Block magic: a prototype bridging digital and physical educational materials to support children learning processes. In: Uskov, V.L., Howlett, R.J., Jain, L.C. (eds.) Smart Education and Smart e-Learning. Smart Innovation, Systems and Technologies, vol. 41, pp. 171–180. Springer, Heidelberg (2015)
4. Di Fuccio, R., Ponticorvo, M., Di Ferdinando, A., Miglino, O.: Towards hyper activity books for children. Connecting activity books and montessori-like educational materials. In: Conole, G., Klobučar, T., Rensing, C., Konert, J., Lavoué, É. (eds.) Design for Teaching and Learning in a Networked World, pp. 401–406. Springer, Heidelberg (2015)
5. Fornefeld, B.: Storytelling with all our senses. In: Using Storytelling to Support Children and Adults with Special Needs: Transforming Lives Through Telling Tales, p. 78. Routledge, London (2012)
6. Miglino, O., Di Ferdinando, A., Schembri, M., Caretti, M., Rega, A., Ricci, C.: STELT (smart technologies to enhance learning and teaching): a toolkit devoted to produce augmented reality applications for learning teaching and playing. Sistemi Intelligenti **25**(2), 397–404 (2013)
7. Miglino, O., Di Ferdinando, A., Di Fuccio, R., Rega, A., Ricci, C.: Bridging digital and physical educational games using RFID/NFC technologies. J. e-Learn. Knowl. Soc. **10**(3), 87–104 (2014)
8. Montessori, M.: The Montessori Method. Transaction Publishers, Piscataway (2013)
9. Muñoz, J., Marchesoni, M., Costa, C.: i-Theatre: tangible interactive storytelling. In: Camurri, A., Costa, C. (eds.) INTETAIN 2011. LNICST, vol. 78, pp. 223–228. Springer, Heidelberg (2012)
10. Susman, E.J., Feagans, L.V., Ray, W.J. (eds.): Emotion, Cognition, Health, and Development in Children and Adolescents. Psychology Press, New York (2013)
11. ten Brug, A., van der Putten, A., Penne, A., Maes, B., Vlaskamp, C.: Multisensory storytelling for persons with profound intellectual and multiple disabilities: an analysis of the development, content and application in practice. J. Appl. Res. Intell. Disabil. **25**(4), 350–359 (2012)

# A Platform for Social Microlearning

Bernhard Göschlberger[1,2](✉)

[1] Research Studios Austria FG, Linz, Austria
bernhard.goeschlberger@researchstudio.at
[2] Johannes Kepler University, Linz, Austria

**Abstract.** In the 21st century the web has evolved from a producer-consumer oriented information source to a prosumer centric social web filled with user generated content. To overcome potential loss of quality assurance on the producer side successful social web solutions came up with methods to ensure content quality using wisdom of the crowd. Although the success of this revolution is undisputed a vast majority of e-learning systems are still producer-consumer oriented and therefore impede engagement potential. We propose to use interaction patterns of successful social web solutions to create a platform that motivates students to create and share learning activities. As we will argue, microlearning activities are especially well suited for such a platform. We also demonstrate how to design such a system open and interoperable by using xAPI and a flexible authentication concept.

**Keywords:** Microlearning · Social learning · Crowd sourcing · Question posing · xAPI

## 1 Introduction

The evolution of the Internet towards a space of more democratic information exchange has ultimately led to its society-changing success. Whilst called Web 2.0 earlier the term *social web* is nowadays used more often, as it better reflects the social nature of the process of creating and sharing information resources. Accordingly the term *social software* has been coined for software that enables groups to form and self-organize in a bottom-up manner (cf. [1, 2]).

As of today social network sites (SNS) are the predominant form of social software on the web. Two success factors for SNS are the simplicity and immediate graspability of its content artifacts. Twitter – considering itself as a micro-blogging service – became more popular than other blogging services as it restricted tweets to 140 characters. Hence, the cognitive load per tweet for both creators and consumers is reduced. This lowers the barrier to initiate social interaction by sharing on the one side and enables the consumers to quickly decide whether content is relevant to them on the other side.

In this paper we present a prototype for social microlearning that tries to incorporate successful strategies and common features of social software.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 513–516, 2016.
DOI: 10.1007/978-3-319-45153-4_52

## 2 Background

Microlearning focuses on short-term and informal learning activities using small, but self-explanatory learning resources that are available via Internet [3, 4]. Microlearning implementations oftentimes use learning activities similar to flashcards (e.g. Mobler Cards [5, 6], KnowledgePulse [7]). Flashcards are generally associated with behaviorist learning style and lower-level cognitive functions. In Bloom's revised taxonomy [8] the act of learning a flashcard (in drill mode) represents an act of remembering. To promote understanding – a higher-level learning objective – the aforementioned microlearning implementations enhanced the traditional flashcards enriching them with explanation, insight and/or feedback. Moreover, they implemented a variety of features aimed at engaging students in higher order cognitive tasks such as reflection, self-regulation, content evaluation and content creation. In order to evaluate or create learning content a learner already needs a good understanding of the subject. Baumgartner [9] proposes the model of a competence spiral. In a first step learners have to absorb basic knowledge about a topic or subject (Learning I), before being able to actively acquire knowledge about that topic in a self-determined manner (Learning II) and finally being able to construct knowledge in a third step (Learning III). With the learner proceeding to more advanced concepts this process is repeated on a higher level (Learning I+). Baumgartner remarks relations between Learning I and behaviorism, Learning II and cognitivism, and Learning III and constructivism.

A key challenge for microlearning systems is to motivate students to progress through these phases as each phase implies different requirements for the system. Learning I requires the software to provide strict guidance and reduce complexity by limiting the degree of freedom. In Learning II phase the learner takes control over his learning process. Guidance is reduced to recommendation. Learning III phase includes the construction of new knowledge. Therefore the system needs to support students to contribute, evaluate and discuss. The prototype presented in the following section is a first step towards a system addressing students' needs throughout the three phases.

## 3 Social Microlearning Platform

To validate the pedagogical model and evaluate best practices in design and usability for social microlearning we decided to prototypically implement a platform for our experiments. The developed platform prototype aims to provide a social space for microlearning activities. Based on analysis of features and strategies of social software in literature (cf. [1, 10, 11]) we decided on an initial feature set for our prototype. Learners can (1) create and share, (2) evaluate, rate, comment and improve, (3) tag and collect, and (4) interact with and solve learning activities.

Before these capabilities are explained in depth, a few remarks about the implementation details are provided. The prototype frontend is developed using AngularJS, Bootstrap 3 and Material Design, providing a mobile first, responsive user interface. It uses a Spring Data REST Backend that uses MongoDB for persistency. All user interactions listed above are logged to a learning record store (LRS) using xAPI. Fine grained user interactions such as mouse clicks are logged directly by the frontend and persistent user interactions such as content creation are logged by the backend. Amongst other options, Shibboleth is used for authentication to facilitate experiments in the tertiary sector.

**Create and Share.** Through a simple interface users can create and share micro learning content. Shared content is presented as an inverse chronological stream in the main view. The system does not separate the processes of creating and sharing. Therefore it is not possible to use the system as a private content repository. The prototype currently supports only multiple-choice cards (single-select and multi-select). However, it is designed to support a great variety of micro learning content types in the future. Creating and sharing learning content aligns with the highest level in Bloom's revised taxonomy.

**Evaluate, Rate, Comment and Improve.** Existing content items can be rated using a simple up/down-vote mechanism commonly used in social software. To enable students to express their thoughts on particular items each item has a comment section. These comments themselves can also be rated by up/down-vote. This approach has been proven very effective and is well accepted on e.g. stackoverflow.com, an online social Q&A system. Authors can edit and improve their content items based on these inputs. A last-edited-remark denotes that an item has been edited. Previous versions remain available as a version history to all users by clicking the last-edited remark. These activities align with the second and third highest level in Bloom's revised taxonomy.

**Tag and Collect.** To organize existing learning content relevant to them, students can tag items. Tags can be chosen arbitrary. The user interface supports the student by offering tags previously used by the student on any content item or by other students on the respective content item as autocompletions. The user can browse through his tags in the myTags-view and through the collection of items annotated with the tag by clicking a tag. Tagging and collecting is an act of curation and aligns with fourth and fifth level in Bloom's revised taxonomy.

**Interact and Solve.** Students can interact with the provided micro-content. In the case of multiple-choice questions this means that they can check and uncheck options. Once they chose an answer they can submit and resolve. This can be repeated any number of times. Interacting and solving simple micro-content items, such as multiple-choice questions is initially a task of remembering and therefore on the lowest level of Bloom's revised taxonomy. However, it triggers any higher order activities described above in students that have passed through the Learning I phase already.

## 4 Future Work

Currently the prototype is used to validate the pedagogical model. It does not yet filter the shared content. To use it beyond isolated experimental settings restricted to certain topics, it is however necessary to identify communities and filter content based on those community structures. For students in Learning I phase additional guidance needs to be provided. Therefore it will be necessary to extract and use information provided by more advanced learners and/or historical data (traces) of other learners. Moreover it is planned to implement user statistics to foster reflection and self-regulation.

## References


1. Ziovas, S., Grigoriadou, M., Samarakou, M.: Supporting Learning in Online Communities with Social Software: An Overview of Community Driven Technologies. INTECH Open Access Publisher (2009)
2. Boyd, S.: Are You Ready for Social Software? Darwin Magazine (2003)
3. Kovachev, D., Cao, Y., Klamma, R., Jarke, M.: Learn-as-you-go: new ways of cloud-based micro-learning for the mobile web. In: Leung, H., Popescu, E., Cao, Y., Lau, R.W., Nejdl, W. (eds.) ICWL 2011. LNCS, vol. 7048, pp. 51–61. Springer, Heidelberg (2011)
4. Hug, T.: Micro learning and narration: exploring possibilities of utilization of narrations and storytelling for the design of "micro units" and didactical micro-learning arrangements. In: Proceedings of Media in Transition (2005)
5. Glahn, C.: Supporting learner mobility in SCORM-compliant learning environments with ISN Mobler cards. Connect. Q. J. **12**(1) (2012)
6. Glahn, C.: Using the ADL experience API for mobile learning, sensing, informing, encouraging, orchestrating. In: 2013 Seventh International Conference on Next Generation Mobile Apps, Services and Technologies (NGMAST). IEEE (2013)
7. Bruck, P.A., Motiwalla, L., Foerster, F.: Mobile learning with micro-content: a framework and evaluation. In: 25th Bled eConference eDependability: Reliable and Trustworthy eStructures, eProcesses, eOperations and eServices for the Future, pp. 17–20 (2012)
8. Anderson, L.W., Krathwohl, D.R., Bloom, B.S.: A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Longman, New York (2001)
9. Baumgartner, P.: Educational dimensions of microlearning–towards a taxonomy for microlearning. In: Designing Microlearning Experiences–Building up Knowledge in Organisations and Companies. Innsbruck University Press, Innsbruck (2013)
10. McLoughlin, C., Lee, M.J.: Social software and participatory learning: pedagogical choices with technology affordances in the Web 2.0 era. In: ICT: Providing Choices for Learners and Learning. Proceedings Ascilite Singapore 2007 (2007)
11. Boulos, K., Wheeler, S.: The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Inf. Libr. J. **24**(1), 2–23 (2007)

# A Framework to Enhance Adaptivity in Moodle

Ioannis Karagiannis(✉) and Maya Satratzemi

Department of Applied Informatics, University of Macedonia, 54006 Thessaloniki, Greece
{karagiannis,maya}@uom.edu.gr

**Abstract.** The purpose of this paper is to present a framework that can be used to embed an adaptivity mechanism to Moodle so as to achieve better learning results. One of the main innovations is that a hybrid dynamic user model is adopted which is built with techniques that are based both on learner knowledge and behaviour. The proposed mechanism adapts the presentation and the recommended navigation within a course, to students' different preferences as they are expressed by their learning styles and their educational objectives.

**Keywords:** AEHS · LMS · E-learning · Blended learning · Learning styles · Adaptivity · Static student modelling · Dynamic student modelling · Progress calculation

## 1 Introduction

E-learning systems can be divided into two categories according to the level of personalized services they offer. More specifically, there are systems like Learning Management Systems (LMS) which totally ignore a student's learning style, and deliver the same set of resources to all students. On the other hand, Adaptive Educational Hypermedia Systems (AEHS) consider learning styles and try to adapt the educational resources in order to enhance the learning process.

Learning styles refer to attitudes and behaviors which determine the way an individual learns something new. There are many references [2, 7] about the significance of learning styles and their impact on the learning process. The Felder-Silverman Learning Style Model (FSLSM) [1] is used far more than any other in AEHS, mainly because it describes learning styles in much more detail. There are four dimensions each with two scales: active/reflective, sensing/intuitive, verbal/visual and sequential/global, according to the way students process, perceive, receive and understand information. The Index of Learning Styles (ILS), which is a 44-item questionnaire, was developed in order to assess FSLSM [1].

The purpose of our paper is to present the design of a framework that can be used to embed an adaptivity mechanism to Moodle so as to achieve better learning results. This mechanism adapts content's presentation and navigation within a course, to students' different preferences as they are expressed by their learning styles and their educational objectives.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 517–520, 2016.
DOI: 10.1007/978-3-319-45153-4_53

The remainder of the paper is organized as follows: related work is presented in Sect. 2. In Sect. 3 a description of the proposed framework is given. Lastly, in Sect. 4 are the conclusions.

## 2 Related Work

A user model can be built, either statically or dynamically, with techniques that are based on the knowledge or behaviour of the learner [6]. More specifically, student modelling could be achieved by analyzing his behaviour data [2, 7]. On the other hand, there are researchers [5] who focus on learners' knowledge, while at the same time considering information about learning style as this arises from questionnaires.

Substantial efforts that took place were [2, 5, 7]. Popescu developed the WELSA system which is an AEHS that adapts educational resources to the learning styles of users [7]. Graf attempted to exploit the advantages of LMS and combine them with those of AEHS, proposing the use of adaptation techniques in Moodle [2]. Although the FSLSM was used in this case, its visual/verbal dimension was ignored in the development of educational resources, mainly because it is time-consuming [2]. This, however, may result in erroneous outcomes as the educational process is not fully personalized. Kazanidis and Satratzemi developed the ProPer system which is a SCORM-based AEHS that adapts presentation and navigation according to a complex user model where learners' knowledge, educational objectives and learning style are represented [5].

## 3 Framework of the Adaptivity Mechanism

Taking into account our research findings [4], it was decided to embed adaptivity techniques in Moodle rather than develop a new AEHS. One of the main innovations is that we decided to adopt a hybrid dynamic user model. The term "hybrid" is used because the model is built with techniques that are based both on learner knowledge and behaviour. In order to implement static modelling, the learner has to answer ILS and to declare his/her objectives at the beginning of the course. Regarding dynamic modelling, data comprising the number of visits to each type of learning object and the duration of these visits, are used as input in a decision tree algorithm. Besides mining behaviour data, dynamic modeling implies knowledge progress calculation.

In order to match diversity of learning styles, it was decided to use seven different types of learning objects: outlines, content objects, videos, solved exercises, quizzes, open-ended questions, and conclusions. Regarding the structure of the course, it was decided to use a sufficiently flexible mechanism [3], which has been modified to correspond to our needs. Thus, the proposed structure of the course consists of sections, each with a different theoretical concept, and its own learning objects. Immediately after the outline at the beginning, there is what we call the "area before content" whose aim is to stimulate the learner to become actively involved in this section. This in turn is followed by the content objects. Then there is the "area after content".

In our model, the adaptation features deal only with the position of the learning objects in the particular section. As regards the "area before content" solved exercises, videos, open-ended questions and short quizzes are chosen as adaptation features. These specific features can attract the learner's attention according to their learning style preferences. The next four features concern the "area after content", which are related to the position of the solved exercises, quizzes, videos and open-ended questions. Two more features concerning the specific area were also included. The first is outlines appearing not only at the beginning of a section but also between the content objects, and the second is the conclusion appearing either right after the content objects or at the end of the section.

A well-established methodology, found in [2] was adopted and modified to suit our needs. Thus, a matrix with one row for each adaptation feature and one column for each dimension pole of the FSLSM is built. The matrix cells are filled in as follows: 1 if the adaptation feature supports the specific learning style, −1 if the feature should be avoided in order to support the specific learning style, and 0 if the feature has no effect on the learning style according to the literature [1]. Learning styles obtained from the ILS questionnaire ($LS_{ILS}$) as well as those that were derived from mining the behaviour data with the decision tree algorithm ($LS_{AD}$) were considered regarding the input of the algorithm. More specifically, we add up the respective values of the adaptation matrix for each of the adaptation features, by firstly considering $LS_{ILS}$ and then the $LS_{AD}$, and the final ranking score is equal to their average.

The "area before content" consists of the learning object, which has the biggest ranking score from the equivalent features. As regards the "area after content", all learning objects are ranked in descending order according to their score, which determines their positioning within the specific area. Besides adaptive sorting, adaptive annotation is also used. According to what the learner state as his/her learning objectives for the specific course, a respective icon appears before the learning object link. Adaptive annotation also implies the different annotation of the links of the objects considered as having been learnt, according to the procedure of calculating the learner's knowledge progress (Fig. 1).

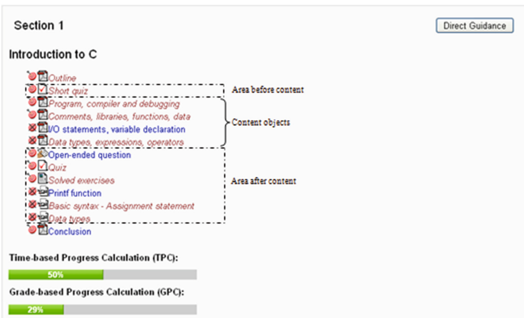



**Fig. 1.** Screenshot of a course section in Moodle

The proposed procedure of knowledge progress calculation implies the use of two different measures: the Time-based Progress Calculation (TPC) and the Grade-based Progress Calculation (GPC). TPC and GPC are depicted in each section with the help of two independent progress bars As regards the first measure, Moodle's authoring tool was extended enabling it to store for each type of learning object two different time values, namely t_min and t_max. Additionally, one more value named w is stored which indicates the weight of importance of the specific learning object and it ranges from 0 to 1. The t_min value represents the minimum time that is required for a learner to study the specific learning object in order for it to be considered as "known". The t_max value represents the maximum time that a learner can study it. Thus, if a time value exceeds the t_max limit, the specific time value will not be considered. Therefore, depending on whether a learning object is considered as "known" or "unknown", it is assigned the value of 1 or 0 respectively. Each value is multiplied by the respective w value and TPC is the average of these values. The second measure of a learner's progress (GPC) refers to his/her performance on Moodle activities that can be graded such as quizzes and open-ended questions. Due to possible different specifications of a course, it was decided that these grades would have adjustable weights in the GPC. Therefore, the GPC is the weighted average of these grades.

## 4 Conclusion

There is a growing tendency to take a student's individual characteristics in e-learning systems into consideration. In our contribution to research, we suggest a system that combines usability, extendibility and the support community of an LMS with the personalization capabilities of an AEHS. In order to achieve this, we embedded an adaptivity mechanism to Moodle so that the sequence and presentation of a course's learning objects can be adapted to students' learning styles. Moreover, the presentation of learnt learning objects is also adapted to students' knowledge level in terms of the two proposed measures, namely TPC and GPC.

## References

1. Felder, R.M., Silverman, L.K.: Learning and teaching styles in engineering education. Eng. Educ. **78**(7), 674–681 (1988)
2. Graf, S.: Adaptivity in learning management systems focusing on learning styles. Ph.D. dissertation, Vienna University of Technology, Vienna, Austria (2007)
3. Graf, S., Kinshuk, C.I.: A flexible mechanism for providing adaptivity based on learning styles in learning management systems. In: 10th IEEE International Conference on Advanced Learning Technologies (ICALT 2010), Sousse, Tunisia, pp. 30–34 (2010)
4. Karagiannis, I., Satratzemi, M.: Comparing LMS and AEHS: challenges for improvement with exploitation of data mining. In: 14th IEEE International Conference on Advanced Learning Technologies (ICALT 2014), Athens, Greece, pp. 65–66 (2014)
5. Kazanidis, I., Satratzemi, M.: Adaptivity in ProPer: an adaptive SCORM compliant LMS. J. Distance Educ. Technol. **7**(2), 44–62 (2009)
6. Kobsa, A., Koenemann, J., Pohl, W.: Personalised hypermedia presentation techniques for improving online customer relationships. Knowl. Eng. Rev. **16**(2), 111–155 (2001)
7. Popescu, E., Bădică, C., Moraret, L.: WELSA: An Intelligent and Adaptive Web-Based Educational System. In: Papadopoulos, G.A., Badica, C. (eds.) IDC 2009. SCI, vol. 237, pp. 175–185. Springer, Heidelberg (2009)

# Refugees Welcome: Supporting Informal Language Learning and Integration with a Gamified Mobile Application

Hong Yin Ngan[1(✉)], Anna Lifanova[2], Juliane Jarke[2], and Jan Broer[2]

[1] University of the Arts Bremen, Bremen, Germany
ngan@uni-bremen.de
[2] University of Bremen, Bremen, Germany
{sergeeva,jarke,jbroer}@uni-bremen.de

**Abstract.** This paper describes *Moin as* a user-centered design process to develop a mobile application to foster informal learning through face-to-face communication, supported with contextual language learning features and employing gamification as a motivator. The aim of *Moin* is to help refugee teenagers to integrate to German culture and specifically to the region of Bremen. The final product requirements are based on the findings from the state-of-the-art, literature analysis and semi-structured interviews. The overall goal of *Moin* proposes that such a gamified digital application can support forming of local communities that create informal learning of local language and culture and, as a result, support local integration of migrants.

**Keywords:** Informal learning · Mobile learning · Gamification · Integration · Refugees · Migrants

## 1 Introduction

In the 14 months since the beginning of 2015 about 600,000 refugees sought asylum in Germany [1]. Bremen, the focus of our research and one of Germany's three city states, was allocated over 6,500 asylum seekers from Syria and neighbouring countries [1]. This migration poses a number of challenges: Different languages, cultural habits, and life experiences as well as a lack of information keep refugees largely separated from the local society [2].

On the other hand, over 86 % of young refugees own a mobile handset, and more than 50 % use the internet at least once per day [3]. This access has the potential to assist with some of the issues of the migration process, such as cultural barriers, social norms, and the integration into a new society. One of the ways to make integration happen smoothly and in a friendly way is to use technologies for creating informal communication opportunities between people while taking into account language barriers and cultural differences [4].

Applications dedicated to migrants are not new. Learning a language, communicating, and making new friends in a new environment can also be supported by a digital


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 521–524, 2016.
DOI: 10.1007/978-3-319-45153-4_54

device. Clough et al. have shown that smartphone users use their devices to support a wide range of informal learning activities [5]. According to Livingston informal learning is any activity involving the pursuit of understanding, knowledge or skill which occurs outside the curricula of educational institutions [6].

Gamification has gained some recognition in the past 5 years as a motivator for learning. Gamification refers to the "use of game-design elements in non-game contexts" [7]. To our knowledge, no studies exist that discuss the use of gamification in informal language learning for integration purposes. We intend to help close this research gap with our development and evaluation of *Moin* - an informal learning and communication interface that incorporates elements of games (see Fig. 1).

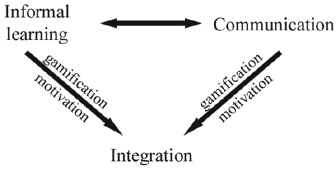



**Fig. 1.** Connection between main terms in a process of integration using *Moin*

## 2 Methods and Implementation

In order to understand the needs of our target group, we performed a total of 33 semi-structured interviews [8]. Refugee teenagers reported few contacts with local people, and high difficulties in communication. All of them expressed the wish to integrate into local culture and to meet people. Most of the interviewees had smartphones with Android platform. According to self-reports, they use the devices mostly for the following activities: learning German, translation from German to their native language, communication via various media, entertainment and information about Germany. The interviewed volunteers pointed out the language barrier – it gets in the way of communication of refugee teenagers with German people. The interviewed communication experts stated that the most important way for teenage communication is a face-to-face informal communication. Interviewed school teachers also highlighted language barriers in the communication with refugees. The interviewed German students did express an interest in communicating with refugee teenagers.

With regards to our state of the art research result on apps which are explicitly for migrants and refugee, only 6 apps target to Germany and 2 apps target Europe overall. Only 2 apps contain communication elements such as registration for refugee events or registration for a job search. However, no communication elements explicitly for refugees and local population were found. None of the applications from above are focused on teenagers or young people.

Therefore, we have developed *Moin* – a mobile application explicitly for Bremen that enables and motivates both local and migrant teenagers to meet for social events and provides some assistance with contextual language learning.

*Moin* allows users to create or join events, with the goal of bringing people together that share the same interests and thereby create opportunities for informal learning through face-to-face communication. Figure 2a shows some examples of such events as seen in the application. Users who join an event can communicate with other participates from the event (see Fig. 2b). *Moin* also contains a direct learning element. Users can choose to learn vocabulary related to various event types – such as festivals in Bremen (see Fig. 2c), scenarios for ordering food, and culture and excursion. *Moin* contains a variety of gamification elements to motivate users to use the application. The progress bar and badges will reward users for their participation and thereby create extrinsic motivation. Users can decide to create and participate in events, make comments, use the chat function to chat with friends and learn vocabulary (see Fig. 2d).

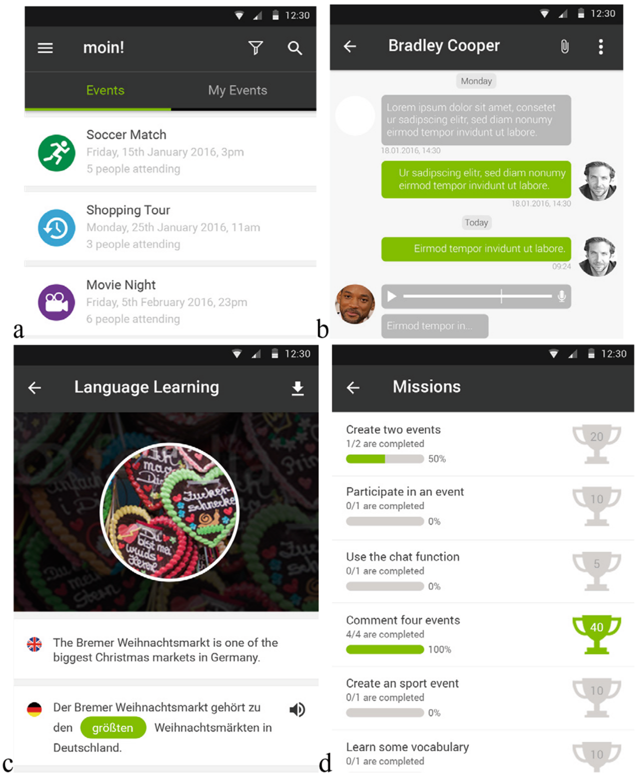



**Fig. 2.** Screenshots from the application. (a) Events, (b) Chatbox, (c) Language Learning, (d) Gamification elements

The prototype was created using Android studio, SDK version 5.0.1 (API 21) and minimum SDK version 4.0.3 (API 15), using PHP 5.5 server-side scripting, apache server 2.4.0 and a MySQL 5.1.61 database.

A standard usability test with the think-aloud method and observations was used for a brief evaluation of the prototype. 10 users participated, five from each of the target groups. The usability tests revealed a satisfactory score on the system usability scale, but also a variety of possible improvements. Especially interesting was that the group of German users had no issues using the application, while the migrants did. This result may well be a function of our low number of participants, but might also hint at a need to keep different usability aspects in mind when developing applications for this target group.

## 3 Conclusion

To sum up, literature analysis revealed that integration depends on the quantity of social relationships of the person and is hindered by language and cultural barriers. The application therefore had to increase real-life communication situations with both local and migrant young people. One of the methods for increasing the motivation to use of such a product is gamification.

While our research has shown that these effects are likely to occur and we have designed the application accordingly, an actual benefit has yet to be shown. So far, only brief usability tests have been performed with the target group. Future research should be aimed at prolonged use of the application and informal learning effects thereof. The usability tests also indicated an unforeseen gap in usability requirements between participating locals and refugees. Further research is needed into the question of user interface design for these target groups.

## References

1. Bundesamt für Migration und Flüchtlinge (BAMF): Asylgeschäftsstatistik 02/2016 (2016). http://www.bamf.de/DE/Infothek/Statistiken/Asylzahlen/asylzahlen-node.html
2. UN High Commissioner for Refugees (UNHCR): Facilitators and Barriers: Refugee Integration in Austria (2013). http://www.refworld.org/docid/5278dc644.html
3. Maitland, C.: A social informatics analysis of refugee mobile phone use: a case study of Za'atari Syrian Refugee Camp. In: TPRC 43: The 43rd Research Conference on Communication, Information and Internet Policy Paper (2015)
4. Strang, A., Ager, A.: Refugee integration: emerging trends and remaining agendas. J. Refugee Stud. **23**(4), 589–607 (2010)
5. Clough, G., Jones, A.C., Mcandrew, P., Scanlon, E.: Informal learning evidence in online communities of mobile device enthusiasts. In: Ally, M. (ed.) Mobile Learning: Transforming the Delivery of Education and Training, pp. 99–112. Athabasca University Press, Edmonton (2009)
6. Livingstone, D.W.: Exploring the icebergs of adult learning: findings of the first Canadian survey of informal learning practices. Can. J. Study Adult Educ. **13**(2), 49–72 (1999)
7. Deterding, S., Sicart, M., Nacke, L., O'Hara, K., Dixon, D.: Gamification: using game-design elements in non-gaming contexts. In: Extended Abstracts on Human Factors in Computing Systems, CHI 2011, pp. 2425–2428. ACM (2011)
8. Personal interviews in refugee camp. Personal interview by A. Lifanova, M. Karayel, M. Yildirim (2015)

# DEDOS-Player: Educational Activities for Touch Devices

David Roldán-Álvarez[1], Estefanía Martín[2(✉)], Óscar Martín Martín[2], and Pablo A. Haya[3]

[1] Universidad Autónoma de Madrid, 28049 Madrid, Spain
david.roldan@uam.es
[2] Universidad Rey Juan Carlos, 28933 Móstoles, Madrid, Spain
estefania.martin@urjc.es, martinm.oscar@gmail.com
[3] Instituto de Ingeniería del Conocimiento, Campus Cantoblanco, 28049 Madrid, Spain
pablo.haya@iic.uam.es

**Abstract.** In recent years, touch devices are used in the educational field. Lots of efforts have been put into creating educational contents and applications for these kinds of surfaces. However, few authoring tools have been developed that allow the creation of educational activities which can be performed in them and other devices. This paper presents DEDOS, a toolset which allows teachers to design their own educational activities which can be performed on several devices (PCs, digital whiteboards, Android devices and multitouch tabletops). The adaptation of the educational activities to the device used is done automatically. Therefore, it is easy to use it since teachers do not need to configure anything.

**Keywords:** Authoring tool · Learning · Tablets · Multitouch tabletops

## 1 Motivation

The inclusion of technology in the classroom has proven to be very useful in the field of education. The use of ICTs in the education field improves the confidence and motivation of the students since computer aided applications promote errorless learning, offer an immediate and personalized assessment and let teachers adapt the rhythm of learning to each student [1]. In the last years, touch devices have emerged as an alternative to the traditional mouse. They allow the user to interact through natural gestures and manipulate the elements directly. Letting students express themselves in a more physical way generates better communication and comprehension [2] which allows them to focus on the contents and solve problems more quickly [3] while they enjoy doing the activities presented [4]. Thanks to the combination of these features and the appropriate multimedia content, users have control of the information and the interaction, which helps them to gain deeper knowledge of the topic presented [5].

In the literature we can find tablet apps such as those designed by Haro [6] or Lingnau [7]. Researchers observed that the number of interactions among the participants increased and that they were more motivated when using touch devices. In addition, some researchers studied how tablets could help students do their daily life activities [8], stating again that students were excited when using tablets and quickly gained independence to perform all the tasks they were asked to. In the literature, we can find few


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 525–528, 2016.
DOI: 10.1007/978-3-319-45153-4_55

web based authoring tools for e-learning, which indirectly can be used with tablets [9, 10]. However, there are not many Android based applications [11].

To solve this issue, we present a new toolset composed by DEDOS-Editor [12], which allows teachers to design their own educational activities, and DEDOS-Player, which let students perform those activities in multiple devices, including Android tablets and multitouch tabletops.

## 2 DEDOS

DEDOS project is formed by two tools. The first one, DEDOS-Editor, will allow the creation of educational and collaborative activities. The designed activities can be performed on multiple devices (PCs, digital whiteboards, Android tablets and multi-touch tabletops) by using DEDOS-Player. These tools put the creative power in the hands of teachers, who will design the activities having in mind the characteristics of the students who will perform them. DEDOS-Player adapts automatically the educational project to the device used by students without teachers having to do any additional configuration steps.

### 2.1 DEDOS-Editor

DEDOS-Editor is the authoring tool used by teachers to create and to share their educational activities in an easy, intuitive and flexible way without needing any technological knowledge. These activities are designed independently of the device where they will be performed, since it is DEDOS-Player the application that adapts the content to the device. With DEDOS-Editor, teachers can design four simple types of activities, which can be combined to create more complex activities, depending on the students' needs:

- Single and multiple choice: The teacher will present a question to the students with one or multiple responses as possible answers. For example, students may be required to choose the mammals among a set of animals provided.
- Pair-matching: Students will have to associate the concepts presented by the teacher. For example, in a recycling exercise, dropping each piece of litter to its corresponding container.
- Point connection: The students will have to follow a point path which has been drawn by the teacher in order to build a picture.
- Math activity: This type is similar to pair-matching activities. Students will have to drag and drop a determined number of elements (each one has a specific number value) until they total the amount requested by the teacher. This type allows creating addition exercises for the kids.

### 2.2 DEDOS-Player

When teachers want to propose the accomplishment of educational activities to their students, they use DEDOS-Player. This tool adapts automatically the contents of the educational project to the specific device used by the student. Figure 1 shows two examples of this tool. On the left side, a child is doing learning activities with a tablet. On the right hand, three students are collaborating doing musical activities using a multitouch tabletop.

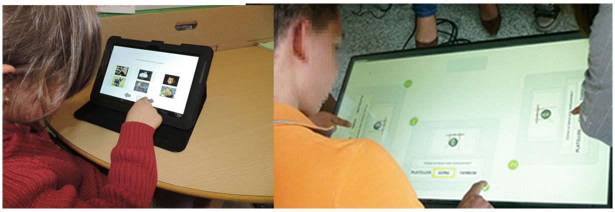


**Fig. 1.** Students performing activities using DEDOS-Player both tablets and tabletop

As it is shown, DEDOS-Player adapts the educational project dynamically to the number of users who are interacting with the device by letting the teacher choose the number of students before starting the project through a panel of options. The maximum number of students doing learning activities with a multitouch tabletop is four (one person per side) and if a tablet is used, DEDOS-Player considers that there is only one student doing the exercises. The activities are displayed in the broadest possible size considering the device characteristics.

These devices are useful in the educational area since students can manipulate the elements displayed with their own hands. In this sense, they do not need to know how to maneuver intermediary devices such as the mouse or the keyboard. This aspect is important especially when teachers are working with children and/or students with special needs.

## 3 Conclusion

This paper presents two complementary applications: an authoring tool (DEDOS-Editor) which allows teachers to become developers of their own educational projects, and a player (DEDOS-Player which allows the students to perform those activities). The flexibility of the authoring tool combined with the support for multiple devices allows teachers to adapt their educational project without extra effort. Furthermore, the use of touch devices in education is becoming more frequent. Therefore, technical solutions for supporting multiple devices are needed in order to facilitate the teachers' work.

**Acknowledgements.** The research presented in this paper has been funded by the Spanish Ministry of Economy and Competitiveness under grant agreement TIN2013-44586-R, "*e-Training y e-Coaching para la integración socio-laboral*" and by Comunidad de Madrid under project S2013/ICE-2715.

## References


1. Ornellas, A., Sancho, J.: Three decades of digital ICT in education: deconstructing myths and highlighting realities. In: Myths in Education, Learning and Teaching: Policies, Practices and Principles, p. 135 (2015)
2. Cantón, P., González, A.L., Mariscal, G., Ruiz, C.: Applying new interaction paradigms to the education of children with special educational needs. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012, Part I. LNCS, vol. 7382, pp. 65–72. Springer, Heidelberg (2012). doi:10.1007/978-3-642-31522-0_10
3. Inkpen, K.M., Ho-Ching, W., Kuederle, O., Scott, S.D., Shoemaker, G.B.: This is fun! We're all best friends and we're all playing: supporting children's synchronous collaboration. In: Hoadley, C.M., Roschelle, J. (eds.) 1999 Conference on Computer Support for Collaborative Learning. International Society of the Learning Sciences (1999). Article no. 31
4. Africano, D., Berg, S., Lindbergh, K., Lundholm, P., Nilbrink, F.: Designing tangible interfaces for children's collaboration. In: Extended Abstracts on Human Factors in Computing Systems, CHI 2004, pp. 853–868. ACM, New York (2004). doi: 10.1145/985921.985945
5. Roldán-Álvarez, D., Márquez-Fernández, A., Rosado-Martín, S., Martín, E., Haya, P.A., García-Herranz, M.: Benefits of combining multitouch tabletops and turn-based collaborative learning activities for people with cognitive disabilities and people with ASD. In: IEEE 14th International Conference on Advanced Learning Technologies, pp. 566–570 (2014)
6. Haro, B.P.M., Santana, P.C., Magaña, M.A.: Developing reading skills in children with down syndrome through tangible interfaces. In: Proceedings of the 4th Mexican Conference on Human-Computer Interaction, pp. 28–34. ACM (2012)
7. Lingnau, A., Zentel, P., Cress, U.: Fostering collaborative problem solving for pupils with cognitive disabilities. In: Proceedings of the 8th International Conference on Computer Supported Collaborative Learning, pp. 450–452. International Society of the Learning Sciences (2007)
8. Edler, C., Rath, M.: People with learning disabilities using the iPad as a communication tool - conditions and impact with regard to e-inclusion. In: Miesenberger, K., Fels, D., Archambault, D., Peňáz, P., Zagler, W. (eds.) ICCHP 2014, Part I. LNCS, vol. 8547, pp. 177–180. Springer, Heidelberg (2014)
9. Gordillo, A., Barra, E., Gallego, D., Quemada, J.: An online e-Learning authoring tool to create interactive multi-device learning objects using e-Infrastructure resources. In: 2013 IEEE Frontiers in Education Conference, pp. 1914–1920 (2013)
10. Little, B.: Effective and efficient mobile learning: issues and tips for developers. Ind. Commercial Training **44**(7), 402–407 (2012)
11. Torrente, J., Serrano-Laguna, Á., Fisk, C., O'Brien, B., Alesky, W., Fernández-Manjón, B., Kostkova, P.: Introducing Mokap: a novel approach to creating serious games. In: 5th ACM International Conference on Digital Health 2015, pp. 17–24 (2015)
12. Roldán-Álvarez, D., Martín, E., García-Herranz, M., Haya, P.A.: Mind the gap: impact on learnability of user interface design of authoring tools for teachers. Int. J. Hum. Comput. Stud. **94**, 18–34 (2016)

# The Booth: Bringing Out the Super Hero in You

Jan Schneider(✉), Dirk Börner, Peter van Rosmalen, and Marcus Specht

Welten Institute, Open University of the Netherlands, Heerlen, The Netherlands
{jan.schneider,dirk.boerner,peter.vanrosmalen,
marcus.specht}@ou.nl

**Abstract.** The acquisition of knowledge is a key aspect for learners. However, in moments of stress the cognitive capacities of learners decrease considerably, making it very difficult for learners to get access and use their already acquired knowledge. Therefore we developed *The Booth*, a prototypical toolkit designed to support learners to prepare for key situations that are foreseen to be stressful. It guides the learner through a series of lectures helping them to gain personal power, and become in touch with their sincere-self. This paper presents the prototype, including the description of its technical aspects and the theory behind its lectures.

**Keywords:** Sensor-based learning · Affective computing · Self-confidence · Demonstration

## 1 Introduction

Common educational practices and research focus mostly on the acquisition of knowledge leveraging the cognitive domain of learning, while regularly ignoring the affective and psychomotor domain of learning [1]. This major focus on content acquisition is also reflected in the research area of technology enhanced learning (TEL), where most TEL technologies focus on the acquisition of content [2]. A key objective for the acquisition of knowledge is to be able to apply it when needed. Through live learners face events that require full use of cognitive capacities. These events are in many cases stressful, leaving learners feeling powerlessness. The feeling of powerlessness activates the behavioral inhibition system, forcing learners to focus on threats rather than on opportunities. Learners therefore tend to become anxious, pessimistic and susceptible to social pressures, forcing them to be less in touch with their sincere-selves [3]. It also undermines executive functions such as reasoning, task flexibility, attention control [4] and keeps learners post processing the event days later [5].

To avoid feeling powerlessness, research has shown that at some point the learner should stop preparing content and start preparing mindset [6]. Therefore, we developed *The Booth* in order to support learners with their mindset preparation for situations that can be foreseen as stressful. *The Booth* is a prototypical tool that guides learners through a set of lectures designed to make them feel in touch with their most sincere-self and regain their personal power.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 529–532, 2016.
DOI: 10.1007/978-3-319-45153-4_56

## 2 The Booth

*The Booth* is a system designed as a confidence booster. Its current version consists of six small lectures or exercises that can be completed by the learner in five to eight minutes. The featured exercises are: Super Hero Posture, Super Powers, Inspiration 1, Inspiration 2, Saving the Planet, and Celebration. In order to interact with the system the learner makes use of postures and gestures. This interaction is possible though the use of the Microsoft Kinect[1] sensor.

### 2.1 Lecture: Super Hero Posture

Body language does not only communicate to others it also communicates to ourselves. Expansive body language increases optimism, assertiveness and resilience while reducing stress [7]. It improves our strengths, skills, decision taking and perception [8]. The study in [9] describes how participants who were asked to stay in expansive body postures that express power prior to a job interview, significantly outperformed participants who did not use the power postures before the interview.

The first lecture consists on teaching the learner the super hero posture (see Fig. 1), which requires the learner to smile, and stand straight, with spread legs, hands on hips. During the remaining lectures the system requests learners to remain in a power posture.

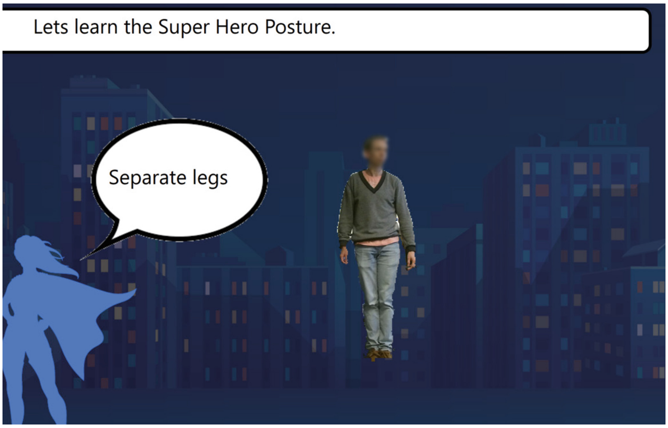



**Fig. 1.** The booth teaching the super hero posture.

### 2.2 Lectures: Super Powers, Inspiration 1 and Inspiration 2

Research has shown that acting powerful, being exposed to words related with power and reflecting about times when one was feeling powerful helps learners to prepare for cognitive by improving their performance [3, 10]. Another preparation strategy that

[1] https://developer.microsoft.com/en-us/windows/kinect/develop.

helps learners to prepare their mindset is to get in touch with their sincere-self [11]. The study in [12] shows that getting in touch with core values through self-affirmation also supports mindset preparation of the learner. This strategy significantly decreases the learner's stress levels.

The purpose of the lectures Super Powers, Inspiration 1 and Inspiration 2 is to help learners to reduce stress and improve their performance. In order to achieve this during the lectures, learners have to select and reflect about concepts that find inspiring and align with their values, while standing in a powerful posture (see Fig. 2).

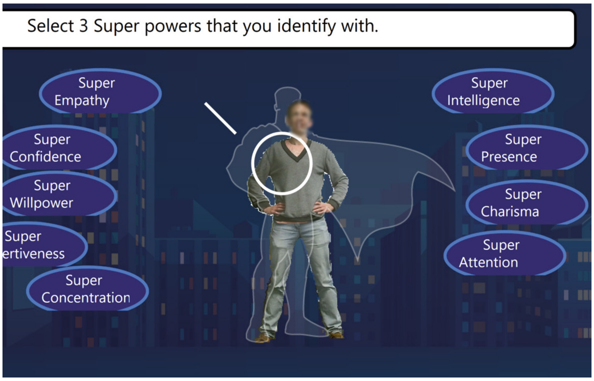



**Fig. 2.** Super power selection lecture

### 2.3 Lecture: Saving the Planet

A warm and trustworthy person who is strong and competent elicits admiration. Nevertheless, only after establishing trust strength and competence become a gift rather than a threat [13]. The saving the planet lecture has the purpose to elicit the sense of kindness and warmth by asking the learner to reflect on how to save the world.

### 2.4 Lecture: Celebration

During this lecture the learner is asked to stand in a celebrating posture raising both arms in a V posture, while remembering and reflecting about winning and achieving goals.

## 3 Conclusions and Future Work

Research has shown that the mindset has a significant influence on the learner's performance [4]. To support learners to obtain a right mindset to approach foreseen challenges we developed *The Booth*. We based its development on research that has already shown to help individuals to regain their personal power to their biggest challenges. For future work we plan to explore the usage of the system for scenarios such as public speaking that are usually considered as stressful.

**Acknowledgement.** The underlying research project is partly funded by the METALOGUE project. METALOGUE is a Seventh Framework Programme collaborative project funded by the European Commission, grant agreement number: 611073 (http://www.metalogue.eu).



## References

1. Wirth, K.R., Perkins, D.: Learning about thinking and thinking about learning. In: Innovations in the Scholarship of Teaching and Learning at the Liberal Arts Colleges, St. Olaf and Carleton College, MN, USA, pp. 16–18 (2007)
2. Schneider, J., Börner, D., van Rosmalen, P., Specht, M.: Augmenting the senses: a review on sensor-based learning support. Sensors **15**(2), 4097–4133 (2015)
3. Cuddy, A.: How powerlessness Shackles the self (and how power sets it free). In: Presence: Bringing Your Boldest Self to Your Biggest Challenges. Hachette, UK (2015)
4. Derakshan, N., Eysenck, M.W.: Anxiety, processing efficiency, and cognitive performance: new developments from attentional control theory. Eur. Psychol. **14**(2), 168–176 (2009)
5. Gaydukevych, D., Kocovski, N.L.: Effect of self-focused attention on post-event processing in social anxiety. Behav. Res. Ther. **50**(1), 47–55 (2012)
6. Dilon, K.: What you should (and shouldn't) focus on before a job interview (2015). Harvard business review https://hbr.org/2015/08/what-you-should-and-shouldnt-focus-on-before-a-job-interview. Accessed Mar 2016
7. Cuddy, A.: The body shapes the mind (So starfish up!). In: Presence: Bringing Your Boldest Self to Your Biggest Challenges. Hachette, UK (2015)
8. Arnette, S.L., Ii, T.F.P.: The effects of posture on self-perceived leadership. Int. J. Bus. Soc. Sci. **3**(14), 8–13 (2012)
9. Cuddy, A.J., Wilmuth, C.A., Yap, A.J., Carney, D.R.: Preparatory power posing affects nonverbal presence and job interview performance. J. Appl. Psychol. **100**(4), 1286 (2015)
10. Smith, P.K., Jostmann, N.B., Galinsky, A.D., van Dijk, W.W.: Lacking power impairs executive functions. Psychol. Sci. **19**(5), 441–447 (2008)
11. Cuddy, A.: Believing and owning your story. In: Presence: Bringing Your Boldest Self to Your Biggest Challenges. Hachette, UK (2015)
12. Cohen, G.L., Sherman, D.K.: The psychology of change: self-affirmation and social psychological intervention. Annu. Rev. Psychol. **65**, 333–371 (2014)
13. Cuddy, A.J., Kohut, M., Neffinger, J.: Connect, then lead. Harvard Bus. Rev. **91**(7), 54–61 (2013)

# DojoIBL: Nurturing Communities of Inquiry

Angel Suarez(✉), Stefaan Ternier, Fleur Prinsen, and Marcus Specht

Welten Institute, Open University of the Netherlands, Heerlen, The Netherlands
{angel.suarez,stefaan.ternier,fleur.prinsen,marcus.specht}@ou.nl

**Abstract.** This paper presents and outlines the demonstration of DojoIBL, a web-based platform that aims at nurturing communities of inquiry by supporting communication and collaboration with emerging technological affordances. The manuscript briefly elaborates on the theoretical underpinning of DojoIBL and describes the functionalities supported. It concludes anticipating the follow up implementation which will consist on the integration of role support in DojoIBL.

**Keywords:** Mobile supported collaborative learning · Inquiry-based learning · Context-awareness · Interoperability · Informal learning

## 1 Introduction

The study [1] emphasized the collaborative nature of learning, arguing that the creation of knowledge can be explained as products of social interactions. Inquiry-based learning (IBL) [2] is certainly this, a collaborative process where students engage in social interactions to co-create knowledge around shared essential questions. Nowadays, these processes are supported by technology, which offers a whole new range of possibilities for learning. Yet, not all have been explored in the context of IBL. Hence, based on existing initiatives [3–5] and studies, this demo paper presents DojoIBL, a platform to nurture communities of inquires [6, 7], which combine essential inquiry elements with emerging technological affordances to support collaboration.

Inquiry-based learning is a methodology that often is characterized as a collaborative process where participants co-create knowledge by engaging in social interactions [8–10]. This was adequately coined in [6] with the term Community of Inquiry (CoI) [6, 7], which emphasizes that the creation of knowledge occurs within a social context and it requires social interactions among participants with different background information.

## 2 DojoIBL Affordances

DojoIBL is a Learning Content Management System, where students construct knowledge collaboratively using *atomic* inquiry elements to structure the inquiry processes. In DojoIBL, users design *blueprints* or *templates* of inquiries, meaning that different groups of students can work with the same inquiry structure in different topics. Those inquiry structures are organized in phases, which represents the steps that the students


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 533–536, 2016.
DOI: 10.1007/978-3-319-45153-4_57

need to follow in the inquiry. Figure 1 shows how the inquiry phases are represented in DojoIBL.

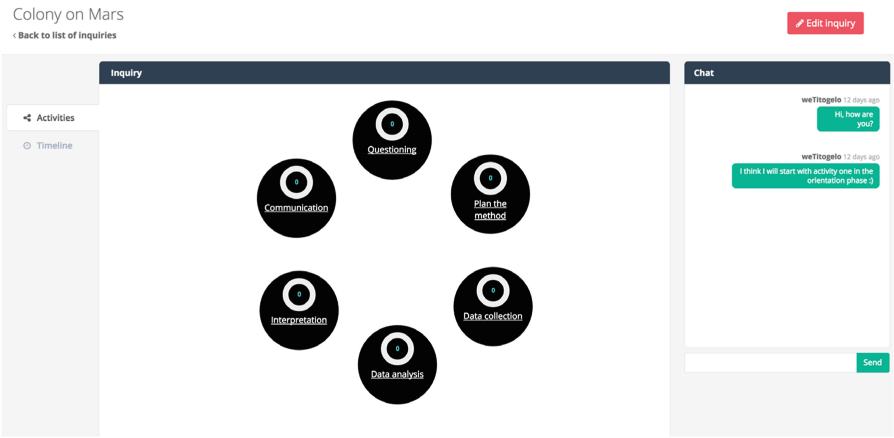



**Fig. 1.** Visualization of the inquiry process on the Colony on Mars activity

Within the inquiry phases, users can add specific atomic inquiry elements that are defined as the smallest re-usable type of resources available in DojoIBL to support specific pedagogical affordances. The selection of these inquiry elements has been done out of the experiences with students during the weSPOT project [17]. The following six types of elements have been implemented in DojoIBL because they were the most used by students:

- Discussion: forms the simplest type of activity which is based on plain text. Students can find a description, a story or a definition that inspire them about the specific topic.
- Research question: is an essential part of IBL where students collaboratively work around a shared question or topic.
- Data collection: enables the visualization and upload of data to DojoIBL.
- Concept map: concept mapping helps students to represent and organize knowledge and concepts around a topic.
- External plugin: enables the integration of external widgets repositories like GoLabs [5]. Those widgets provide the possibility to conduct scientific experiments in a virtual environment.
- Multimedia: similar to discussion activity but it adds the possibility to incorporate a multimedia element to inspire students.

Every atomic inquiry element is supported by a discussion functionality that the students can used to reflect, share and discuss about the activity itself.

Comparing DojoIBL with other existing platform, it adds value to the students' experience by integrating emerging technological affordances to support collaborative inquiries. The instant messaging system offers a contextualized communication channel to enable just in time text-based communications. It addressed, the three essential components of any educational transaction [11–13]; cognitive, social and teacher presence. In addition, DojoIBL implements an inquiry timeline and a notification system. They complement the support of collaboration by offering functionality to enable collaborative awareness. Two, social and action awareness, out of three types of awareness described in [14] are supported in DojoIBL. The third one, activity awareness will be addressed in future development cycles.

DojoIBL implements a *notification system* and an *inquiry timeline*. Both the timeline and the notification system, promote collaboration awareness based on social, action and activity awareness described in [15]. Inspired by patterns found in current existing social networks, DojoIBL integrates several functionalities to facilitate students' collaboration and communication combined with atomic inquiry elements.

The proposed inquiry workflow is based on the pedagogical IBL model developed in the weSPOT project [17]. This inquiry workflow can be adjusted and modified in DojoIBL using the edit inquiry function. Therefore, designers are able to add, remove and modify inquiry phases and activities in order to follow other IBL models.

## 3 Conclusions and Future Work

This manuscript presented DojoIBL, a Learning Content Management System that aims at nurturing 'Community of Inquiry' (CoI), by helping students to co-create knowledge through social interactions. It combined essential elements to support inquiry-based learning (IBL) with social collaborative tools in order to facilitate better collaborative processes. In short, DojoIBL focused on adding value to teachers and students' IBL experiences by providing a simple, intuitive and flexible tool.

As a future work, DojoIBL will implement the integration of role support [16] to enable testing the role taking strategy in CoI.

To conclude, this manuscript contributed DojoIBL, an open source platform that aims at fostering communities of inquiry for driving students' success facilitating the acquisition of the so called 21st century skills, e.g. communication and collaboration.



## References

1. Vygotsky, L.: Mind in Society. Harvard University Press, London (1978)
2. Bruder, R., Prescott, A.: Research evidence on the benefits of IBL. ZDM Math. Educ. **45**, 811–822 (2013)

3. Mikroyannidis, A., Okada, A., Scott, P., Rusman, E., Specht, M., Stefanov, K., Boytchev, P.: weSPOT: a personal and social approach to inquiry-based learning. J. Univ. Comput. Sci. **19**(14), 2093–2111 (2013)
4. Mulholland, P., Anastopoulou, S., Collins, T., Feisst, M., Gaved, M., Kerawalla, L., Paxton, M., Scanlon, E., Sharples, M., Wright, M.: nQuire: technological support for personal inquiry learning. IEEE Trans. Learn. Technol. **5**, 157–169 (2012)
5. Gillet, D., de Jong, T., Sotirou, S., Salzmann, C.: Personalized learning spaces and federated online labs for STEM education at school. In: 2013 IEEE Global Engineering Education Conference (EDUCON). pp. 769–773. IEEE (2013)
6. Peirce, C., Buchler, J.: Philosophical Writings of Peirce. Dover, New York (1955). Selected and Edited, with and Introduction, by Justus Buchler
7. Pardales, M.J., Girod, M.: Community of inquiry: its past and present future. Educ. Philos. Theor. **38**, 299–309 (2006)
8. Scardamalia, M., Bereiter, C.: Higher levels of agency for children in knowledge building: a challenge for the design of new knowledge media. J. Learn. Sci. **1**, 37–68 (1991)
9. Dillenbourg, P.: What do you mean by collaborative learning. Collaborative Learn. Cogn. Comput. Approaches **1**, 1–15 (1999)
10. Bell, T., Urhahne, D.: Collaborative inquiry learning: Models, tools, and challenges. Int. J. Sci. Educ. **32**(3), 349–377 (2010)
11. Garrison, D., Anderson, T., Archer, W.: Critical thinking, cognitive presence, and computer conferencing in distance education. Am. J. Distance Educ. **15**, 7–23 (2001)
12. Rourke, L., Anderson, T.: Assessing social presence in asynchronous text-based computer conferencing. Int. J. Distance Educ. **14**(3), 51–70 (2007)
13. Anderson, T., Rourke, L., Garrison, D., Archer, W.: Assessing teaching presence in a computer conferencing context (2001)
14. Carroll, J., Neale, D., Isenhour, P., Rosson, M., McCrickard, D.: Notification and awareness: synchronizing task-oriented collaborative activity. Int. J. Hum. Comput. Stud. **58**, 605–632 (2003)
15. Carroll, J., Rosson, M., Convertino, G., Ganoe, C.: Awareness and teamwork in computer-supported collaborations. Interact. Comput. **18**, 21–46 (2006)
16. Strijbos, J.-W., De Laat, M.F.: Developing the role concept for computer-supported collaborative learning: An explorative synthesis. Comput. Hum. Behav. **26**, 495–505 (2010)
17. Specht, M., Bedek, M., Duval, E., Held, P., Okada, A., Stefanov, K., Parodi, E., Kikis-Papadakis, K., Strahovnik, V.: weSPOT: inquiry based learning meets learning analytics (2013)

# Poster Papers

# Towards an Automated Assessment Support for Student Contributions on Multiple Platforms

Oula Abu-Amsha(✉), Nicolas Szilas, and Daniel K. Schneider

TECFA, Faculté de psychologie et de sciences de l'éducation,
Geneva University, Geneva, Switzerland
oula.abuamsha@heig-vd.ch, {nicolas.szilas,daniel.schneider}@unige.ch

**Abstract.** Varying learning activities beyond existing LMS's can improve the learning experience [1]. However, managing student interactions and productions across multiple platforms can be very time consuming. This contribution proposes a novel approach to monitor student productions on varied online platforms, such as social networks, Wiki pages, Google Docs. We rely on a combination of techniques: data is collected through web scraping or web APIs, then synthetic information and varied analysis are applied, and finally the results are presented through a web application. We applied our approach to a course where the students contribute on a private social network, Google Docs, and on a MediaWiki. The pilot is built with the R programming language.

## 1 The Context

The master program MALTT delivered by TECFA - University of Geneva builds on an active pedagogy: course-long analysis, synthesis and development activities, group work, interim feedback on deliverables, etc. Over the last few years, enrollment for classes did increase. In order to keep up with the active pedagogy, one possible solution is to rely on (semi) automatic methods to support the instructors' tutoring, monitoring and assessment activities.

Examples of desired automatic analysis include descriptive summaries of the contributions such as numbers, sizes and dates of submissions. Deeper analysis of the written content include lexicometric text analysis (e.g. lexical diversity [9], readability [5]), plagiarism detection [7], detection of concepts and keywords, use of references and citations, etc.

Monitoring tools already exist in most educational platforms. However, the active pedagogy applied at TECFA requires a monitoring approach that preserves the flexibility in platforms selection and contribution analysis. No full turn-key solution was required, but rather a modular and flexible system that could easily be adapted to emerging needs. The present paper informs about an approach that collects student contributions from different platforms, prepares the data for analysis, and then presents the results through a web application. The current system is a proof of concept and needs to be further developed.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 539–542, 2016.
DOI: 10.1007/978-3-319-45153-4_58

## 2 The Proposed Approach

Processing the student contributions goes through three consecutive phases represented in Fig. 1.

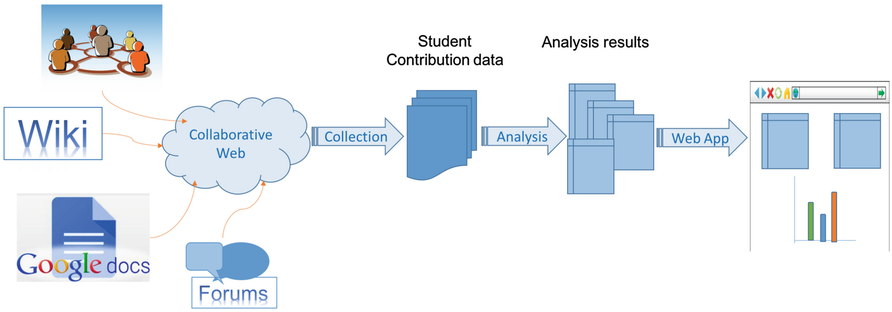



**Fig. 1.** Phases of the student contributions processing

The approach was tested on student productions in a course called "Educational games" where the students submit a variety of written analysis and contributions in discussion forums on a private social network created on Yooco[1], on Google Docs, and on EduTech Wiki[2], a MediaWiki platform. The solution is built using the R language and its "Shiny" library.

### 2.1 Phase 1: Data Collection

The data collection approach depends on the type of the platform: whether the platform offers a web API to collect information or not. For instance, MediaWiki's Web API allows to query the server for well-structured information with respect to content, changes and user actions. Other platforms of interest, such as the private social network Yooco, do not necessarily propose an API. In that case, we rely on web scraping, where R allows to sign in with the instructor credentials, read the HTML pages and extract the information. The issue with web scraping is that it is tightly tied to the HTML layout and might require continuous adaptation if the scraped pages are changed.

Google Docs files require a different approach because their content cannot be scraped, but R allows to download the documents in different formats provided that the files are accessible to anyone who has their links.

### 2.2 Phase 2: The Analysis

Initially, we produced tables that synthesize an overview of the student submissions including for instance, the dates of submission, the number of thread

[1] www.yooco.org.
[2] edutechwiki.unige.ch.

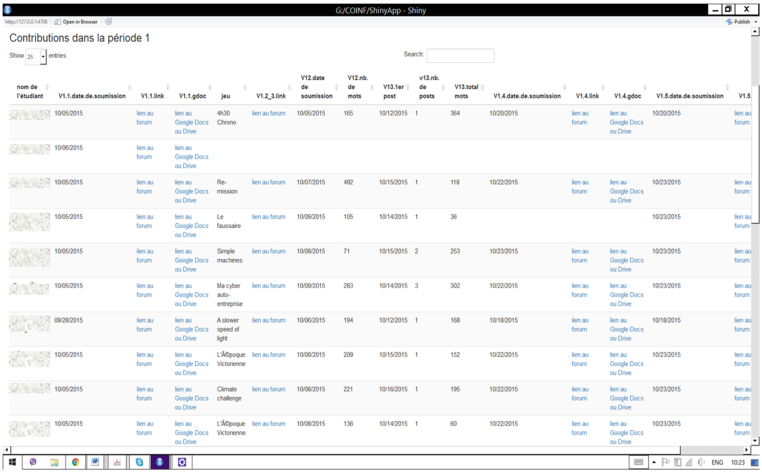



**Fig. 2.** Submissions summary of the first phase of the course "Educational Games"

interactions, an estimation of the number of words in forum posts, etc. The instructors were also interested in grouping in one place the hyper-links to the student work submitted on the different platforms. These links allow for a quick access to pieces of work without the need to navigate through Yooco discussion forums. Figure 2 shows an example of a comprehensive table summarizing all the student contributions to the first phase of the course called "Période 1".

Rich lexicographic analysis of the contributions can also be done with R. Future prototypes will add more advanced text analysis methods (e.g. [6]).

### 2.3 Phase 3: Presentation of the Results

Dynamic web pages can be created with the "Shiny" [8] library. These "Shiny apps" can be tested locally from within R, deployed on a web server. Typically, the Shiny app collects the instructor credentials, connects to Yooco, and collects the data that she/he wishes to visualize/analyze and then presents the results on the web page. In our case, we mostly relied on displaying tables synthesizing the information (see. Figure 2) with the possibility to save these tables in local files for later use. Future developments will enrich the web application with other types of visualizations.

## 3 Conclusion and Future Work

This work was motivated by a need for flexible tools to support the evaluation and assessment processes while using platforms and collaborative tools outside of the limitations and restrictions of the traditional LMS's. In addition, this approach offers instructors access to a richer set of analysis and tracking tools than what LMS platforms usually offer.

The functional prototype indicates that R might be an appropriate programming language as it allows for quick prototyping and testing. Most importantly, it has vast libraries for text analysis, text mining and statistical analysis that allow answering interesting questions regarding the students' written contributions. These aspects are not yet fully explored, we first wanted to be sure that the language offers all what an instructor might need to (semi)-automatically collect information about student work, regardless of the used collaborative platform. Nevertheless, the approach we propose here is independent of the programming language. Any other programming environment can be used provided it also offers flexibility and diversity in the analysis tools.

Our future endeavour will focus on developing the core analysis functionalities through the exploration of how to use existing text analysis methods and tools to meaningfully support instructors in the contribution assessment process. Another aspect of interest concerns the design of more compelling visualizations of the analysis results.

## References

1. Dalsgaard, C.: Social software: E-learning beyond learning management systems. Eur. J. Open Distance E-learning (2006)
2. koRpus package Manual, 8 March 2016. https://cran.r-project.org/web/packages/koRpus/koRpus.pdf
3. rvest package Manual, 11 November 2015. https://cran.r-project.org/web/packages/rvest/rvest.pdf
4. Mediawiki API Tutorial, 25 March 2016. https://www.mediawiki.org/wiki/API:Tutorial
5. Senter, R.J., Smith, E.A.: Automated readability index. Cincinnati University (1967)
6. Roy, S., Narahari, Y., Deshmukh, O.D.: A perspective on computer assisted assessment techniques for short free-text answers. In: Ras, E., et al. (eds.) CAA 2015. CCIS, vol. 571, pp. 96–109. Springer, Heidelberg (2015). doi:10.1007/978-3-319-27704-2_10
7. Seifried, E., Lenhard, W., Spinath, B.: Plagiarism detection: a comparison of teaching assistants and a software tool in identifying cheating in a psychology course. Psychol. Learn. Teach. J. **14**(3), 236–249 (2015)
8. Shiny: A web application framework for R. http://shiny.rstudio.com/. Accessed 25 Mar 2016
9. Torruellaa, J., Capsadab, R.: Lexical statistics and tipological structures: a measure of lexical richness. Procedia - Soc. Behav. Sci. **95**, 447–454 (2013)
10. TreeTagger - a part-of-speech tagger for many languages. http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/. Accessed 25 Mar 2016

# Experiments on Virtual Manipulation in Chemistry Education

Shaykhah S. Aldosari[1,2] and Davide Marocco[2,3](✉)

[1] College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia
ssaldossary@pnu.edu.sa

[2] School of Computing, Electronic and Mathematics, Plymouth University, Plymouth, UK
{shaykhah.aldosari,davide.marocco}@plymouth.ac.uk

[3] Department of Humanistic Studies, University of Naples Federico II, Naples, Italy

**Abstract.** The virtual reality technology is getting enhanced day by day as simulation and haptic together opened a new dimension to education system. It is a positive movement which has the potential to change the way of teaching from traditional style to computer and simulation enhanced teaching methods. This work discusses the achievement and testing of an educational system based on interaction and 3D visualization. This educational system is a prototype of a haptic system for chemistry experiment simulations and molecular visualization that exploits the functionalities of a gesture-based device that could be applied both in research and e-learning. The qualitative analysis of the obtained results is presented in this paper.

**Keywords:** Interaction · 3D visualisation · Technology enhanced learning · Chemistry education · e-learning · Simulation · Molecular visualization

## 1 Introduction

In recent years, the evolution of technology has influenced education in several ways. Various tools and methods based on visual interaction technologies have been introduced in order to help students acquiring knowledge based on simulation. It has been realized that the virtual vision, e.g., could solve the difficulty in understanding certain aspects of the real world and help students to fully understand the idea behind scientific rules and laws or other subjects rather than just having theoretical knowledge. Recent research, indeed, has found that the use of acoustic and visual presentation techniques works exceptionally well in educational domains [1]. However, in [1] it is shown that displays for educational purposes are primarily vision based even in the field of virtual reality. While many difficulties in understating the boundaries and connections between complex concepts could be solved by allowing people to directly manipulate objects and, up to a certain extent, concepts. Despite these considerations, the main methodology currently used in the education system is largely based on a rather passive approach of teaching, and even in most modern educational tools, such as MOOCs and online virtual labs, those tools do not allow student to actually be active in their learning as they often


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 543–546, 2016.
DOI: 10.1007/978-3-319-45153-4_59

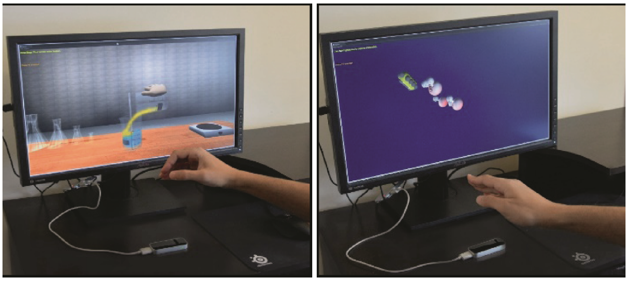


**Fig. 1.** A student using the haptic system with the leap motion controller

tend to display virtual contents without providing a concrete possibility to provide an effective learning-by-doing pedagogy [2]. In contrast, teaching practices based on technologies that allow the experience of natural interactions with the subject matters, such as haptic and gesture-based technologies, encourage active participation of students to provide a better understanding by means of direct interaction with the course material, as it exists in the real world. Therefore, we believe that the application of such technologies in education may play a substantial role, as it could be beneficial wherever an experience of realistic simulation is actually required.

## 2 Gesture-Based Educational System

In the field of education, the process of studying is traditionally based on visual and auditory cues. Only recently, scientists and technologists explored other ways and other senses through which learning processes can be developed. Currently, we are developing an educational gesture-based and manipulative system for integrated chemistry experiment simulations and molecular visualization. The system uses a 3D graphical user interface coupled with a Leap Motion controller or the Mouse, depending on user choice of input device selection. The potential of the Leap Motion controller is huge and it is expected to be used in education and simulation environments to simplify learning issues. Leap Motion technology can provide this often complicated service in an efficient convenient mode. A student using the system by the Leap Motion with its PC is shown in Fig. 1. Our system introduces two chemical experiments that can be used in teaching chemistry at an introductory level. In each experiment, a student can interact with the haptic device, the Leap Motion or the Mouse based on the user choice, and complete the experiment steps "as-if" she was in a real lab, what they called in chemistry macroscopic level [3]. At the macroscopic level, a user can use physical movements to translate, rotate and swipe glassware and chemical substances to act and perform experimental procedures as in a real lab. Then, the user can zoom into the chemical compound and see and interact with virtual representations of molecules and chemical bonds, what they called in chemistry microscopic level [3]. At the microscopic level, the system shows a graphical representation of atoms and molecules. The user can use gestures to rotate the 3D molecular models of the chemical compounds See [4] for a detailed description.

For example, in one experiment the objective is to study the molecular changes of three different types of salts when they are added to water, chemically called Salt Hydrolysis. Our system allows the user to select one of three different types of salts - ammonium chloride salt, sodium nitrate salt or potassium fluoride salt. Each one of these salts gives a different result when it is added to water. At the macroscopic level of the system, the user completes the experiment steps by using the haptic device and then observe the change in the mix appearance as in a real lab. The chemical explanation of this change can then be discovered by moving to the microscopic level which can lead to understanding this chemical reaction, as shown in Fig. 2(a) and (b).

(a) (b)

**Fig. 2.** (a) The macroscopic level of the second experiment; (b) The chemical reaction between ammonium chloride and water molecules at the microscopic level

## 3 System Testing

The purpose of the test was to assess the usability of the system's interface, information flow and architecture. The effectiveness of the learning will be addressed in the next phase of the study. Each participant filled out a survey with her opinions. The collected information about the behaviors and opinions using surveys are used to find out attitudes and reactions, to measure user's satisfaction, and to gauge opinions about the system. The testing was conducted at Princess Nora Bint Abdul-Rahman University in Riyadh, Saudi Arabia in November 2015. In this study, two different setups were used to test the system. The first test (LM_DEVICE) is applied on the system while using the Leap Motion controller as an input device to interact with the interface. The second test (MOUSE_DEVICE) has been carried out using the Mouse as an input device. 90 participants were involved in the testing to ensure stable results. 45 users participated under the LM_DEVICE setting and 45 users participated in MOUSE_DEVICE one. They fill out a brief background questionnaire and rate the system by using a 5-point Likert scale (Strongly Disagree, Disagree, Neutral, Agree and Strongly Agree) for 13 subjective measures.

As a result, all participants successfully completed all the tasks. Overall, irrespectively of the settings experienced, 91.78 % of them found the system easy to use, 90.44 % liked the interface of this system and 91.33 % would like to use this system again. In addition, 91.56 % thought the system is effective in helping them to understand the experiments' tasks and scenarios. 92.89 % of the participants in LM_DEVICE agreed that the system was easy to use, which is greater than the agreement percentage in

MOUSE_DEVICE, which was 90.67 %. Most of the participants (91.11 %) liked the interface of this system in LM_DEVICE, which is higher the agreement percentage in MOUSE_DEVICE. In addition, most of the participants in LM_DEVICE (93.33 %) think that they would like to use this system again, while only 89.33 % of the participants in MOUSE_DEVICE has the same opinion. This shows that the Leap Motion has a higher potential to engage users with these experiments.

Based on a comparative study of the result of the two different usability testing, MOUSE_DEVICE's results were a little better than LM_DEVICE's results and there is only a slight difference between them. Although participants' average agreement rating was 4.51 for the Mouse testing results, 4.41 was the participants' average agreement rating for the Leap Motion controller testing results. Even though all participants have not seen or used the Leap Motion before, more participant in LM_DEVICE agreed that the system was easy to use and the Leap Motion was easier to use more than the Mouse. On the other hand, more participant in LM_DEVICE liked the interface of this system and they would like to use this system again, which implies using the system with the Leap Motion controller is more interesting for the users and it gave them a better understanding of the experiments while using the simulation system.

## 4 Conclusion and Future Work

This paper introduces a molecule virtualization system and virtual labs along with the technology implemented within it and the results of a system testing survey conducted in Saudi Arabia. Our educational system combines three approaches: virtual environments, simulations and interaction based design. Thus, students will be able to apply lab procedures in a virtual lab environment and explore the underlined molecular structures by only downloading the system. In future works, more effort would be required to enhance and improve the interface, particularly in the attempt to bridge more closely the gesture-based approach and the haptic/manipulation aspects by exploring the robustness of the system. On the pedagogical side, more efforts would be required to improve the system test and to study how such devices may actually aid students understanding.

## References

1. Barfield, W., Danas, E.: Comments on the use of olfactory displays for virtual environments. Presence Teleoper. Virtual Environ. **5**, 109–121 (1996)
2. Waldrop, M.M.: Education online: the virtual lab. Nature **499**, 268–270 (2013)
3. Petrucci, R.H., Harwood, W.S., Herring, G.F., Madura, J.D.: General Chemistry: Principles and Modern Applications. Pearson Prentice Hall, New York (2010)
4. Aldosari, S.S., Marocco, D.: Using haptic technology for education in chemistry. In: 2015 Fifth International Conference on e-Learning (ECONF), Manama, pp. 58–64 (2015)

# A Survey Study to Gather Requirements for Designing a Mobile Service to Enhance Learning from Cultural Heritage

Alaa Alkhafaji[1](✉), Sanaz Fallahkhair[2], Mihaela Cocea[1], and Jonathan Crellin[1]

[1] School of Computing, University of Portsmouth, Buckingham Building, Portsmouth, UK
{alaa.alkhafaji,mihaela.cocea,jonathan.crellin}@port.ac.uk

[2] School of Computing, University of Brighton, Cockcroft Building, Brighton, UK
S.Fallahkhair@brighton.ac.uk

**Abstract.** This study was carried out to gather user requirements using a questionnaire survey. The study has investigated how people may use mobile location-aware technologies for learning purposes in cultural heritage contexts. This paper presents the results of this survey study and outlines a number of challenges for further development.

**Keywords:** Mobile learning · Informal learning · Location-based · Cultural heritage

## 1 Introduction

Learning occurs while people are experiencing and being engaged in different types of activities [1]. Learning from experiences is a notion that was originally developed by the theorist John Dewey in his book "Experience and Education" (Dewey 1938). Dewey's theory has served as a foundation stone for informal learning that was developed by Malcolm Knowles in 1950, in his publication "Informal Adult Education" [2].

Engaging in aspects of cultural heritage forms an important facet of the informal learning process. Since cultural heritage reflects the identity of most societies [3], it is important for people to learn more about the historical significance of heritage sites. This may help people appreciate their history, which could further promote a sense of loyalty and engagement [4]. Technologies, such as mobile learning, have already been used to support learning from cultural heritage sites, which help learning independent of time and place [5].

This study was conducted in the form of a survey, with data being gathered using a questionnaire technique to elicit user requirements for developing a mobile location-based learning service to be used at cultural heritage sites. The results of this study act as a cornerstone in designing the first version of the user requirements for developing a mobile location-based service to support informal learning at cultural heritage sites.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 547–550, 2016.
DOI: 10.1007/978-3-319-45153-4_60

## 2 The Survey Study

The survey study was conducted to elicit user requirements for developing a mobile location-based learning service with respect to cultural heritage sites.

A questionnaire technique was used to gather user requirements within the user-centered design approach. The questionnaire was designed based on the themes that emerged from a previous focus group study [6]. The "Convenience Method" of sampling was used to recruit participants [7]. The study was carried out between 17th Feb and 17th March 2015. The data was analyzed using the SPSS software [8]. A simple statistical analysis was used to obtain frequencies of the nominal data.

189 participants responded to this survey. The participants' age ranged from 18 to 70+ years old. 47 % of participants were male and 52 % were female. 47 % of participants were students, 33 % were employed and 12 % were retired. The remainders were: unemployed (4 %) or self-employed (3 %). 3 % of participants stated different occupations such as researcher, independent, and semi-retired.

## 3 The Results

This study has investigated how people my use mobile technology for learning purposes in cultural heritage contexts. The section presents the summary of the results.

The results show some features and services that people would like to use through their mobile device which include information in multiple modes: images (74 %), texts (70 %), audio (49 %) and video (47 %). Participants claimed that they would like to use different services at cultural heritage sites. The most popular services include: (1) to get directions (75 %), (2) to find the nearby cultural heritage places (65 %), (3) to find the nearest services (56 %), (4) both services, to get historical information while people walk around, and finding out extra information about the sites (53 %), (5) to pre-organized a visit (50 %). Furthermore, 62 % of participants said they would like to customize their mobile app based on their interest.

Participants were asked to choose the service that they think of as a type of learning, in order to understand how they construe learning. The results revealed that 85 % of respondents consider online courses as learning, 78 % of respondents said *accessing online services* is a type of learning, 76 % of respondents think that *accessing specific information* is a type of learning and 67 % of respondents consider *using a dictionary* is a type of learning. Interestingly, only 36 % and 31 % of respondents believe that *accessing general information* and *getting directions* respectively are types of learning.

Some challenges were highlighted by the results regarding using mobile devices at cultural heritage sites. Some participants said that they do not use mobile devices at cultural heritage sites (23 %); respondents stated several reasons for that: (1) 57 % of them claimed that the mobile device distract them during the tour, (2) 20 % of them do not use mobile devices due to a poor network quality, (3) 13 % of them reported that it is not easy to follow the instructions, (4) 11 % of respondents said that the available applications do not meet their needs. In addition, 15 % of respondents reported different reasons such as: weather limitation "*would need a waterproof tablet*".

## 4 Discussion

The questionnaire technique used in this study allowed the gathering of a wide range of data. This in turn gives a clear understanding of how people differ in the way they use mobile devices at cultural heritage sites. The findings have significant implications for the development mobile learning services to be used in cultural heritage contexts.

The results indicate factors that will be useful in designing mobile-location based leaning services. These can be summarized as: considering user profile and adapt services based on users' interests, presenting information in multiple modalities, and providing instant information based on the user location which supports situated learning.

An interesting issue that was revealed is the different perceptions of learning. People have different understanding about the meaning of 'learning' [9]. Based on the results, learning could be classified into several categories: (1) acquiring formal information such as accessing online courses, which could help to enhance an individual's professional life; (2) acquiring information that could enhance an individual's skills; (3) acquiring informal information that could be helpful to enhance an individual's personal knowledge; (4) acquiring general information that could assist in individual's daily life. Learning from experiences could include all aforementioned learning categories. Since learning interweaves with people's daily life, it could be hard for it to be distinguished as learning [10]. We can infer that learning could be happening incidentally with the learning showing little awareness that learning is taking place.

Finally, the current study has underlined some challenges regarding using mobile devices for learning purposes at cultural heritage sites. The challenges include physical aspects of the devices, such as the screen size and the network. The increasing capabilities of tablets and smartphones may reduce the importance of these factors. Furthermore, a minority find mobile devices distract from the enjoyment the visit. A possible explanation for this issue the interruption caused by the switching visual attention between the device and the exhibit. Using Smart Glasses could help with this by overlaying data on a user's visual field. In addition, an interesting issue has emerged through the study such as the weather, as some people reported that the need for a waterproof device given how frequently it rains in the UK. Possibly smart glasses might help with this too. Finally, an issue regarding the quality of network, not considered in this research, but could be an important issue for further research.

## 5 Conclusion and Further Work

A summary of a survey study has been presented in this paper, which was carried out as a part of a series of studies designed to gather user requirements. A questionnaire technique was used in this study. This study forms a stage of a research project, which is intending to develop a mobile location-based learning service with respect to cultural heritage contexts. There are a number of areas that we envision to carry out further work: First, to conduct further steps to fully eliciting of users requirements, by conducting interviews with end-users and museum staff, to gain in-depth details regarding using mobile devices for learning purposes. Second, to design a task model based on the results of the combination of this study and the interview study. Third, to develop a prototype mobile as a proof of concept based on the task model. Next, usability evaluation will be conducted. Finally, a list of guidelines will be for future mobile application development in this domain.

## References

1. Schunk, D.H.: Learning Theories. Printice Hall Inc., New Jersey (1996)
2. Smith, M.K.: Malcolm Knowles, informal adult education, self-direction and andragogy (2002)
3. González, M.V.: Intangible heritage tourism and identity. Tourism Manage. **29**, 807–810 (2008)
4. UNESCO: Managing Cultural Wold Heritage: The United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, France (2013)
5. Sharples, M.: The design of personal mobile technologies for lifelong learning. Comput. Educ. **34**, 177–193 (2000)
6. Alkhafaji, A., Fallahkhair, S., Cocea, M.: Towards gathering initial requirements of developing a mobile service to support informal learning at cultural heritage sites. In: Cognition And Exploratory Learning In The Digital Age (CELDA 2015), p. 51 (2015)
7. Barnett, V.: Sample Survey Principles and Methods. Edward Arnold, London (1991)
8. Greasley, P.: Quantitative Data Analysis Using SPSS: An Introduction for Health & Social Science. McGraw-Hill Education, London (1991)
9. Schmeck, R.R.: Learning Strategies and Learning Styles. Springer, New York (1988)
10. Vavoula, G.: KLeOS: a knowledge and learning organisation system in support of lifelong learning, Ph.D. thessis, University of Birmingham, UK (2003)

# Inspiring the Instructional Design Process Through Online Experience Sharing

Grégory Bourguin, Bénédicte Talon(✉), Insaf Kerkeni, and Arnaud Lewandowski

University of Lille Nord de France, ULCO, LISIC, Calais, France
{bourguin,lewandowski}@lisic.univ-littoral.fr,
{Benedicte.Talon,Insaf.Kerkeni}@univ-littoral.fr

**Abstract.** A lot of pedagogical resources are available thorough the Web. Paradoxically, it is hard for instructional designers to discover and decide which ones will best fulfill their needs. End-users' experience has been identified as a major source of information in the resource selection process. However, no solution totally fulfills the needs and end-users' experience can hardly be browsed while being dispersed over the web. Our research prototype called EVOXEL can help instructional designers by completing current web solutions. Built upon ontological mechanisms, EVOXEL provides teachers a mean to share experience they have developed during their instructional activities. This experience is crystallized in the educational resources assemblages they have built. This experience can then be browsed, inviting others to be inspired from it.

**Keywords:** Instructional design · Experience sharing · Open educational resources · Ontology

## 1 Introduction

Open resources platforms and the web in general, offer an increasing set of online or downloadable pedagogical resources. Re-using these resources facilitates the instructional design process but the infrastructures should provide better means to find adequate educational resources [1]. An important issue is how to help pedagogical designers finding and selecting the most appropriate resources. Studies show that resources have different meanings according to context and users [2]. Basic domain categorization does thus not necessarily help users in discovering a resource and understanding what it can serve for. Categorization of the resources around sets of generic tasks augments description but a user doesn't know if the generic task will match their own in the instructional activity. To overcome these issues, end-users' experience through social tagging [3] and experience sharing systems [4, 5] has been identified as a major source of information.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 551–554, 2016.
DOI: 10.1007/978-3-319-45153-4_61

## 2 End-Users' Experience

### 2.1 Rating, Commenting and Tagging Systems

Rating system provides information about quality of a shared resource, but a rank cannot be fully understood without context. Comments sections offer another form of information but it is mostly implicit, unstructured and hard to extract. Tagging systems and folksonomies are integrated in educational resources repositories. They serve for indexing shared resources [2], and can be used as a key concept for collaborative learning [6]. However, folksonomies cannot be considered as directly matching a particular user's culture [6] and are reflections of dominant cultural groups only. Researchers [4] have proposed to augment tagging systems with ontological features in order to keep track of the link between a tag and its creator, and authors [2] have shown that investigating one's universe facilitates resource appropriation and helps the inspiration process. This is what EVOXEL has been designed for.

### 2.2 Sharing Resources and Context of Use

Weblogs, forums, wikis or video channels are widely used by end-users to let others discover their universe. They offer experience sharing about resources assemblages and facilitate discovering and understanding of someone's context. Designers look for inspiration by browsing teachers or institutions' spaces where they share their experiences about resources assemblages. For example, in MERLOT [7], users create and share *ePortfolios* describing resources assembled to perform a pedagogical task. But the existing system doesn't provide structure and semantics allowing deeply searches.

## 3 EVOXEL Prototype

### 3.1 Online Personal Ontologies

EVOXEL is an online tool based inspired on by Activity Theory [8] through its description of the mechanism of experience crystallisation inside activity's mediating artefacts. Its ontological meta-model has been successfully used in various application domains [9]. Users feed their personal ontology by describing and tagging resources and activities. Each element can refer to web resources associated with it. Users can insert links towards ontological elements into existing web solutions (repositories, weblogs, etc.). Links serve as *entry points* from where users can browse others' pedagogical universe.

Elements of the personal ontologies can be freely tagged and take benefits from the expressiveness and power of ontological tools. Tags are not just keywords, but can be commented, structured in hierarchies and use inheritance. The reasoner infers and reveals relations between the elements, and advanced search mechanisms can be set up.

### 3.2 EVOXEL Architecture

The server side is realized in JEE and applies the OWL API for creating, managing and querying personal ontologies. Angular JS is used to provide a client side single page application. The environment dynamically provides a fixed URL for each element (activities, resources and tags) of a universe. An URL opens on a synthesis of all information concerning the pointed element. The JFact ontological reasoner plugged in the JEE server infers all the links that can be deduced from an element. An ontological search tool completes the browsing system.

### 3.3 Activity Modeling - Sharing Pedagogical Experience in EVOXEL

A teacher involved in learning design was asked to use EVOXEL to describe her teaching universe. Her resulting personal ontology is called *MyTeachingUniverse*. It describes resources, activities and a set of tags corresponding to her pedagogical work (Fig. 1).

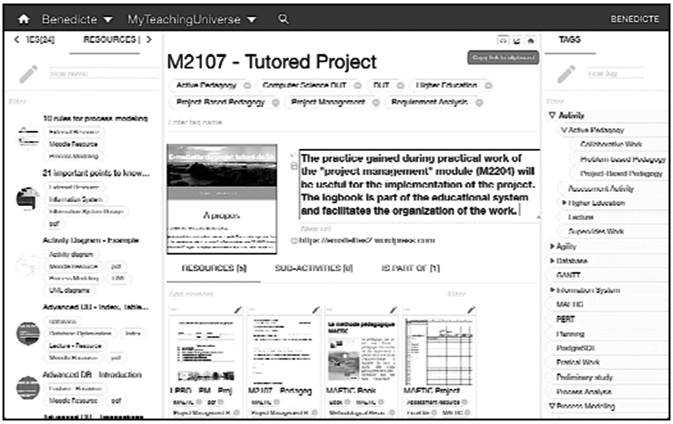



**Fig. 1.** The EVOXEL Modeler

### 3.4 Browsing Shared Experience

Once the URL(s) inserted in a web media, other users can discover *MyTeachingUniverse* and be inspired from resources assemblages that crystallize her pedagogical experience. For example, an URL corresponding to the *MAETIC Book* she used in the *Tutored Project* was inserted in her Weblog. Following this link, one discovers (Fig. 2(a)) this resource (description, two links towards the book) and its surrounding elements in the teacher's universe inferred by EVOXEL (tags, all activities and resources). A search panel completes this browsing mechanism and takes benefits from the semantics carved into personal ontologies. In Fig. 2(b), a user looking for all *Lectures* using *SCRUM* in an *Active Pedagogy* process enters elements ***tagged*** with *Lecture*, ***including*** elements ***tagged*** with *SCRUM*, and that are ***part of*** elements ***tagged*** with *Active Pedagogy*.

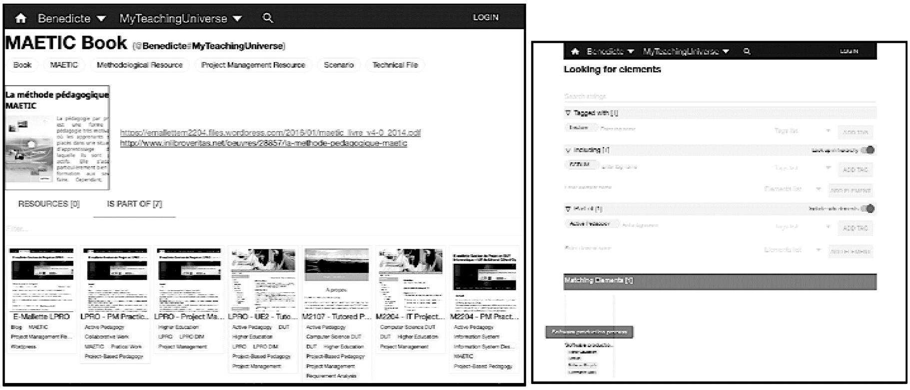



**Fig. 2.** (a) The *MAETIC Book* resource in EVOXEL (b) A complex search

### 3.5 Perspectives

EVOXEL still need to be developed. We are currently testing new EVOXEL features like importation of elements from other one's universe. We also are organizing a deep evaluation of the current features with pedagogical designers and students.

## References


1. Hylén, J.: Open educational resources: opportunities and challenges. In: Proceedings of Open Education, pp. 49–63 (2006)
2. Draxler, S., Jung, A., Stevens, G.: Managing software portfolios: a comparative study. In: Piccinno, A. (ed.) IS-EUD 2011. LNCS, vol. 6654, pp. 337–342. Springer, Heidelberg (2011)
3. Conole, G., Culver, J.: The design of cloudworks: applying social networking practice to foster the exchange of learning and teaching ideas and designs. Comput. Educ. **54**, 679–692 (2010)
4. Kim, H.L., Scerri, S., Breslin, J.G., Decker, S., Kim, H.G.: The state of the art in tag ontologies: a semantic model for tagging and folksonomies. In: International Conference on Dublin Core and Metadata Applications, North America (2008)
5. Knerr, T.: Tagging Ontology – Towards a Common Ontology for Folksonomies (2008). http://tagont.googlecode.com/files/TagOntPaper.pdf
6. Lavoué, E.: Social tagging to enhance collaborative learning. In: Proceedings of the 10th International Conference on Advances in Web-Based Learning, 08–10 December, Hong Kong, China (2011)
7. Moncada, S.M.: Rediscovering MERLOT: a resource sharing cooperative for accounting education. J. Higher Educ. Theory Pract. **15**(6), 85–95 (2015)
8. Kuutti, K.: Notes on systems supporting "Organisational context" – an activity theory viewpoint, COMIC European project, D1.1, pp. 101–117 (1993)
9. Bourguin, G., Lewandowsky, A.: Using online personal ontologies to share experience about web resources. In: Proceedings of Information Systems 2015, Madeira, pp. 177–184 (2015)

# An Approach to the TEL Teaching of Non-technical Skills from the Perspective of an Ill-Defined Problem

Yannick Bourrier[1(✉)], Francis Jambon[2], Catherine Garbay[2], and Vanda Luengo[1]

[1] UPMC - LIP6, Paris, France
{yannick.bourrier,vanda.luengo}@lip6.fr
[2] UGA - LIG, Grenoble, France
{francis.jambon,catherine.garbay}@imag.fr

**Abstract.** In this paper we take a look at the difficulties raised by the teaching of the technical and non-technical skills mobilized during a critical situation, in the context of TEL within virtual environments. We present the advantages of using a combined enactive and situated learning approach to this problematic, and take an ill-defined perspective to raise important designing issues in this respect. We show that some aspects of this problem have not been encompassed yet in the ill-defined domains literature, and should be further studied in any attempt at teaching behaviours inducing technical and non-technical skills in a virtual world.

**Keywords:** Ill-defined domains · Non-technical skills · Critical situations · Virtual reality environment

## 1 Modelling the Interaction Between Learner and VE

In most domains involving expert knowledge, there is a number of cognitive and social factors influencing human performance, which are commonly described as Non-Technical Skills (NTS), and whose impact is most important on perceptual-gestural activities performed under critical situations. In this paper, we showcase the challenges raised by the learning of NTS inside a Virtual Environment (VE), and discuss the potential and limitations of a number of approaches recently used in this domain, in the light of an ill-defined perspective, under the scope of their application to the domains of driving and medical surgery. Based on this analysis, we point out ill-defined dimensions of our domain and discuss some of the corresponding designing issues. NTS can be defined as the "cognitive, social, and personal resource skills that complement technical skills, and contribute to safe and efficient task performance" [1]. They have an influence on a worker's technical skills and include situation awareness, decision-making, leadership, stress and fatigue management. The strong links between NTS and critical situations

This research was supported by the MacCoy Critical project (ANR-14-CE24-0021).


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 555–558, 2016.
DOI: 10.1007/978-3-319-45153-4_62

underlie the necessity to put the interaction between learner and VE at the centre of our approach. Two approaches may be used in this respect.

The first approach is based on the principles of enaction. In virtual reality, an enactive system is a system constructing a world, while being constructed by it [2]. In this view, the coupling between the VE and the user's perceptual-gestural activity is central; an individual's actions will result in a modification of the virtual world by the system, and reciprocally. Knowledge becomes the result of this interaction between individual and virtual world, and can be found in the perceptions and actions that this interaction creates. In this approach, knowledge is purely empirical, and the focus is put on what is directly experienced by the learner. The main benefit when it comes to the teaching of NTS can be seen in this phenomenological focus to learning, which becomes highly specific to an individual. However, while the benefits in terms of learner modelling are important, the fact that no specific skill is targeted may result in a loss of efficiency when it comes to the choosing of a new learning situation. Being able to assess which NTS should be improved could greatly improve the training effectiveness and therefore some modelling of the learner's knowledge is in order.

Another interaction-centric approach is Situated Learning (SL). Applied to VEs, such an approach often comes with an important background work in order to understand the knowledge underpinnings of a domain. A task is evaluated in the specific knowledges involved in each of the learner's actions or strategies [3]. Knowledge being de facto represented for a learner, targeting specific elements of the domain become possible. For example, in TELEOS [3], an ITS for the learning of orthopaedic surgery built within a SL paradigm, the feedback type will change for whether it is an empirical or a declarative aspect which is targeted. The benefits of this approach are seen in this understanding of which type of knowledge is used by the learner, allowing a system to target the skills which need the most to be improved. Drawbacks come from the lack of a clear pedagogic strategy and a deficit of efficiency when it comes to integrating the training in a learning curve. In this approach, the interaction between learner and system is important but may sometimes be overshadowed by knowledge of the domain itself. Given the unique links between NTS and critical situations, we argue that they should be best taught by experiencing a large number of critical situations. While domain knowledge is key to know which skill should be targeted, we hypothesize that the teaching of NTS, in the context of a perceptual-gestural activity, should be done by a succession of empirical experiences, and not through post simulation feedback.

Because NTS are non-procedural by nature and appear precisely to cope with the lack of an adapted procedure to deal with a situation [4], our approach should be underpinned in the principles of enaction applied to VE. However, some critical situations may simply be too hard or not critical at all for a learner, given a certain degree of technical and non-technical expertise. We therefore orient ourselves in the direction of an enactive VE including a degree of SL inspired learner's knowledge modelling. The evaluation of a learner's skill level when confronted to a given problem has been explored to lengths in the ITS literature, let us take a look at the difficulties such a knowledge modelling problems poses in an interaction-centred VE such as ours.

## 2 Teaching Non-technical Skills Inside of a Virtual World: Issues from an Ill-Defined Perspective

In 2006, Lynch et al. [5] attempted to cover the different aspects making a teaching problem ill-defined in an ITS. In this section, we take a look at the challenges raised by the modelling of NTS inside of a virtual world, following [5]'s definition of what characterizes an ill-defined problem. We also point out some further aspects that have not been encompassed yet in the ill-defined problems literature. When it comes to modelling learner's knowledge, a central criteria characterizing ill-defined domains is the absence of a complete formal theory of the domain. Here we aim at evaluating two different domains, which are technical and non-technical skills. In themselves, both have a degree of formal theory. We argue that our domain is still ill-defined knowledge-wise from a perspective that has not been covered by [5]'s definition, because technical and non-technical skills are involved in a single perceptual-gestural activity and can only be observed together. While separately well-defined, together they become ill-defined as the ties between them are diffuse and can change from one individual to another. We ask that this is a new form of problem not yet identified by previous approaches to ill-defined domains, and which should be encountered whenever evaluating a perceptual-gestural activity expressing multiple skills being used.

The challenges raised by adopting an enactive approach to NTS learning in VE can also be considered as ill-defined, for two reasons. Firstly, the sub-problems overlap, as any of the learner's actions on the virtual world will result in a change of the situation, which will either increase or decrease the importance of further actions. Secondly, the task structure is ill-defined, and more accurately, it becomes analytical since the number of possible correct paths changes as a result of this pseudo-real-time coupling between learner and world. Rather than a definite task structure, the issue is to model the singular experience of a learner trying to maintain his or her TS in front of a critical situation. The role of the ITS is then to drive the learner in a personalized "journey through criticality", assessing the coverage of a number of critical situations, and the involvement of a number of NTS. Determining the position of a problem in a continuum of solution spaces, as proposed by [6], can provide insights as to which technique should be best used in order to teach a problem. Because of sub-problems overlapping and the task being analytical, we ask that there must be a quantifiable number of appropriate solution strategies in response to a given critical situation, but an indefinite number of ways to apply these strategies. Similar challenges with regards to performance evaluation have been treated by the use of hybrid approaches including systems and model tracing for the more defined aspects of the problem-solving task, and datamining approaches in order to learn the more uncertain parts [7]. These approaches however focus solely on the evaluation of a learner's performance. In our case, because the domain in itself also has ill-defined specificities, the learner's performance will need to be looked at from the scope of his knowledge state and the situation characteristics, to determine the actual influence NTS had in such a performance. This influence may hold with different degrees: intuitively, the effect of situation awareness or stress management on the learner's performance may appear very different. The situation characteristics may also result in varying degrees of criticality impacting the learner's performance.

## 3 Discussion

We have shown the challenges raised by the teaching of NTS for perceptual-gestural activities performed during critical situations inside of a virtual world and showcased why, given the characteristics of such skills, it is necessary to adopt an interaction-centred approach coupled with a modelling of knowledge, in order to maximise efficiency and to explore as many dimensions of criticality as possible. We have highlighted the reasons why this combined approach is an ill-defined problem, both from the point of view of the interaction between learner and virtual world, and the point of view of knowledge modelling. Some aspects of our problems were already partially explored in the ITS literature. TELEOS [2] deconstructed a technical activity as a coupling of different types of knowledge in order to target the best feedback. CANADARMTutor [7] used a hybrid approach including educational data mining techniques to learn a number of correct behaviours for the usage of a robotic arm. Both of these works shared some characteristics of ill-defined domains similar with ours, yet [7] focused on the perception of a technical performance, while [3] aimed at proposing the most appropriate knowledge-type based feedback. The learning of NTS in critical situations inside of a virtual world will need to enfold both of these ITS' characteristics while considering a new aspect of an ill-defined problem, which is taking into account the merging barriers between technical and non-technical expertise.

## References

1. Flin, R.H., O'Connor, P., Crichton, M.: Safety at the Sharp End: A Guide to Non-technical Skills. Ashgate Publishing Ltd., Aldershot (2008)
2. Varela, F., Thompson, E., Rosch, E.: L'inscription corporelle de l'esprit in Paris, Seuil (1993)
3. Luengo, V.: Take into account knowledge constraints for TEL environments design in medical education. In: 8th IEEE International Conference on Advanced Learning Technologies, ICALT 2008, Santander, Cantabria, Spain, pp. 839–841 (2008)
4. Marchand, A.-L.: Les retours d'expériences dans la gestion de situations critiques, pp. 100–113 (2011)
5. Lynch, C., Ashley, K., Aleven, V., Pinkwart, N.: Defining ill-defined domains: a literature survey. In: Proceedings of Intelligent Tutoring Systems for Ill-Defined Domains Workshop, ITS 2006, pp. 1–10 (2006)
6. Le, N.T., Loll, F., Pinkwart, N.: Operationalizing the continuum between well-defined and ill-defined problems for educational technology. IEEE Trans. Learn. Technol. **6**(3), 258–270 (2013)
7. Fournier-Viger, P., Nkambou, R., Nguifo, E.M.: A knowledge discovery framework for learning task models from user interactions in intelligent tutoring systems. In: Gelbukh, A., Morales, E.F. (eds.) MICAI 2008. LNCS (LNAI), vol. 5317, pp. 765–778. Springer, Heidelberg (2008)

# Towards a Context-Based Approach Assisting Learning Scenarios Reuse

Mariem Chaabouni[1,2(✉)], Mona Laroussi[1(✉)], Claudine Piau-Toffolon[2], Christophe Choquet[2], and Henda Ben Ghezala[1]

[1] RIADI, Manouba University, Manouba, Tunisia
Mariem.Chaabouni@univ-lemans.fr, Mona.Laroussi@univ-lille1.fr, Henda.Benghezala@ensi.rnu.t

[2] LIUM, Maine University, Orono, France
{Claudine.Piau-Toffolon,Christophe.Choquet}@univ-lemans.fr

## 1 Introduction

Nowadays, learning design has become one of the principle research topics in the domain of Technology-Enhanced Learning (TEL). In fact, it is important to organize and capitalize the teacher's practices essentially with the emergence of various teaching modalities and the high integration of the technology in the learning processes. This will allow the teachers to reuse their own practices and share them with others (teachers or students). Certain factors may affect the reuse of shared scenarios such as the use of heterogeneous formalisms and representation approaches of scenarios, the definition of rigid scenarios, the high variation of learning contexts from one situation to another one and the high variability of areas/resources used in the scenarios.

A first axis of works has proposed methods and techniques to help users (teachers or students) to identify and select adapted learning objects or scenarios for reuse. We can cite the LOM standard [1] specifying a set of learning object metadata or works using semantic web and ontologies [2] to the learning scenarios. There is also the SCORM standard [3] to promote reusability and interoperability of learning content across platforms. A second axis has been oriented to the proposal of methods and techniques to design adaptable and customizable objects and scenarios in order to promote their reusability. In this axis, some works promote the use of design patterns to assist learning designers in the expression of adaptable scenarios depending on the context. For example, the COLLAGE project [4] proposed patterns of activities of collaborative learning (CLFPs: Collaborative Learning Flow Patterns). These patterns are reusable and customizable good practices used by practitioners according to particular learning situation specifications.

This poster is subscribed in the first axis and mainly focuses on enhancing reuse by treating essentially the context aspect related to the learning scenario. It proposes a method to model this context in a multi-layered approach. An author tool based on this modeling approach is also introduced, assisting the scenario design by suggesting the most appropriated scenarios to a given learning situation. The adopted approach is reinforced by the observation of the past learning experiences to support the pertinence of reuse.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 559–562, 2016.
DOI: 10.1007/978-3-319-45153-4_63

In order to represent the learning experiences and to have a good perception of the scenario in a real learning situation, strategies and techniques of scenarios observation have been implemented in existing works [5, 6]. These works mainly use the concept of pedagogical indicators. Such an indicator is considered *"as a significant variable able to help in understanding the effective activities performed during a learning session"* [5]. In the present work, it is proposed to use the observation by the indicators for indexing purposes. The resulting indexes aim to support the reuse of scenarios. Thus, the observation allows the teacher-designer to assess the progress of the scenario and to determine whether it was successful in a specific context and if it has been effectively adapted to this context.

## 2 Assisted Construction of a Contextual Learning Scenario Index

To assist the construction of contextual learning scenarios, we define an approach of representation of the learning scenario context. For that, we adopt a multi-layered modeling approach. In fact, each work in literature, as it was defined in the previous section, models the diversity of context on their specific scopes and their particular representation. The learning scenario context is rich and open. It is impossible to obtain a stable and complete modeling. We need general context modeling approach without restricting properties. Therefore, we opt for a meta-modeling approach. As illustrated in Fig. 1, we choose the four-layered architecture defined by the Object Management Group (OMG) that separates the different conceptual levels for defining a model. We propose the 4 Layers known as M0, M1, M2, and M3 as following:

– M3: represents the MOF meta-meta-model [7], useful to the description of meta-models proposed by the OMG;

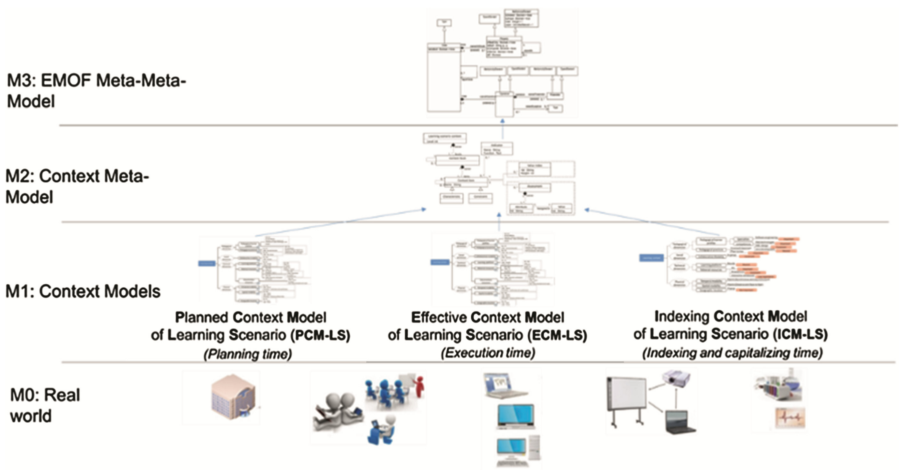



**Fig. 1.** Learning scenario context layers

- M2: represents our proposed meta-model of the learning scenario context, compliant to MOF meta-meta-model;
- M1: represents the context models related to a particular learning situation. Different context representation levels (ECM-LS, ICM-LS and PCM-LS) are compliant to the same context meta-model of M2;
- M0: represents the real world consisting here in the context elements in a learning situation.

## 3 An Authoring Tool to Reuse Adapted Learning Scenarios

The identified levels of the scenario context have been formalized and integrated into an authoring tool "Capture-tool" designed for teachers-designers. This tool integrates a recommender system for adapted learning scenarios to a specific planed context. Firstly, the tool helps designers in retrieving the most relevant scenarios to planned learning situation, and so to enhance the learning scenario reuse. The teacher informs its planned context in which the scenario will be implemented (see *"Inform my context"* part of Fig. 2).

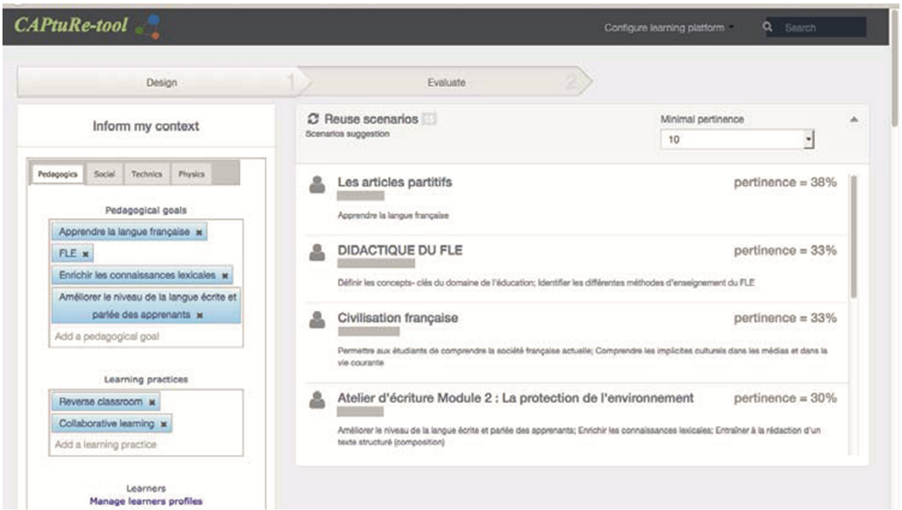



**Fig. 2.** The Capture-tool

The tool implements a context-based similarity algorithm detailed in a previous work [8]. This algorithm calculates the similarities between a planned context in a learning situation and contextual indexes associated with capitalized learning scenarios. The tool identifies and suggests the capitalized scenarios having been indexed by the most similar contexts ICM-LS to the planned context (see *"Reuse scenarios"* part of Fig. 2). Parallel to the design, the teacher plans the observation through specifying the indicators to be calculated during the execution of the scenario.

The tool also offers other interfaces allowing the teacher-designer to analyze the context and the scenario progress through calculated indicators, and then index the scenario for future reuse.

## 4 Results and Conclusion

In order to experiment the developed tool implementing the similarity algorithm, we start firstly by analyzing the existing BASAR scenarios (currently 106 scenarios [9]) and then the extraction of the related contexts. We complement these contextual data through the establishment of a survey addressed to BASAR designers. This survey has allowed us to collect the contexts information in which the scenarios was effectively executed, and that we have not been able to extract from the capitalized scenarios. Therefore, we construct, from the collected data, the associated contextual indexes of the existing BASAR scenarios. Then, we proceed to different simulations of the algorithm of the Capture-tool. These simulations have been conducted with teachers placed in a situation of design by reuse.

## References

1. LOM specification: Learning Object Metadata. http://ltsc.ieee.org/wg12/index.html. Accessed 12 Jan 2015
2. Paquette, G.: An ontology and a software framework for competency modeling and management. Educ. Tech. Soc. **10**(3), 1–21 (2007)
3. ADL Technical Team. Sharable Content Object Reference Model - SCORM. Documentation 1st edn., Advanced Distributed Learning (ADL) (2004)
4. Hernández-Leo, D., Villasclaras-Fernández, E.D., Asensio-Pérez, J.I., Dimitriadis, Y., Jorrín-Abellán, I.M., Ruiz-Requies, I., Rubia-Avi, B.: COLLAGE: a collaborative Learning Design editor based on patterns. J. Educ. Tech. Soc. **9**(1), 58–71 (2006)
5. Ngoc, D.P.T., Iksal, S., Choquet, C., Klinger, E.: UTL-CL: a declarative calculation language proposal for a learning tracks analysis process. In: 2009 Ninth IEEE International Conference on Advanced Learning Technologies, ICALT 2009, pp. 681–685. IEEE, July 2009
6. Dimitrakopoulou, A.: State of the art on interaction and collaboration analysis. Information society technology, Network of Excellence Kaleidoscope, (contract NoE IST-507838), project ICALTS: Interaction and Collaboration Analysis, 2004 (2004)
7. MOF: Meta Object Facility Core Specification, OMG Available Specification, version 2.0. Object Management Group (2006)
8. Chaabouni, M., Laroussi, M., Piau-Toffolon, C., Choquet, C., Ben Ghezala, H.: A context-based similarity algorithm for enhancing learning scenarios reuse. In: 13th International Conference on Intelligent Tutoring Systems, ITS 2016, June 2016
9. BASAR project: A Database of French blended-learning scenarios. http://www.projetbasar.net/index.php/fr/. (Consulted on January 2016)
10. Chaabouni, M., Piau-Toffolon, C., Laroussi, M., Choquet, C., Ben Ghezala, H.: Indexing learning scenarios by the most adapted contexts: an approach based on the observation of scenario progress in session. In: 15th International Conference on Advanced Learning Technologies, pp. 39–43. IEEE, July 2015

# Revealing Behaviour Pattern Differences in Collaborative Problem Solving

Mutlu Cukurova(✉), Katerina Avramides, Rose Luckin, and Manolis Mavrikis

UCL Knowledge Lab, University College London, London, UK
{m.cukurova,k.avramides,r.luckin, m.mavrikis}@ucl.ac.uk

**Abstract.** The identification of effective Collaborative Problem Solving (CPS) strategies for practice based learning would make an important contribution to a better understanding of how to support the CPS process and how to design effective interventions. In this paper, we present a method for identifying effective CPS strategies using learner behaviours as the key to data to unpack this complex learning process. In order to distinguish learner behaviour patterns, we deployed an analysis framework for CPS that identifies fine-grained actions in practice-based learning activities. Then, using cumulative time plots we compared expert (those who have more experience in working together) behaviours with novice behaviours. Results show that participants with different levels of expertise in working together, present different behaviour patterns in collaborative problem solving.

**Keywords:** Collaborative problem-solving process · Practice-based learning · Analysis frameworks · Cumulative time plots

## 1 Introduction

Collaborative problem solving is an important process that triggers specific cognitive mechanisms, such as argumentation, debating and the building of shared understanding, which in turn increases the likelihood that learning may occur [1]. In STEM education, collaborative problem solving is often promoted within the context of practice-based learning activities. However, although technology enhanced learning researchers have provided us with plenty of research on the processes of collaborative problem solving; studies focusing on practice-based learning contexts are scarce. This type of learning in STEM education includes a broad range of activities. In the research study reported here, we focus on open-ended, hands-on, physical computing design tasks. Using a commercial physical computing kit[1], participants first go through a few introductory tasks such as blinking an LED on/off with a timer, blinking an LED on/off with a button, using a potentiometer to control an LED, in order to be familiarized with the kit. Then participants follow an open-ended investigation. For instance, using the materials provided they are asked to build an airplane which flies through different

[1] https://www.samlabs.com/


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 563–569, 2016.
DOI: 10.1007/978-3-319-45153-4_64

cities of the world (which requires participants to measure angles using a protractor), or they are asked to build a small system that changes the motor speed depending on the amount of light it receives (similar to the fact that the amount of $O_2$ produced in a green plant through photosynthesis changes depending on the amount of sunshine the plant receives). Participants are provided with different sensors and actuators to control events in their open-ended investigations. This type of practice-based activities are increasingly used in schools, particularly since the ‘maker movement’ emerged [2] (Fig. 1).

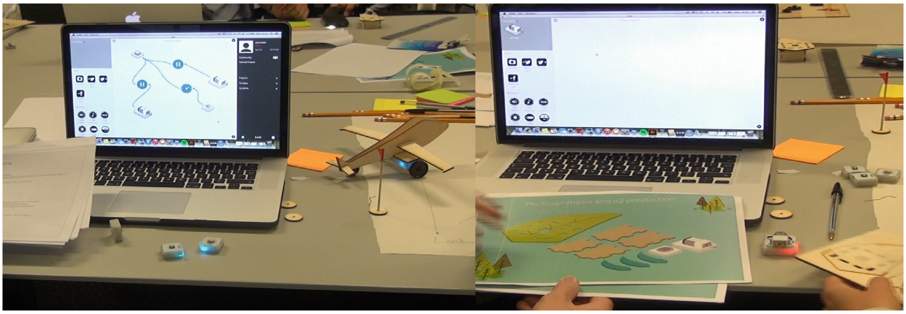


**Fig. 1.** Pictures from the practice-based learning activities

The nature of open-ended learning activities requires appropriate guidance, and this need is more significant for novices. Research shows that allowing novice students to work independently on open-ended practice-based activities without appropriate guidance does not lead to meaningful learning outcomes e.g. [3]. However, it is not easy for teachers to provide the appropriate support for students, since they are rarely aware of the learning processes followed by students [4]. In such complex learning environments as practice- based learning, it is even more challenging to support effective strategies that lead to better outcomes, both in terms of the objective to be achieved and the quality of the CPS processes. We argue that the differences in learners’ CPS processes can be revealed through the investigation of behaviour patterns that occur during the learning activities. With the capability of monitoring differences in behaviour patterns, teachers can adopt appropriate intervention techniques and then positively influence and support CPS process. The power of monitoring differences between groups of students as well as individuals working in a group would allow teachers to identify when and how to intervene in order to facilitate the accomplishment of a higher quality product and/or more satisfying learning experiences. Hence, the aim of this paper is to suggest an appropriate method to reveal the behaviour pattern differences in learners’ CPS processes. In this paper, we first briefly describe our approach to developing an analysis framework for the systematic investigation of students’ collaborative learning processes in the context of practice-based learning activities. Then, using this framework we compare novice and experts’ behaviour patterns. In this paper, we define expertise with respect to experience of working in a group. Hence, in our comparison expert participants are those who have significantly more experience in working together in a group compared to their partners with whom they collaborate. We finalize the paper with a discussion of the behaviour pattern differences in experts and novices CPS processes.

## 2 Analysis Framework

We adopted a mixed-methods approach to develop our analysis framework for CPS processes. We believe that this approach has the potential to generate frameworks that are both theory-driven and therefore broad enough to observe learning processes on the basis of theoretical assumptions; and data-driven, and therefore grounded enough to be applicable to real-life learning contexts. The main value of our framework is that it defines observable actions rather than broad definitions. Such broad definitions are hard to identify, track and interpret in data analyses. In this paper, considering the space limitation we do not get into the details of the development process of the analysis framework. Please see [5] for detailed discussion of the topic.

## 3 Application of the Analysis Framework

Our dataset consists of video and audio recordings from a workshop event in which two pairs of participants (one novice and one expert) worked on two different open-ended, hands-on physical computing projects. Both of the practice-based learning tasks were specifically designed to be accessible to all participants, regardless of their level of STEM subject-specific knowledge. However, participants differed in their expertise relevant to working together in a group. Expert participants were those who have more experience in collaborative projects. This separation was done with self-declared information. Two researchers coded the data using a multi-step qualitative methodology, taking into account the procedures and techniques developed in the qualitative content analysis method. First, two researchers used the analysis framework (Table 1) and the same data set to code students' actions with the ELAN annotation software. Any disagreements between researchers were resolved through discussion and revised the coding was accordingly. Then, the amount of time spent on each coded action within 10 s intervals logged in an excel document, which was used to generate cumulative time plots.

### 3.1 Presentation of the CPS Processes with Cumulative Time Plots

A cumulative time plot (CTP) is essentially an *x*-*y* line plot where each code is represented as a separate curve. In our case, there were in total eighteen different codes stemming from the three competencies of the CPS process relating to collaboration and six competencies relating to problem solving aspect. The *x*-axis represents the total time the learner spent on the activity, and the *y*-axis represents the duration of time spent on a code. Each curve in a CTP therefore represents the amount of time a learner spent on actions described by the code specified to date. With the purpose of comparing

**Table 1.** Final analysis framework

| | (1) Establishing and maintaining shared understanding | (2) Taking appropriate action to solve the problem | (3) Establishing and maintaining team organization |
|---|---|---|---|
| (A) Identifying facts | (A1) Vocalizing knowledge; Confirming shared understanding; Communicating regarding an answer to a question; Asking questions to verify a suggested solution; presenting skills | (A2) Identifying a problem (a situation which stops/hampers students from the natural progression of the practice-based activity) | (A3) Confirming the actions to be taken, engaging with rules |
| (B) Representing and formulating | (B1) Sharing the identified problem with other teammates; Explaining an hypothesis/suggestion in detail | (B2) Communicating about actions to take | (B3) Assigning roles to team mates; Giving responsibilities to team |
| (C) Generating hypotheses | (C1) Critically analyzing a problem; Critically analyzing a suggestion | (C2) Suggesting a solution to a problem; Hypothesizing about a problem | (C3) Suggesting an improved version of an hypothesis |
| (D) Planning and executing | (D1) Negotiating on actions to take; Approving a suggested solution | (D2) Taking actions to progress | (D3) Prompting other team members to perform their tasks; Taking actions regarding suggestions |
| (E) Identifying knowledge and skill deficiencies | (E1) Identifying individual deficiencies | (E2) Making knowledge or skill deficiency explicit | (E3) Identifying a team mistake |
| (F) Monitoring, reflecting and applying | (F1) Verifying what each other knows; Asking questions regarding the actions being taken; Observing an agreed action being taken or a teammate solving a problem | (F2) Testing a solution to check its validity; Reflecting on previous actions; Correcting simple mistakes of others | (F3) Warning teammates regarding a possible mistake |

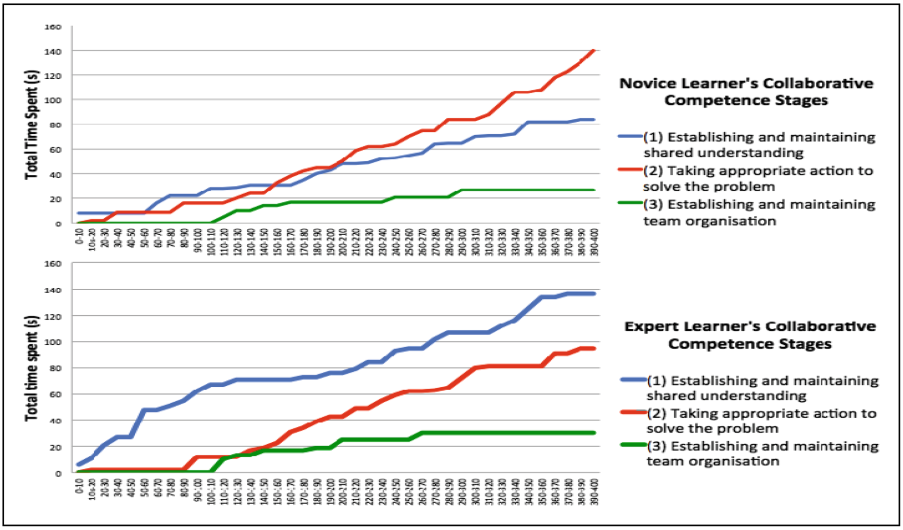



**Fig. 2.** Comparison of behaviour patterns regarding the problems solving dimensions (Novice vs. Expert)

experts and novice learners' behaviours, we compared the accumulated time each learner spent on different competency dimensions of the analysis framework. In Fig. 2, we compared the novice and experts' CPS process behaviours from the aspect of collaborative competency dimensions. As the figure reveals, novice learners spend most of their time on 'taking appropriate actions to solve the problem', while experts spend most of their time on 'establishing and maintaining shared understanding'. This is a surprising result because, one would expect both learners to spend most of their time on 'taking appropriate actions to solve the problem' due to the hands-on nature of the practice based activity. However, experts seem to spend most of their time in 'establishing and maintaining shared understanding', which is an aspect that relates more to keeping the team together than solving the problem at hand. Furthermore, both learners spend little time on the 'establishing and maintaining team organization' dimension. This result could be interpreted as indicating that both learners have high motivation to solve the problems collaboratively.

In Fig. 3 we compared the learners' behaviours regarding the problem solving competencies of the CPS process. First of all, it is clear that 'planning and executing' eventually becomes the dominant dimension in the process of both learners. Considering the hands-on nature of the practice-based activity this result is not surprising.

However, 'planning and execution' appears to start from the beginning of the process for the novice learner, but relatively later for the expert who seems to spend the initial part of the activity on identifying facts and generating hypotheses. Second, the 'monitoring, reflecting and applying' dimension starts early and stays part of the learning process for the novice learner, yet this dimension does not come until later for the expert. Finally, both learners spend only a small amount of time on 'identifying knowledge and skill deficiencies'.

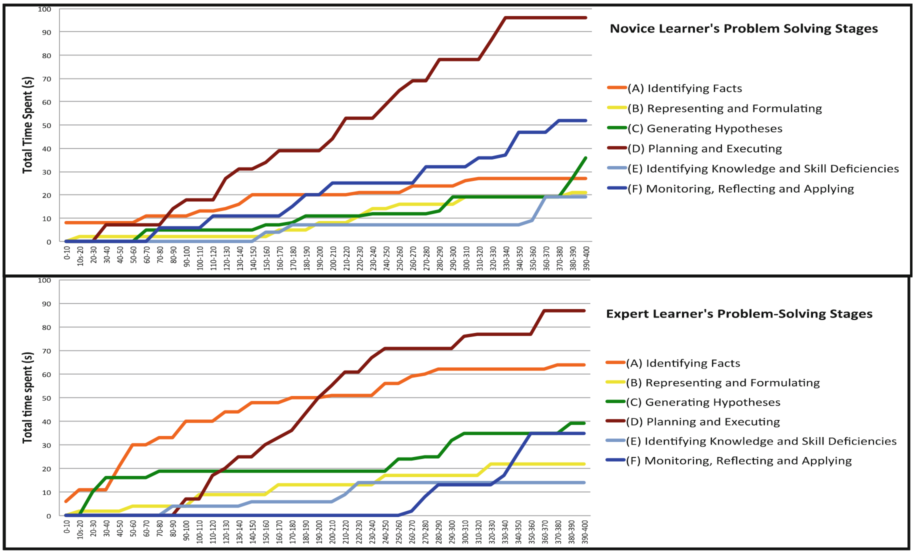



**Fig. 3.** Comparison of behaviour patterns regarding the collaboration dimensions (Novice vs. Expert)

# 4 Conclusions

In this research paper, we presented a method to identify effective strategies for CPS using learner behaviours as the key to unpacking this complex learning process. Using the analysis framework encompassing fine-grained actions of practice-based learning activities, we generated learner behaviour patterns for learners with different levels of expertise in working together. As the analysis framework offers fine-grained actions, its application is relatively easier than other frameworks with more coarse level definitions such as the OECD's CPS assessment framework. Our results show differences, which could be used to interpret effective strategies to solve problems collaboratively. For instance, experts appear to spend a significant amount of time 'identifying facts' and 'establishing and maintaining shared understanding' which appear to be practices that are less followed by novice learners. Some of these findings are similar to previous findings in expert and novice comparisons, which show that experts spend more time on problem scoping activity compared to novices [6]. For future research, we are currently working on mobile tools for on-the-fly coding of learner behaviour patterns so that the use of this method at scale and in classroom settings is more feasible.

# References

1. Dillenbourg, P.: What do you mean by 'collaborative learning'? Cognitive and Computational Approaches, pp. 1–19 (1999)
2. Worsley, M., Blikstein, P.: Analyzing engineering design through the lens of computation. J. Learn. Analytics **1**(2), 151–186 (2014)

3. Clark, R.E.: How much and what type of guidance is optimal for learning from instruction? In: Tobias, S., Duffy, T.M. (eds.) Constructivist Theory Applied to Instruction: Success or Failure?, pp. 158–183. Routledge, Taylor and Francis, New York (2009)
4. Race, P.: A Briefing on Self, Peer and Group Assessement. Higher Education Academy, York (2001)
5. Cukurova, M., Avramides, K., Spikol, D., Luckin, R., Mavrikis, M.: An analysis framework for collaborative problem solving in practice-based learning activities: a mixed-method approach. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK 2016). ACM, New York (2016). doi: http://dx.doi.org/10.1145/2883851.2883900
6. Atman, C.J., Adams, R.S., Cardella, M.E., Turns, J., Mosborg, S., Saleem, J.: Engineering design processes: a comparison of students and expert practitioners. J. Eng. Educ. **96**(4), 359–379 (2007)

# DevOpsUse for Rapid Training of Agile Practices Within Undergraduate and Startup Communities

Peter de Lange(✉), Petru Nicolaescu, Ralf Klamma, and István Koren

Advanced Community Information Systems (ACIS) Group,
RWTH Aachen University, Ahornstr. 55, 52056 Aachen, Germany
{lange,nicolaescu,klamma,koren}@dbis.rwth-aachen.de
http://dbis.rwth-aachen.de

**Abstract.** Establishing a common practice between (startup) companies and universities in applied computer science labs has been tackled by pedagogical approaches based on the communities of practice theory. However, modern agile and distributed software engineering methods and recent developments like DevOps demand focused training of undergraduate students to enable them joining practices in companies. In this paper, we present the Community Application Editor (CAE) embedded in a DevOpsUse methodology supporting this form of basic training for bachelor students of computer science. We have evaluated the methodology and the tool usage in a first-stage undergraduate lab course. The results indicate that the students had a much smoother transition when later joining the second-stage lab with real companies.

**Keywords:** Community of practice · MDWE · End user development · Case study · Entrepreneurship

## 1 Introduction

Universities with a technical focus or curriculum have an important influence on the knowledge and experience of their students, building their theoretical and practical foundation to be later used in the industry. There is a two-way benefit established from cooperations between academia and companies: real-world practice and requirements can be incorporated into university courses and teaching, preparing the students for their later employment; and companies can later make use of innovations and state-of-the-art practice that result from university research projects. Following these aspects, previous research showed that computer science students can take contact with industry within the curricula and encourage entrepreneurship following socio-cultural theories of learning [1]. Based on these foundations, a series of lab courses were held yearly at RWTH Aachen University, where groups of students were forming communities of practice together with local start-up companies for developing IT projects [1]. These


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 570–574, 2016.
DOI: 10.1007/978-3-319-45153-4_65

follow the examples from universities with a tradition in developing entrepreneurship teaching, like the MIT Entrepreneurship Lab [2] and facilitate several groups of computer science students at a Master of Science level to work on a concrete project task for and together with startup companies [1].

## 2 Supporting DevOpsUse in CoPs: A Methodology

DevOps is an emerging paradigm in software development which minimizes the gap between development and operation during agile software engineering processes. It tries to establish a new culture by a tighter integration of software development and deployment resulting in faster release cycles. The term DevOps comprises not only the methodology, but a mindset of working towards the same goal, and a collection of software tools that support this collaboration culture. Due to the missing notion of end user involvement in DevOps, we introduce the extended DevOpsUse approach that aims to unify agile practices of developers, operators and end users. We have used the DevOpsUse methodology in our practical course for teaching agile community-oriented software development. For this purpose we used Requirements Bazaar, a browser-based platform for prospective feedback, developed at our institute [3]. The Bazaar aims at supporting all stakeholders in reaching their particular goals with a common base: end users in expressing their particular needs and negotiating realizations in an intuitive, community-aware manner; service providers in prioritizing requirements realizations for maximized impact.

Further, to support a CoP in scaffolding their own Web applications and rapidly prototype ideas and architectures, we developed the CAE [4] for modeling and generating widget-based, collaborative community Web applications. We use near real-time collaborative modeling such that developers and community users can benefit from a structured approach to redesign existing applications or develop new ones. The architecture of the resulting community applications is constructed with regard to the following three key aspects: A RESTful microservice architecture backend based on las2peer[1] Web-services, a widget based frontend composed of multiple Web widgets running in a widget space and finally near real-time communication and collaboration support via the integration of collaboration frameworks.

Figure 1 shows the integrated perspective of our methodology. Our main target is to consider the learning aspect of using such a methodology, by involving students into the agile process and enable them to pursue it with the gained knowledge in industry. The rationale is to support the CoP in its development process by providing a coherent and integrated set of resources, mainly reflecting the tasks of DevOpsUse with a focus on collaboration and community practice, for speeding up the overall workflow.

[1] https://las2peer.org.

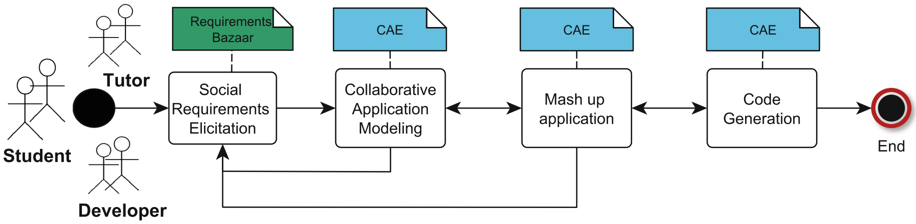



**Fig. 1.** Joint social requirements engineering and CAE methodology

## 3 Application in Lab Courses

For more than 15 years, RWTH Aachen University hosts a yearly five-months practical course on entrepreneurship for graduate students. Since 2011, we introduced a course for undergrads, that first gets students acquainted with our methodology described in the previous section, before they join the master students' projects to apply their knowledge in practice. In the following, we focus on the undergraduate course.

The lab starts with forming groups of about three students, each group working independently on a given project. This project is split up into different subtasks with an average working time of two weeks. At the end of each subtask, a review takes place where students and advisors come together to evaluate the current state of the project. The subtasks build up on each other, starting with a requirements analysis and design phase, then going over to teaching basic infrastructure setup and the basics of Web services.

About half way through the semester, the modus operandi of the course is changed from the structured tasks in the 'sandbox' environment to the real-world problems of local startups. This way, the students of the undergraduate course can apply their gained knowledge of the first half of the course in the context of a bigger project. The master students learn how to deal with the real-world situation of people coming into a project at a late stage, where much of the work is already done and the CoP has already evolved and established their working practices. In parallel, the undergrads continue working on their last subtask by refining it. At this stage, no strict requirements are enforced, giving the students the opportunity for creative problem-solving. The course finishes with a joint presentation of the produced software artifacts performed in short pitches.

## 4 Evaluation

We evaluated our teaching methodology, tools and the MDWE approach in the Winter semester of 2015–2016 with five bachelor students from RWTH Aachen University, split into two groups. Students were required to refine a short initial description in a requirements elicitation phase, via collaborative collection and discussion of requirements using the Requirements Bazaar (c.f. Fig. 1). Later on, we introduced CAE to each group in an one hour collaborative session. This was designed as an example for community formation around the software artifacts and had also the goal to familiarize the students with the technology. As the session was conducted by a tutor (i.e. expert), students could ask questions. After this, students were required to work within their groups to design and realize the complete microservices and the corresponding frontends. After handing in the final application, students were required to complete a questionnaire about their experience with our methodology.

At the time we performed the evaluation using CAE, students were already familiar with Web development using RESTful services, Javascript and HTML from previous tasks. However, as the questionnaire showed, they were not familiar with collaborative modeling as a tool for requirement analysis, system architecture design and MDWE. The teaching with CAE was rated very high from the understandability of separation of concerns between components (4.4/5) and the simplicity of the modeling framework that lead to a quick understanding of the concepts which were explained or designed (4.4/5). Results above average were obtained for learning how to design widget environments, understanding how the relations between microservices and frontend code are realized and the usability of the modeling framework for application redesign and development.

The code generation aspects of CAE were also considered to be relevant in speeding the development process and understanding technical notions. Among the advantages of CAE, students mentioned the redesign of applications, for which the tool is very useful. The collaborative work on the same resource at the same time was also considered to be helpful for learning purposes. Students suggested to improve the usability by adding a Wiki and emphasizing use-cases for classic Web applications, which do not involve widgets.

## 5 Conclusion and Future Work

In this paper, we presented a methodology and tool support for teaching undergraduate students state-of-the-art approaches for requirements elicitation, design and development of Web applications using cutting edge practices and technologies. Our main findings are that relevant tool support, social requirements, near real-time collaboration and collaborative MDWE approaches provide a solid foundation for bridging the gap between academia and industry and is able to rapidly train students for joint work with agile startups. In the future, we plan to further evaluate our method within our practical courses and to deeper investige the role of near real-time collaboration and end user development in formal teaching scenarios.

## References

1. Rohde, M., Klamma, R., Jarke, M., Wulf, V.: Reality is our laboratory: communities of practice in applied computer science. Behav. IT **26**(1), 81–94 (2007)
2. Roberts, E.B.: Entrepreneurs in High Technology: Lessons from MIT and Beyond. Oxford University Press, Oxford (1991)

3. Renzel, D., Behrendt, M., Klamma, R., Jarke, M.: Requirements Bazaar: social requirements engineering for community-driven innovation. In: 21st IEEE International Requirements Engineering Conference, RE 2013, pp. 326–327(2013)
4. de Lange, P., Nicolaescu, P., Derntl, M., Jarke, M., Klamma, R.: Community application editor: collaborative near real-time modeling and composition of microservice-based web applications. In: Modellierung 2016 (2016)

# Towards an Authoring Tool to Acquire Knowledge for ITS Teaching Problem Solving Methods

Awa Diattara[1,2](✉), Nathalie Guin[1], Vanda Luengo[3], and Amélie Cordier[1]

[1] Université de Lyon, CNRS Université Lyon 1, LIRIS, UMR5205, 69622 Lyon, France
{awa.diattara,nathalie.guin,amelie.cordier}@univ-lyon1.fr
[2] Univ. Grenoble Alpes, CNRS, LIG, 38000 Grenoble, France
[3] Sorbonne Université, UPMC Université Paris 6, CNRS, UMR 7606, LIP6, 75005 Paris, France
vanda.luengo@lip6.fr

**Abstract.** We propose a process of knowledge acquisition and an authoring tool to assist teachers who are not IT specialist to explicit knowledge needed to design ITS teaching solving problems methods. This paper describes our authoring tool and the type of knowledge to acquire.

**Keywords:** Knowledge acquisition · Authoring tool · Intelligent tutoring system · Teaching methods for problem solving

## 1 Introduction

The challenge of knowledge acquisition in Intelligent Tutoring Systems (ITS) is one of the main obstacles to their development. To overcome this problem, authoring tools have been proposed to reduce the cost of ITS design.

The purpose of AMBRE project [1] is to design ITS teaching problem solving methods [2]. AMBRE ITS are based on knowledge-based systems and use two main type of knowledge: knowledge about the methods to teach and knowledge to guide the learner when he/she solves problems, providing assistance and diagnosing his/her answers. However, designing and implementing AMBRE ITS is difficult and costly, particularly because the knowledge has to be described in Prolog, a programming language for knowledge representation. We wish to help teachers who are not IT specialist to design an AMBRE ITS in domain they are interested in. For this, we propose a process of knowledge acquisition implemented through an authoring tool. This paper presents this authoring tool and the type of knowledge to acquire.

## 2 Acquisition of Knowledge

To solve problems in a given domain of learning, AMBRE ITS rely on three knowledge containers: classification knowledge, reformulation knowledge and resolution knowledge. Classification and reformulation knowledge are used to (i) determine the class of the problem and (ii) to build a new model of the problem – called the operational model.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 575–578, 2016.
DOI: 10.1007/978-3-319-45153-4_66

Then, the solution is obtained by applying the resolution knowledge suited for the class of the problem to the operational model.

**Classification knowledge.** Problems are organized in a classification tree where a class C2 is subclass of a class C1 if any problem of C2 is also a problem of C1. The root class is defined as the most general class, and the leaves, the most specific ones.

For each class, a *discriminating attribute* is defined. This attribute must have different values in each subclass. Non discriminating attributes – called *problem attributes* – can also be defined if they make sense for problems of the class. These attributes are useful for the resolution and their values depend on the problem to solve. Classes that are specific enough so that we can assign them a resolution technique are called *operational classes.*

**Reformulation knowledge.** In order to identify the class of a problem, an AMBRE ITS uses the classification tree and a set of rules allowing, given the statement of a problem, to determine the values of the attributes (discriminating or not), thus allowing to locate the most specific class to which the problem belongs in the classification tree. A rule is defined by its *name*, a *set of premises* related to the elements of the statement and the problem attributes, and a *set of conclusions* enabling to calculate or to modify the values of the attributes.

**Resolution knowledge.** Each operational class in the classification tree has an associated solving technique. These techniques constitute the resolution knowledge. They are specific to domains of learning. For example, in the domain of arithmetic problems, a resolution technique provides a plan for solving an exercise and a formula for calculating its numerical solution.

These three knowledge bases are needed when designing an AMBRE ITS. We considered using existing authoring tools, but they do not meet our needs either because they do not match to AMBRE principle, or techniques used do not allow representing all knowledge needed by AMBRE ITS. Indeed, Pedagogy-oriented tools such as CREAM-TOOL [3] do not match to AMBRE ITS principle because of their lack of knowledge on the domain and the learner [4]. Performance-oriented tools as far as we know do not also meet our needs. ASPIRE [5] for example, is limited to models based on the constraints. Authoring tools developed around the Cognitive Tutor Authoring Tool (CTAT) [6] are the closest tools to our needs. However, tutors produced by CTAT are limited to domains where the task of problem resolution is made step by step and where all the domain knowledge can be represented in the form of production rules and consequently do not enable to acquire the knowledge needed for AMBRE ITS.

None of these authoring tools allow representing all the knowledge needed by AMBRE ITS. This is why we designed an authoring tool dedicated to the AMBRE project.

### 2.1 AMBRE-KB: An Authoring Tool to Acquire Knowledge Needed to Build an AMBRE ITS

AMBRE-KB (AMBRE-Knowledge Builder) enables to acquire knowledge of several types from the teacher and generates a Prolog version of these knowledge models.

On Fig. 1, the blue boxes represent the meta-models of knowledge to acquire.

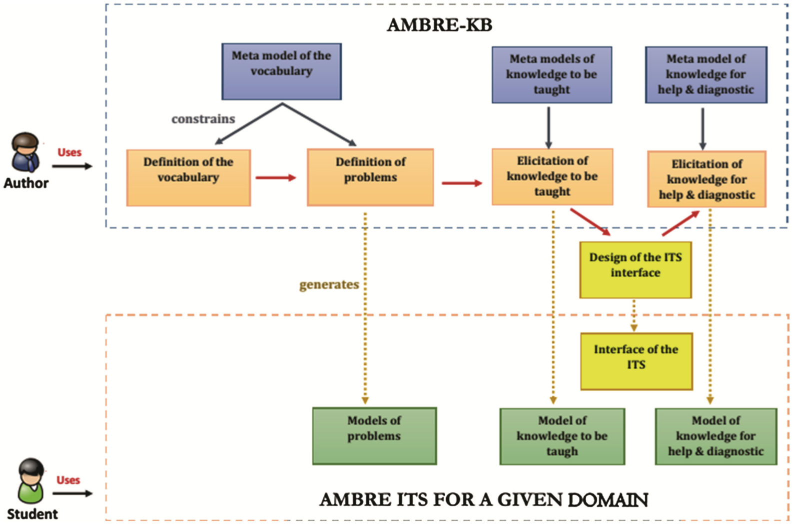



**Fig. 1.** General approach of AMBRE-KB (Color figure online)

These meta-models constrain knowledge models to be defined, and the design process allowing the user to do so. The red arrows show the process to be followed in order to explicit knowledge in a given domain of learning:

1. First, the author defines the vocabulary. In AMBRE, problems are given to the system as models that we call descriptive models. Such a model describes a situation which is the one presented in the statement of the problem to solve. To describe these models, we need to define a vocabulary.
2. The author then uses the vocabulary to define problems to solve by the system and the learner.
3. Next, the author defines the knowledge about the method, using AMBRE-KB. He/she defines knowledge of classification, reformulation and resolution.
4. The next step includes the interface design - by the teacher- and specially the tasks that the learner must perform to solve problems, based on the AMBRE cycle [1]. The development of this interface must be done by an IT specialist.
5. Finally, using AMBRE-KB, the teacher defines knowledge to guide the learner.

The green boxes represent the generated knowledge models.

For each of these types of knowledge, the system checks that the knowledge defined by the user is in accordance with the meta-models. For classification knowledge, for example, the system verifies if all non-operational classes have at least one subclass. Operational classes can have subclasses that are more specific or not. When two classes have the same discriminating attribute, the system suggests to the author to define the attribute at the level of their lowest common ancestor class.

The system offers also flexibility when defining knowledge. To define the classification tree for example, the author can choose to build the graph from the root to the leaves or vice-versa. He/she has the possibility to define all classes, and then organize them into a hierarchy. He/she can also organize the classes into a hierarchy as the definition of classes progresses. Some classes can be defined by adapting other classes.

## 3 Conclusion and Future Work

We presented AMBRE-KB, an authoring tool to acquire knowledge needed to build an AMBRE ITS. We described the knowledge needed for an AMBRE ITS to solve problems in a given domain of learning, and we propose a process to acquire this knowledge.

We are testing AMBRE-KB to evaluate its utility. It seems that the knowledge acquisition process is independent from domain of learning, and that teachers are able to explicit knowledge needed using AMBRE-KB. However, observations of the system use suggest improvements about the interactivity of the system in order to facilitate the knowledge acquisition.

This work focused on the acquisition of knowledge about the method to teach. Future work includes the acquisition of knowledge intended to guide the learner during his/her learning providing assistance and diagnosing his/her answers.

**Acknowledgments.** Thanks to Rhône-Alpes region for the scholarship that supports this thesis work.

## References

1. Nogry, S., Guin, N., Jean-Daubias, S.: AMBRE-add: an ITS to teach solving arithmetic word problems. Technol. Instr. Cogn. Learn. **6**(1), 53–61 (2008)
2. Schoenfeld, A.H.: Mathematical Problem Solving. Academic Press, New York (1985)
3. Nkambou, R., Frasson, C., Gauthier, G.: CREAM-Tools: an authoring environment for knowledge engineering in intelligent tutoring systems. In: Murray, T., Blessing, S.B., Ainsworth, S. (eds.) Authoring Tools for Advanced Technology Learning Environments: Toward Cost-Effective Adaptive, Interactive and Intelligent Educational Software, pp. 269–308. Springer, Netherlands (2003)
4. Murray, T.: Authoring intelligent tutoring systems: An analysis of the state of the art. In: Murray, T., Stephen, B., Shaaron, A. (eds.) Authoring Tools for Advanced Technology Learning Environments, pp. 98–129. Kluwer Academic Publishers (2003)
5. Mitrovic, A., Martin, B., Suraweera, P., Zakharov, K., Milik, N., Holland, J., McGuigan, N.: ASPIRE: an authoring system and deployment environment for constraint-based tutors. Int. J. Artif. Intell. Educ. **19**(2), 155–188 (2009)
6. Aleven, V., McLaren, B.M., Sewall, J., van Velsen, M., Popescu, O., Demi, S., Ringenberg, M., Koedinger, K.R.: Example-tracing tutors: intelligent tutor development for non-programmers. Int. J. Artif. Intell. Educ. **26**(1), 224–269 (2016)

# Kodr: A Customizable Learning Platform for Computer Science Education

Amr Draz[1(✉)], Slim Abdennadher[1], and Yomna Abdelrahman[2]

[1] Computer Science and Engineering Department,
German University in Cairo, New Cairo, Egypt
{amr.deraz,slim.abdennadher}@guc.edu.eg

[2] Institut fr Visualisierung und Interaktive Systeme,
University of Stuttgart, Stuttgart, Germany
Youmna.Abdelrahman@vis.uni-stuttgart.de

http://met.guc.edu.eg
http://vis.uni-stuttgart.de

**Abstract.** There are innovative systems designed for computer science education that teach programming concepts. However, many of them lack formal testing and comparison in a real course setting. This work intends to introduce a tool for teaching, evaluating, and assessing computer science students. Kodr is a modular gamified learning platform designed to evaluate varying problem types through gathering data about students performance. We conducted two studies in the wild with more than one thousand students to evaluate the initial design of Kodr. The first study evaluated two methods of teaching. The first method is to solve programming problems from scratch, the second, is to debug an incorrect solution of those problems. The results of the study yielded no significant difference between the two styles. The second study found significant positive correlations between Kodr's activity data and student's final course grades. Qualitative feedback gathered from students also evaluated Kodr as quite helpful.

**Keywords:** Computer education · Gamification · Debugging · Python · Web-based · Offline-ready

## 1 Introduction

It is a general conception that computer science is difficult to learn and that most of the programming courses have a very high failure rate [5]. There are several factors that can explain students' failure to acquire programming skills such as problem solving abilities, self-efficacy and an inability to form the correct mental model [8]. Therefore, recent years have witnessed a growing interest in fostering computer science education due to a large demand in labor force, as well as a call to develop computational thinking abilities in young students. This interest fostered an environment for innovation in computer science teaching pedagogy. In order to investigate various teaching methods suitable for teaching programing


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 579–582, 2016.
DOI: 10.1007/978-3-319-45153-4_67

in an introduction to computer science course, we developed a teaching tool named Kodr. Kodr is a modular customizable gamified learning platform used to run and track student performance through coding challenges. Kodr combines several features from tools like Coding Bat, Python Tutor [4], Pythy [3], PILeT [1], as well as TurningCraft's CodeLab [2]. Kodr is web-based and offers offline execution of Python code and programing assignments similar to Pythy. Kodr possesses PILeT's capability of accommodating varying problems types designed to test different teaching methods. Kodr hosts a wide variety of programing problems with varying difficulty, similar to Coding Bat. Kodr also offers teachers the option to fully customize the coding problems similar to CodeLab (Fig. 1).

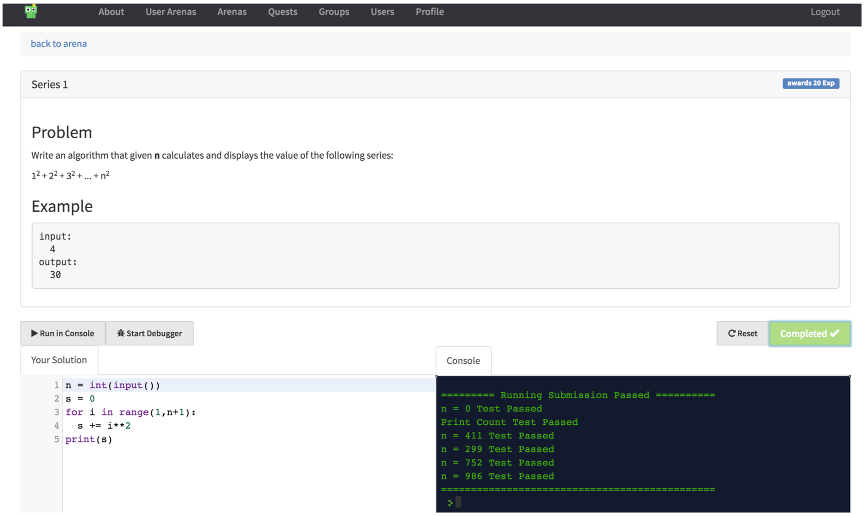



**Fig. 1.** Example of a Challenge in Kodr. Here you see an editor with debugging capabilities, a description section for the problem, and a console output section where output and submission result is printed.

Kodr extends on tools such as Pythy and Python Tutor by having an offline debugger. It helps novices while tracing their code. Additionally, it provides problems and assignments in a gamified context presented as challenges in arenas and quests with achievements respectively, awarding points on completion. Kodr has support for Javascript, Java, and Python programing challenges. In addition Kodr was designed with a completely modular challenge module capable of hosting programing game challenges similar to games like Help Gidget [6] as well as media manipulation challenges, accommodating a wider possible set of problems types than PILeT[1]. Kodr uses programmable tests suits similar to behavior driven tests which offers more flexibility for teachers than tools such

[1] A stand alone version of Kodr's python challenges can be viewed at pythondebugger.xyz.

as CodeLab and Coding Bat. The test suite allows submission evaluation not just through input and output but also through static analysis of the submitted code itself. The test suit also features tags that can be used to award students additional points and badges. Kodr keeps track and record data about students behavior patterns as they solve problems in order to collect evidence for evaluating student's problem solving abilities. It reports on student's progress for teachers and provides a training data set for turning the tool into an adaptive tutor that can automatically adapt itself to the student based on their perceived performance.

## 2 Evaluation

For Kodr's first design iteration, a study was carried out on 1078 engineering and business informatics students. 830 were male students and the rest were female students with an average age of 18. All students had no previous background in computer science/programing. The testing phase lasted for one semester spanning over 4 months.

A Pearson product-moment correlation coefficient was computed to assess the relationship between the number of challenges completed on Kodr and course grade. There was a positive correlation between the two variables: $r = 0.572$, $n = 1076$, $p < 0.01$. The positive correlation between challenge completion and final grade presents an indication that the more students engage on the Kodr system, the higher the likelihood the student would get a good grade.

250 students also agreed in a questionnaire administered post course. voting 58 % for "I had no trouble figuring out how to use Kodr". 88 % for "Kodr made it easy for me to find out how well my programs' were close to the problem solution". 80 % for "I generally found Kodr helpful in supporting my studies".

The most preferred features in Kodr for students were, being able to revisit previous lab problems (76 %), getting automatic feedback about their solution (65 %), being web-based (64 %), and being able to step through code using the debugger (63 %).

Kodr was designed to evaluate teaching methods. Accordingly, we carried out during the semester an experiment comparing the effect of whether, when faced with solving programing problems, starting from scratch (code first) as opposed to debugging a buggy solution of the same problem (debug first), would aid novice programmers in developing better understanding of programing concepts.

The study was carried out in our course delivered by two lecturers and thirteen teaching assistants. To control for the large variability, a semi random assignment was carried out across tutorial groups and lecture groups such that every teaching assistant and instructor taught both groups equally across majors. Only 449 of the 1078 students; all freshmen average 18 years of age, with minimal to no knowledge of computer science or programing, fit the criteria and opted in for completing both pre and post test. The experiment followed a between-subjects design with a pre-post test. Participants where administered the pre-test in their first lab prior to any exposure to programing concepts. The post-test was administered after the midterm (2 month), which marks the end of the programing and algorithms section of the course. The questionnaire administered was taken from [7], as it was validated. An independent-samples t-test for difference in pre-post tests noted an almost significant difference between the control ($M = 1.81, SD = 1.701$) and experiment groups ($M = 2.11, SD = 1.75$); $t(448) = 1.86$, $p = 0.064$, which was insufficient to reject the null at $<0.05$ but lied in the 90th percentile, which presents some room for future research.

## 3 Conclusion

Kodr has been used to evaluate a teaching method and aid students in learning programing. The testing results show that the tool contributed positively in the delivery of the course content. Once the data gathered through out the semester will be analyzed, we will be capable to answer more questions about students' learning patterns and to train Kodr to become an adaptive tutor capable of modulating challenge type and difficulty.

## References

1. Alshaigy, B., Kamal, S., Mitchell, F., Martin, C., Aldea, A.: Pilet: an interactive learning tool to teach python. In: Proceedings of the Workshop in Primary and Secondary Computing Education, WiPSCE 2015, pp. 76–79. ACM, New York (2015). http://doi.acm.org/10.1145/2818314.2818319
2. Barr, V., Trytten, D.: Using turing's craft codelab to support CS1 students as they learn to program. ACM Inroads **7**(2), 67–75 (2016). http://doi.acm.org/10.1145/2903724
3. Edwards, S.H., Tilden, D.S., Allevato, A.: Pythy: improving the introductory python programming experience. In: Proceedings of the 45th ACM Technical Symposium on Computer Science Education, SIGCSE 2014, pp. 641–646. ACM, New York (2014). http://doi.acm.org/10.1145/2538862.2538977
4. Guo, P.J.: Online python tutor: embeddable web-based program visualization for CS education. In: Proceedings of the 44th ACM Technical Symposium on Computer Science Education, SIGCSE 2013, pp. 579–584. ACM, New York (2013). http://doi.acm.org/10.1145/2445196.2445368
5. Guzdial, M., Soloway, E.: Teaching the nintendo generation to program. Commun. ACM **45**(4), 17–21 (2002). http://doi.acm.org/10.1145/505248.505261
6. Lee, M.J.: How can a social debugging game effectively teach computer programming concepts? In: Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research, ICER 2013, pp. 181–182. ACM, New York (2013). http://doi.acm.org/10.1145/2493394.2493424
7. Lee, M.J., Ko, A.J.: Comparing the effectiveness of online learning approaches on CS1 learning outcomes. In: Proceedings of the Eleventh Annual International Conference on International Computing Education Research, ICER 2015, pp. 237–246. ACM, New York (2015). http://doi.acm.org/10.1145/2787622.2787709
8. Ramalingam, V., LaBelle, D., Wiedenbeck, S.: Self-efficacy and mental models in learning to program. SIGCSE Bull. **36**(3), 171–175 (2004). http://doi.acm.org/10.1145/1026487.1008042

# A Reflective Quiz in a Professional Qualification Program for Stroke Nurses: A Field Trial

Angela Fessl[1(✉)], Gudrun Wesiak[1], and Viktoria Pammer-Schindler[1,2]

[1] Know-Center, Inffeldgasse 13, 8010 Graz, Austria
{afessl,gwesiak}@know-center.at

[2] Knowledge Technologies Institute, Graz University of Technology, Inffeldgasse 13, 8010 Graz, Austria
viktoria.pammer-schindler@tugraz.at

**Abstract.** Reflective learning is an important strategy to keep the vast body of theoretical knowledge fresh, stay up-to-date with new knowledge, and to relate theoretical knowledge to practical experience. In this work, we present how reflective learning prompts can enhance a medical quiz used in a qualification program for stroke nurses in Germany. In the seven-week study, 21 stroke nurses used a quiz on medical knowledge as additional learning instrument. The quiz contained typical quiz questions ("content questions") as well as reflective questions presented at different points in time. The latter aimed at stimulating nurses to reflect on the practical relevance of the learned knowledge. The results show that by playful learning and presenting reflective questions at the right time, the participants were motivated to reflect and to transfer theoretical knowledge into practice.

**Keywords:** Game-based learning · Reflective learning · Reflection

## 1 Introduction

Today's health care professionals have to work in fast-paced and changing health care environments. They have to keep the vast body of knowledge and skills fresh and up-to-date and solve complex health care problems, especially when working at stroke units. Therefore, for nurses, who embrace lifelong learning, reflective learning and reflective practice is viewed as an important strategy [5]. While reflective practice can be seen as the reconstruction and re-evaluation of experiences with the goal to learn for the future, reflective learning means to derive new insights, a change in behaviour or perception [1].

In this work, we will present the results of a field study, in which a quiz with reflective questions was integrated as additional learning instrument in a qualification program for stroke nurses. The reflective questions were presented at the beginning, during and at the end of a quiz to motivate the user to reflect at different points in time during the quiz play. The aim of the evaluation was to investigate the usefulness of the implemented reflective questions with regard to learning support and reflective learning. More in detail, we will particularly focus on the answers given to the integrated reflective questions in the quiz.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 583–586, 2016.
DOI: 10.1007/978-3-319-45153-4_68

## 2 Background and Related Work

Following Boud et al. [1], we see reflective learning as *"those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to new understandings and appreciations"*. Reflective practice is of crucial relevance for nurses, because they have to assess the health status of individuals, provide care to their patients to the best of their abilities and constantly keep-up-to-date the professional skills and social competences [7].

Initiating reflective learning with technological support is extensively investigated in formal learning environments, where prompts are used to organise, retrieve, monitor or evaluate knowledge as well as to reflect on student's learning [2,4]. At work, technology enhanced reflective learning is less investigated [3].

Quizzes are widely used in e-learning since they represent a familiar way to play and motivate students to reflect by adding meta-cognitive questions [6].

## 3 The Medical Quiz: Playful Reflective Learning

The Medical Quiz was developed for nurses who are in education to become a nurse working at a stroke unit in German hospitals. The goal of the quiz is to provide an easy and playful way of refreshing knowledge (via the content questions) and to connect theoretical knowledge with practical prior experience (via the reflective questions). The quiz was implemented with the eLearning platform Moodle[1] and four different quiz types were created: A Quiz-against-time, the Quiz-of-20 (answer 20 questions), the Quiz-of-10 and the Quiz-of-5. Altogether 142 content questions were developed by nurses and physicians working at the German stroke unit.

*Reflective Questions:* Three different types of reflective questions were implemented: "Learning progress reflective questions" at the beginning of all quizzes, "work-related reflective questions" during the Quiz-of-20 and "general reflective questions" at the end of the quizzes, except the Quiz-against-Time. The reflective questions at the beginning intend to motivate users to reflect about their knowledge status (based on previously quiz results) and their play frequency (how often the user played the quiz), for example *"You are very motivated and you play the quiz at least once per week - your results are really very good. What is your recipe for success?"*. The in-between reflective questions aim at relating the previous content question (presented together with the reflective question) to the users work practice, for example *"To what extent is the question stated above relevant for your work?"*. The question posed at the end of the quiz asks explicitly for gained insights or new knowledge with regard to the currently played quiz, for example *"Reflect on the currently played quiz. Have you gained any special insights for yourself?"*.

[1] https://moodle.org.

## 4 Method

The study was integrated into a qualification course dealing with special care at stroke units. The course took place at a German neurological clinic from October 2013 to January 2014 with one course week per month. During the first week, the Medical Quiz was introduced to the participants and they completed a pre-questionnaire to gather demographic data. During the next three months, participants could play the quiz as often as they wished and in the fourth course week, a half-day workshop and interviews were conducted at the hospital's site.

*Participants:* Twenty-one nurses (2 male, 19 female) participated in this evaluation, fourteen were aged from 20 to 29 years, seven from 30 to 59 years. The average time in their current position was 6.3 years, 81 % worked full time. 18 participants played the Medical Quiz at least once.

*Evaluation Tools:* Objective usages rates of the quiz were captured via users' log data and the written answers to the reflective questions were collected within the quiz. Demographic data was gathered in the pre-questionnaire. The interviews and the workshop provided additional information about the gained insights.

## 5 Results

Over a period of 7 weeks, 18 participants answered altogether 8314 questions, ranging from 25 to 1358 questions per user ($M = 461.9$, $SD = 341.0$). The Quiz-of-20 was clearly preferred: 18 participants played the quiz, answered on average 320.6 ($SD = 304.9$) questions and finished altogether 239 quiz attempts (on average 13.3 per user, $SD = 12.9$). The other three quiz types were played by 13 users, answering on average 24.3 ($SD = 32.9$) to 59.7 ($SD = 76.9$) questions. From all presented reflective questions, 52 % were answered in a meaningful way. In the Quiz-of-20 over 110 of the 205 presented reflective questions at the beginning were answered. For the Quiz-of-5, 38 % out of 37 posed questions were answered, for the Quiz-of-10 and the Quiz-against-time only 18 % and 13 % out of the 53 and 51 starting questions, respectively. An example for a concrete answer is *"I can recognize my state of knowledge by answering the questions several times and enhance my knowledge accordingly."* Summarizing all given responses we looked for the most frequent words to get a general impression of participants' thoughts: repetition (40), learning (27), yes (19), practice (10), retain knowledge (7), and nothing (17). Except for the Quiz-against-Time, each quiz included a reflection question presented at the end. The percentage of answered questions amounts to 54 % for the Quiz-of-20, 32 % for the Quiz-of-10, and 45 % for the Quiz-of-5. Most frequently used words in those answers were: yes (55), practice (13), learning (11), no (7), and recognise progress (5). The two in-between questions in the Quiz-of-20 have been only shortly answered in about half the cases, e.g. yes (145), no (38), very relevant (9), and combine theory with practice (4).

Especially the reflective questions at the beginning and end indicate that participants did benefit from the quiz and that reflective learning was triggered.

In the interviews the participants confirmed that they could improve their state of knowledge with regard to their work.

## 6 Discussion and Conclusion

For health care professionals like stroke nurses it is of crucial relevance to keep their knowledge up-to-date and connect theoretical knowledge with practical experiences. Thus, we implemented a Medical Quiz enhanced with reflective questions and the corresponding results confirmed that the quiz triggers reflective learning. Participants could be motivated to reflect with the "learning progress reflective questions" at the beginning and the "general reflective questions" at the end of the quiz. Especially by answering the reflective questions at the end the participants confirmed that they gained clear benefits and insights for themselves. Unfortunately, these learning outcomes were not inserted into the quiz. The "work-related reflective questions" during the Quiz-of-20 were perceived as rather disruptive for the learning process. We view the Medical Quiz with the integrated reflective questions as a viable concept for initiating reflective learning, especially where theoretical knowledge needs to be transferred into practice.

**Acknowledgement.** The project "MIRROR - Reflective learning at work" is funded under the FP7 of the European Commission (project number 257617). The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

## References

1. Boud, D., Keogh, R., Walker, D.: Reflection: turning experience into learning. In: Promoting Reflection in Learning: A Model, pp. 18–40. Routledge Falmer, New York (1985)
2. Davis, E.A.: Prompting middle school science students for productive reflection: generic and directed prompts. J. Learn. Sci. **12**(1), 91–142 (2003)
3. Fessl, A., Wesiak, G., Rivera-Pelayo, V., Feyertag, S., Pammer, V.: In-app reflection guidance for workplace learning. In: Conole, G., Klobucar, T., Rensing, C., Konert, J., Lavoué, E. (eds.) EC-TEL 2015. LNCS, vol. 9307, pp. 85–99. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24258-3_7
4. Ifenthaler, D.: Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. Educ. Technol. Soc. **15**(1), 38–52 (2012)
5. Mann, K., Gordon, J., Macieod, A.: Reflection and reflective practice in health professions education: a systematic review. Adv. Health Sci. Educ. **14**(4), 595–621 (2007)
6. O'Hanlon, N., Diaz, K.: Techniques for enhancing reflection and learning in an online course. MERLOT J. Online Learn. Teach. **6**(1), 43–54 (2010)
7. Somerville, D., Keeling, J.: A practical approach to promote reflective practice within nursing. Nursing Times **100**(12), 42–45 (2004)

# Helping Teachers to Help Students by Using an Open Learner Model

Blandine Ginon[1(✉)], Matthew D. Johnson[1], Ali Turker[2], and Michael Kickmeier-Rust[3]

[1] School of Engineering, University of Birmingham, Birmingham, UK
b.ginon.1@bham.ac.uk
[2] SEBIT Education and Information Technologies, Ankara, Turkey
[3] Knowledge Technologies Institute, Graz University, Graz, Austria

**Abstract.** The benefits of Open Learner Model for learners have been widely demonstrated: supporting learning and metacognition, facilitating self-monitoring and planning, improving self-assessment skills… In this paper, we investigate the benefits of using an OLM for teachers. 10 teachers have been using the OLM in order to monitor their class in the context of a 12 day intensive course using the speed reading application Hizligo and involving 87 students. The OLM have been regularly used by teachers, using different visualisations, mainly in the aim to identify the strengths and weakness of both their class and their individual students. Teachers found the OLM easy to use and to understand and helpful for their teaching.

**Keywords:** Open learner model · Learning analytics · Teaching analytics

## 1 Introduction

An Open Learner Model (OLM) is a learner model that is accessible to a user, in an understandable way [2]. The aims to make the model accessible to learners are to support learning and cognition, and to facilitate self-monitoring and planning [4]. OLMs can also be useful as well for other stakeholders of learning, like teachers and parents, in order to help them help learners and facilitate learners monitoring [8, 10]. Access the learner model can help teachers to identify learners' strengths and difficulties and to plan and adapt their teaching [11]. Thus, several OLM are intended for both teachers and learners (e.g. [7, 12]), some OLMs offer different visualisations for learners and teachers (e.g. [5]), especially in the cases where the learners are children (e.g. [6]). However, in these OLMs the model cannot be built from data coming from an external data source, with a competency-based approach.

In this paper, we investigate the benefits for teachers to use a competency-based OLM, in the context of a speed reading course. First, we introduce the LEA's Box OLM, a competency-based OLM intended for both teachers and learners. Then, we present how the OLM have been used in the context of a 12 day intensive course with Hızlıgo, an online speed reading application, involving 10 teachers and 87 students.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 587–590, 2016.
DOI: 10.1007/978-3-319-45153-4_69

## 2 LEA's Box Open Learner Model

The LEA's BOX OLM is a competency-based open learner model that provides teachers and learners with 12 visualisations [3], from the most simple like skill meters (Fig. 1) to more complex multidimensional visualisations like across time (Fig. 2). They can be used to visualise different information: groups' overall level, students' overall level, the level of one or several students or groups for each competency in the model, and the data coming from activities or information sources.

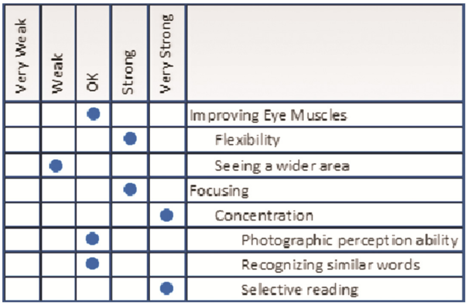



**Fig. 1.** Visualisation of the competencies using Table.

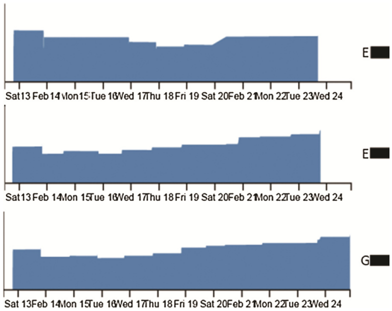



**Fig. 2.** Visualisation of the evolution of the students' models across time.

## 3 Evaluation

Hızlıgo (www.hizligo.com) is an online application intended to help learners to improve their speed reading competencies using 20 types of activities. Using Hızlıgo, learners and teachers can visualise statistics regarding the completion rate of the course and the activity scores, however, it does not provided information with a competency-based approach.

In the context of a 12 day intensive speed reading course in Turkey, 87 secondary-school students from grade 7 to 11 have been using Hızlıgo. They have been encouraged to use Hızlıgo daily, on the base of 30 min per day. Teachers have defined in the LEA's Box OLM 50 competencies and sub-competencies related to the speed reading and divided into 5 area (improving eye muscles, seeing rapidly, focusing, reading and understanding), that have then been linked to the activities provided by Hızlıgo. Every time a learner performs an activity in Hızlıgo, the outcome, using several measures, is sent to the OLM as a piece of evidence for each competencies linked to this activity. In order to monitor their students' engagement in the course and the evolution of their competencies, their 10 teachers had the possibility to use the LEA's Box OLM. Students also had the possibility to use the OLM for self-monitoring. At the beginning of the course, students and teachers have been introduced to Hızlıgo and the LEA's Box OLM. All usages have been logged. At the end of the course, a questionnaire has been send to participants about the OLM. In this section, we focus on how the OLM has been used by teachers.

The 87 students have performed an average of 61,76 activities in Hızlıgo (median = 33, minimum = 1, maximum = 275). The usages of the OLM by the teachers are presented in Table 1. The 10 teachers have been using the OLM in an average 7,9 times during the course; a session of use of the OLM lasted in average 17 min. All teachers have been using several visualisations, 3 in average, but only two visualisations have been very regularly used: the across time (used in 86 % of the OLM sessions) and skill meters visualisations (used in 56 % of the OLM sessions). Teachers frequently used the filters, mainly to monitor a given students, in 33 % of the OLM sessions.

**Table 1.** Use of the OLM by teachers.

| | Average | Median | Range |
|---|---|---|---|
| Session of use of the OLM | 7.9 | 5 | 2–29 |
| Time per session (in min) | 17 | 12 | 3–104 |
| Number of visualisations used | 3 | 2 | 1–10 |

In the final questionnaire, teachers claimed several reasons to use the OLM: 9 teachers used it to identify the weaknesses and strengths of individual students and of the group, 8 teachers used it to identify the weaknesses of the group and 7 teachers used it to identify the strengths of the group. 5 teachers also used the OLM to compare individual students' levels or the group's in different competencies. Most teachers found LEA's OLM easy to use and useful: 6 teachers found it easy to use and found the interaction with the system clear and understandable, 5 teachers found it useful for their teaching and 6 teachers claimed that using LEA's OLM make their teaching easier and enhance their effectiveness. In their comments, teachers also claim an interest of in monitoring the students' engagement in the course and their regularity.

## 4 Discussion and Conclusion

Using the LEA's Box OLM, it has been possible to define a set of 50 competencies related to speed reading, and to link them to the activities provided by Hızlıgo. The OLM provided teachers with learning analytics that were not available in Hızlıgo, in order to help them in their teaching. Although it was not the case in this first study, the LEA's Box OLM can gather information from different data sources, like several online learning applications, teacher assessment and student self-assessment.

10 teachers have been using the LEA's Box OLM in order to monitor their class in the context of a 12 day intensive involving 87 secondary school students. The teachers have been using the OLM regularly during the course. They were particularly interested in using the across time visualisation in order to see the overall evolution of a student or a group, the evolution of the level of a competency and the evolution of the scores to an activity. Teachers were also interested in using the filters facility, in order to focus on one student or competency. Most of teachers found the LEA's Box OLM easy to use and to understand, and helpful for their teaching, notably to identify the strengths and weaknesses of their class as a group or of individual students.

These promising results show an Open Learner Model intended for teachers can be a powerful tool for teachers in order to help them in their teaching by providing relevant learning analytics in a suitable way. Teachers seem to be particularly interested in seeing an overview of their students' levels and their evolution across time, but there are also interested in focusing on one student or one competency.

**Acknowledgments.** This project is supported by the European Commission (EC) under the Information Society Technology priority FP7 for R&D, contract 619762 LEA's Box. This document does not represent the opinion of the EC and the EC is not responsible for any use that might be made of its contents.

## References

1. Bull, S.: Negotiated learner Modelling to Maintain Today's Learner Models. Research and Practice in Technology Enhanced Learning (in press)
2. Bull, S., Kay, J.: Student models that invite the learner in: the SMILI:() open learner modelling framework. Int. J. Artif. Intell. Educ. **17**, 89–120 (2007)
3. Bull, S., Ginon, B., Boscolo, C., Johnson, M.D.: Introduction of learning visualisations and metacognitive support in a persuadable open learner model. In: Proceedings of Learning Analytics and Knowledge (in press)
4. Bull, S., Kay, J.: Open learner models as drivers for metacognitive processes. In: Azevedo, R., Aleven, V. (eds.) International Handbook of Metacognition and Learning Technologies, vol. 28, pp. 349–365. Springer, New York (2013)
5. Bull, S., McEvoy, A.T., Reid, E.: Learner models to promote reflection in combined desktop PC/mobile intelligent learning environments. In: Workshop on Learner Modelling for Reflection, Sydney, pp. 199–208 (2003)
6. Bull, S., McKay, M.: An open learner model for children and teachers: inspecting knowledge level of individuals and peers. In: Lester, J.C., Vicari, R.M., Paraguaçu, F. (eds.) ITS 2004. LNCS, vol. 3220, pp. 646–655. Springer, Heidelberg (2004)
7. Ginon, B., Jean-Daubias, S.: Models and tools to personalize activities on learners profiles. Ed-Media, Portugal (2011)
8. Lee, S.J., Bull, S.: An open learner model to help parents help their children. Technol. Instr. Cogn. Learn. **6**(1), 29–51 (2008)
9. Siemens, G., Long, P.: Penetrating the fog: analytics in learning and education. EDUCAUSE Rev. **46**(5), 31–40 (2011)
10. Van Leeuwen, A.: Learning analytics to support teachers during synchronous CSCL: balancing between overview and overload. J. Lear. Analytics **2**(2), 138–162 (2015)
11. Vatrapu, R., Teplovs, C., Fujita, N., Bull, S.: Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, pp. 93–98. ACM (2011)
12. Zapata-Rivera, J.D., Greer, J.: Externalising learner modelling representations. In: Proceedings of Workshop on External Representations of AIED: Multiple Forms and Multiple Roles, pp. 71–76 (2001)

# Personalized Rooms Based Recommendation as a Mean for Increasing Students' Activity

Veronika Gondova, Martin Labaj, and Maria Bielikova(✉)

Faculty of Informatics and Information Technologies,
Slovak University of Technology in Bratislava, Ilkovicova 2, 842 16 Bratislava, Slovakia
{veronika.gondova,martin.labaj,maria.bielikova}@stuba.sk

**Abstract.** In this paper we present a novel method of navigation in an educational system based on game mechanics levels. We propose a concept called rooms. More precisely, we introduce a navigation based on personalized rooms as a part of gameplay design. The room is represented by a set of items (learning objects) selected adaptively. Its main purpose is a presentation of the recommended items in a series of small sets, which supports activity of students. In gameplay design we focus on supporting students' motivation which is the key to increase students' activity. We evaluate our approach using mobile version of an adaptive learning system ALEF in software engineering domain.

**Keywords:** Personalized navigation · Gamification · Motivation · Support of activity · Personalized recommendation · Levels · Gameplay design

## 1 Introduction and Related Work

Important problem in domain of education is a lack of students' motivation associated with a low activity of students. In accordance with the fact that motivation is the source of any human activity [2], it is necessary to support it. Gamification by Zichermann and Cunningham can increase students' motivation up to 40 % [11]. The concept of gamification is not new [11]. Many systems use different mechanisms such as leaderboards, points, levels or badges to support the motivation of users [3].

The idea of levels is used in several educational systems. Even though with different forms such as status of student or a level of a game [8], its main idea is always the same – a progress of the student [3, 11]. Level, as a status of a student, represents the position of the student in the system [8]. This type of level is used also by an educational system Moodle [6]. The second form of levels represents typical levels in games. In this case content of system is organized into smaller units called levels. One of the systems that use both types of levels is system Memrise.

Another way to increase the activity of students is personalization. Personalization can cause an increase in students' satisfaction [5], which is associated with an increase of students' activity. One of the most popular ways of personalization is personalized recommendation. The recommendation aims to simplify and streamline users' activity in the system [7, 10]. Currently there are several methods of recommendation including collaborative filtering, content-based filtering or hybrid recommendation [9].


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 591–594, 2016.
DOI: 10.1007/978-3-319-45153-4_70

Educational systems with this types of the recommendation are Wayang Outpost, ALEF, Coursera or Moodle [4].

## 2 Navigation Based on Personalized Rooms

Existing approaches use the levels as a mean to express the progress. However, there is also a potential to use them as a tool for navigation while the original concept is exactly used for the motivation. In order to support the activity of students in the system we propose a method of navigation between those groups. Our method is based on dynamic personalized distribution of items (learning objects) into smaller groups called rooms and navigation between these groups.

The main difference of rooms and typical levels is in distribution of items into the rooms that is based on personalized recommendation of items. The items are selected adaptively based on the configuration of two recommenders. The first recommender hides already solved items. The personalization of rooms is fully realized in the second recommender (realized as IRT recommender) that recommends items from simpler to more complex ones for each student separately. The probability of students' correct answer to a question is computed through two-parametrical model of an item response theory (2P IRT) that provides information about this probability.

The navigation between the rooms is a basis for the gameplay principle. At the beginning of a week each student has only one room available. Achieving the necessary activity in the current room is a condition for an opening of the next room. Every room can be used to open a new one no more than once. If a student is active enough in current room he/she can open a new room, otherwise he has to work again with the items in the current room (Fig. 1.).

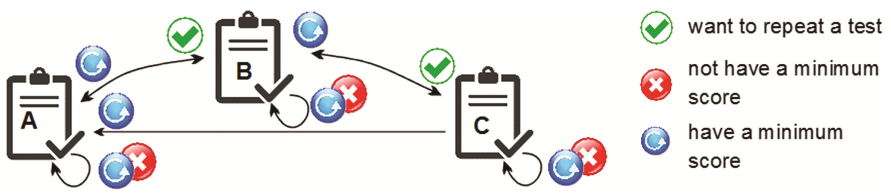



**Fig. 1.** Principle of personalized navigation between the rooms. After completing a test in room A, score of A is compared with a threshold score, which can cause the creation of a new room or repetition of current test.

Success of the student's try is determined by comparing two types of scores – threshold score and score of current try. Threshold score reflects minimal activity that student has to demonstrate to open a new room. This score represents the score obtained for M average correctly answered items. Score of the current try is a sum of two types of score, score for commenting and score for answering. Every type of score is regularly calculated and depends on the actual difficulty and importance of items in the current room.

## 3 Evaluation

We integrated our method of navigation in the recommended items into the mobile version of Adaptive Learning Framework ALEF [1] (aleftng.fiit.stuba.sk). ALEF is used by students during the semester as their preparation for entry tests in the course of Software Engineering. It contains a set of questions for every week selected manually by a teacher based on the identification of concepts that are taught that week.

We organized a three-week experiment with 250 students. We divided students into two groups based on the activity of students in the system before the experiment and on their study results aimed to make the groups equivalent. Students in the control group worked with the original version of ALEF and students in the experimental group worked with a new version of ALEF with implemented personalized rooms. We monitored students' activity expressed by the interactions of students in ALEF.

After the first week of the experiment we provide a questionnaire for students to determine if personalized rooms cause some problems. This questionnaire was answered by 64 students (44 from experimental group and 20 from control group). Depending on the results of the questionnaire we can claim that our method reduces the number of students for which number of items in the system was causing frustration by 21 % which is a significant result ($H_0$: The percentage of students which said that the number of items in the system caused frustration is same for both groups; Mann-Whitneyho U test; $p = 0{,}03412$; $p < 0{,}05$ - $H_0$ is rejected). The second interesting result of the questionnaire is that up to 86 % of students with personalized rooms said that this version of ALEF is better than original version of ALEF.

After three weeks of the experiment we observed 124 active students (61 in control group and 63 in experimental group), 21 674 of students' logs in the system (including 8580 interactions with learning objects) and 37 comments. Our results show that our method increased activity of students (activity = number of interactions with learning objects). The number of interactions in the experimental group was higher by 8 % compared to control group. However, this result was not significant.

Despite this our method was able to significantly increase the proportion of interactions to the logs ($H_0$: The proportion interactions/logs is same for both groups; Mann-Whitneyho U test; $p = 0{,}00548$; $p < 0{,}05$ - $H_0$ is rejected). It means that our method increased the percentage of the activity that consists of answering to items to total activity of student in the system. Total activity is equivalent to logs and includes interactions with the questions and also the display of a question or correct answer to a question. Another interesting result was a significant increase of comments in the system, while students in the control group added 7 comments, students in the experimental group added 30 comments ($H_0$: The amount of added comments is same for both groups; Mann-Whitneyho U test; $p = 0{,}0463$; $p < 0{,}05$ - $H_0$ is rejected).

The last result is a significant reduction of the interactions of type "I do not know" by 67,81 % ($H_0$: The number of interactions of type "I do not know" is same for both groups; Mann-Whitneyho U test; $p = 0{,}03412$; $p < 0{,}05$ - $H_0$ is rejected). This type of interaction is recorded as explicit feedback from the students by clicking on the button "I do not know". This result means that our method motivates students to solve questions and not only click on some button to see the result. This difference is due to calculating the actual score in room. Students get higher score for answering question (correct or incorrect) than clicking the button “I do not know”.

## 4 Conclusions

The goal of our work is to support activity of students. For this purpose we proposed a method of navigation within items (learning objects) based on a distribution of recommended items into the rooms. We evaluate our method through an experiment with two groups of students (experimental condition = mobile ALEF + adaptive rooms and control condition = mobile ALEF without rooms). The results show that our method increased activity by 8 %. Our method also significantly decreased number of students, who said that the number of learning objects in the system caused frustration. Another significant result is an increase of proportion logs/interactions and number of comments in the system. The last significant result is reduction of the interactions of type “I do not know” by 67,81 %.

**Acknowledgement.** This work was partially supported by grants APVV-15-0508, KEGA 009STU-4/2014 and it is the partial result of the collaboration within the SCOPES JRP/IP, No. 160480/2015.

## References

1. Bieliková, M., et al.: ALEF: from application to platform for adaptive collaborative learning. In: Manouselis, N., Drachsler, H., Verbert, K., Santos, O.C. (eds.) Recommender Systems for Technology Enhanced Learning, pp. 195–225. Springer, New York (2014)
2. Deci, E.L., Ryan, R.M.: Intrinsic motivation. Wiley, Hoboken (1975)
3. Deterding, S., Sicart, M., Nacke, L., O’Hara, K., Dixon, D.: Gamification. using game-design elements in non-gaming contexts. In: Extended Abstracts on Human Factors in Computing Systemspp, CHI 2011, pp. 2425–2428. ACM (2011)
4. Drachsler, H., et al.: Recommendation strategies for e-learning: preliminary effects of a personal recommender system for lifelong learners (2007)
5. Ferrer, F.: Personalisation of education. In: Personalisation of Education in Contexts, pp. 109–127. Sense Publishers (2012)
6. Kiryakova, G., Angelova, N., Yordanova, L.: Gamification in education. In: Proceedings of 9th International Balkan Education and Science Conference, pp. 32–39 (2014)
7. Michlik, P., Bielikova, M.: Exercises Recommending for Limited Time Learning. Procedia Computer Science **1**(2), 2821–2828 (2010). Elsevier
8. Mullins, W. L.: Game with multiple incentives and multiple levels of game play and combined lottery game with time of purchase win progressive jackpot. U.S. Patent No 6,210,276 (2001)
9. Pazzani, M.J.: A framework for collaborative, content-based and demographic filtering. Artif. Intell. Rev. **13**(5–6), 393–408 (1999)
10. Tintarev, N., Masthoff, J.: Designing and evaluating explanations for recommender systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 479–510. Springer, New York (2011)
11. Zichermann, G., Cunningham, C.: Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps, pp. 35–66. O’Reilly Media, Sebastopol (2011)

# Detecting and Supporting the Evolving Knowledge Interests of Lifelong Professionals

Oluwabukola Mayowa Ishola(✉) and Gordon McCalla

Department of Computer Science,
University of Saskatchewan, Saskatoon, Canada
bukola.ishola@usask.ca, mccalla@cs.usask.ca

**Abstract.** Our research is tackling a challenging problem in lifelong learning: helping practicing professionals to identify emerging gaps in their knowledge as the knowledge base of their profession evolves and changes over time. Our specific goal in this paper has been to see if the later knowledge interests of programmers (as exhibited by their behavior in the StackOverflow (SO) forum) could be predicted from their earlier behavior in SO. We examined the past behavior of each programmer over a long-term (4 year) baseline as well as a short-term (6 month) baseline, and used a Bayesian approach to predict each programmer's later knowledge interests. When comparing these predictions to their actual later interests (as demonstrated in SO), we achieved recall values of 0.70 (using the long term baseline) and 0.93 (using the short term baseline) with precision values of 0.61 (long term) and 0.81 (short term), implying that a short term baseline is better for prediction than a long term one. This is promising for creating a system that can automatically track the evolving changes in professional knowledge by observing the questions and answers of professionals as they themselves interact about these changes.

**Keywords:** Personalization · Lifelong learning · Professional development

## 1 Introduction

Rapid technological advances are leading to massive ongoing change in society and work, driving the need for lifelong learning of the new skills and knowledge needed to succeed in this changing world [1]. In the advanced learning technology research community, increasing interest in personalizing learning technology tailored to lifelong professionals according to their evolving learning needs is consequently on the increase [2]. As knowledge evolves, professionals will need to continually update their knowledge to effectively participate in professional development.

Knowledge can be classified [3, 4] into the things we know we know, the "known knowns" (KK); the things we know we don't know, the "known unknowns" (KU); the things we are not aware we know but we do know, the "unknown knowns" (UK); and, lastly, the things we don't know we don't know, the "unknown unknowns" (UU). Both KU and UU signify gaps that exist in the knowledge of the professional. In supporting the lifelong professional whose knowledge interests evolve over time, our long-term research goal is to predict the future knowledge needs (KU and UU) of each


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 595–599, 2016.
DOI: 10.1007/978-3-319-45153-4_71

user in order to help them identify these emerging gaps in their knowledge. Our short-term goal in the research reported upon in this paper is to try to predict the changing knowledge interests of users of Stack Overflow, who are programmers seeking to ask or answer questions about software and programming issues. We wanted to diagnose from each user's past interactions in SO what their knowledge interests were, and then to predict how these interests would evolve going forward. To measure the quality of our predictions, we then wanted to compare how these interests actually emerged in SO.

## 2 Tag Classification

Tags are employed in SO to describe the question being asked which also helps users to determine the questions they will be able to answer. While creating a question, a maximum of 5 tags and a minimum of 1 tag can be employed. We classified the various tags employed in SO into 19 suitable computing related classes, which represent the possible knowledge interests of the users. The top tag classes with corresponding tags mapped to them are shown in Table 1.

**Table 1.** Tag classification

| Tag Class | Number of Tags | Example of Tags |
|---|---|---|
| General computing | 31409 | Rounding, detection, vocabulary |
| Software coding | 1567 | Java, C#, python |
| Web application | 1293 | Php, html, asp.net |
| Mobile application development | 951 | Android, ios, blackberry |
| Database systems | 731 | Mysql, sql, oracle |
| Framework and library | 440 | .net, playframework, zend-framework2 |
| Operating system | 330 | Windows, powershell, active directory |

## 3 Inferring a User's Knowledge Interests

Using the tag classes discussed in Sect. 2 we mapped each user's question post to a tag class based on the tag associated with the question post. In cases where more than one tag was used, we counted the number of tag classes that occurred in the post and the post was assigned to the class with the highest frequency of occurrence. We then wanted to look at evolving interests over a short term and long term baseline. Questions asked from January 2009 to December 2011 were used to infer each user's long term knowledge interests while questions asked between March 2014 and July 2014 were used as a basis for inferring their short term knowledge interests. The specific knowledge interests of the user were determined by mining all tags employed in questions asked by the user during the long and short term time periods. To determine the tag classes where the interest of the individual user lies, we computed the tag distribution D(u,t) employed in question posts for each user as described below:

$$D(u,t) = (\frac{N_1}{N_{total}}, \frac{N_2}{N_{total}}, \ldots, \frac{N_n}{N_{total}}), \quad \text{where } N_{total} = \Sigma_i N_i$$

The count of questions asked by user *u* for the tag-class *i* is represented by $N_i$, while $N_{total}$ shows the total number of questions asked by the user for the defined time frames (long and short term) for all the tag classes represented in their profile. The tag with the highest tag distribution as computed using both the long and short-term data samples is inferred as the genuine knowledge interest of the user in the long and short term respectively.

Having inferred the knowledge interest of each user from their historical learning activities, we then looked at what their knowledge interests were in the time period right after their baseline interests were examined. For the long term analysis, we selected question posts made in 2012, the year after the long term baseline; and for the short term analysis we looked at the posts made in August 2014, the month after the short term baseline. The 100 most popular question posts for each user were selected from the test data as having tags that might represent the future knowledge interests of the user. Popularity of a question was determined by the number of views the question had (information that is available in SO). Selecting the most popular posts helps to tailor predictions of knowledge interests so they align with trends within the learning community, and allows for the possibility of tracking the evolving knowledge of the discipline over time. It should be noted that while a similar number of question posts might be selected for users with the same knowledge interests, with post ranking the set of posts containing tags which will be predicted as a given user's future interests will differ based on the historical activities of each individual user.

## 4 True Bayesian Estimation

The 100 most popular posts for each user as discussed in the previous section were ranked using a True Bayesian estimate [3]. The True Bayesian estimate is computed as shown below:

$$w = \frac{v}{v+m}R + \frac{m}{v+m}C.$$

In this equation w = weighted rating, R = average rating of observed data, v = number of votes for the observed data, m = weight given to the prior estimation, C = the mean vote across the whole pool. Using the equation above, each post is assigned a computed weighted rating which shows the relevance of the post to the user in comparison to all selected posts. Tags from selected question posts with a weighted rating greater than 0.7 are predicted as the future knowledge interests of the user. It should be noted that the *'R'* and *'C'* components of the Bayesian estimation take into consideration the previous rating of individual users and that of the professional community respectively.

Therefore, as the knowledge interests of individual users and the community evolve over time, the values of *'R'* and *'C'* will also change accordingly to adapt to the current interests of each user.

## 5 Evaluation

In evaluating the results of this study for our long-term prediction, we compared all the tags actually used by an individual user in the year 2012 with the predicted tags. Likewise, we evaluated our prediction of the short-term knowledge interests of the user by comparing the tags employed by the user in August 2014 with the predicted interests of each user. Precision and Recall were computed for each user based on this comparison for their long and short-term knowledge interests

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fp} \qquad \text{F}-\text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}.$$

$T_p$ represents *True Positive* (which is number of tags used and recommended), $f_p$ is *False Positive* (which is the number of tags recommended but not used) while $f_n$ is false negative (which is the number of tags used but not recommended). Since F-measure is computed using both precision and recall, it allows the overall effectiveness of the recommender system to be determined. Table 2 shows the average recall, precision and F-measure for the long-term and short-term learning needs.

**Table 2.** Evaluation of results

| Time duration | Recall | Precision | F-measure |
|---|---|---|---|
| Long term | 0.70104 | 0.61485 | 0.56553 |
| Short term | 0.92959 | 0.80909 | 0.83273 |

We observed higher precision and recall with the predictions made using the short-term learning data as compared with the long-term data. These seem to be fairly good levels of accuracy, particularly predictions made from the short-term baseline.

## 6 Discussion

Being able to predict how a user's knowledge interests evolve from their SO behaviour is a first step along the road to being able to build an open user model that could inform the user of their impending knowledge needs (their KUs and UUs). This is especially important for the UUs, of course, since knowledge that a professional needs to know (but that they don't know they need to know) will be a serious impediment to maintaining their professional competence. Of course, our current work needs further confirmation. We need to explore more sophisticated mappings of tags to tag classes, and more elaborate ontologies of tag classes that better capture the professional body of knowledge. We need to conduct further experiments on varying baselines. We need to explore how other information in SO can augment our diagnoses and predictions (in fact we have already carried out preliminary experiments drawing on user reputation and badges). Even so, we believe that the general approach we have taken is very promising since it relies on actual interactions among practicing professionals and can potentially track not only ongoing changes in individual user knowledge, but also emerging new knowledge important to the profession.

**Acknowledgments.** The authors wish to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the U of Saskatchewan for funding this research project.

## References

1. Sharples, M.: The design of personal mobile technologies for lifelong learning. Comput. Educ. **34**(3), 177–193 (2000)
2. Tang, T., McCalla, G.: Smart recommendation for an evolving e-learning system: architecture and experiment. Int. J. E-learn. **4**(1), 105–129 (2005)
3. Dunning, D.: The Dunning-Kruger effect: on being ignorant of one's own ignorance. Adv. Exp. Soc. Psychol. **44**, 247 (2011)
4. Ishola, O., McCalla, G.: Tracking and reacting to the evolving knowledge needs of lifelong professional learners. In: Proceedings of the 6th Workshop on Personalization Approaches in Learning Environments (PALE 2016) at the 24th International Conference on User Modeling, Adaptation, and Personalization (UMAP 2016), pp.68–73 (2016)

# Boosting Vocational Education and Training in Small Enterprises

Miloš Kravčík(✉), Kateryna Neulinger, and Ralf Klamma

Advanced Community Information Systems (ACIS), RWTH Aachen University, Informatik 5, Ahornstr. 55, 52056 Aachen, Germany
{kravcik,neulinger,klamma}@dbis.rwth-aachen.de

**Abstract.** Learning and training at the workplace is critical for economic development of companies and their competitiveness. Nevertheless, it is known that especially small firms have difficulties with long term planning and systematic cultivation of employees' knowledge and skills. The challenge is to integrate learning and training activities into the work process and to provide benefits and incentives for both managers and employees, which would motivate both to use such services. The main aim of our study was the development of a Web-based learning environment that supports this objective as well as piloting and evaluation in real settings. The outcomes have shown that although it is not easy to get small enterprises involved in such experiments, there is a potential to use personal learning environments for supporting workplace learning in small companies.

**Keywords:** Workplace learning · Personal learning environments · Design

## 1 Introduction

Small enterprises represent the vast majority of companies in Europe, employ a huge number of people, and provide a large portion of European's economic power. As their participation rates in Vocational Education and Training (VET) are declining in the EU, this is a big problem and there is a real need to engage them in developing a positive attitude towards training [1]. The EU Leonardo-Da-Vinci BOOST (Business perfOrmance imprOvement through individual employee Skills Training) project aimed to improve the participation of small enterprises (up to 20 employees) at vocational education and training programs. It integrated results from two predecessor EU projects (LLL Leornado-Da-Vinci BeCome and EU FP7 Integrating Project ROLE). The solution enables small enterprises to identify their critical business needs and then to organize the learning process in order to meet them. Of course, it is crucial to consider the interests of all stakeholders in order to motivate them to use the tools. Another important requirement in this context is the seamless integration of learning into work processes. In this paper, we first introduce related work. Then an explanation of the BOOST methodology and technology follows. The core is a presentation of the outcomes from the qualitative evaluation. We conclude the paper summarizing our main findings.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 600–604, 2016.
DOI: 10.1007/978-3-319-45153-4_72

## 2 Related Work

Workplace learning in small enterprises has been reviewed in [2] where the author specified main problems associated with engaging these enterprises in training activities. One of them is lack of internal capacity and motivation to provide learning opportunities for employees. This requirement is supported in [3] by the claim that workplace learning takes place in work processes and on a just in time basis, is multi episodic, often informal, and problem based. In the context of lifelong and informal learning at the workplace, also Self-Regulated Learning (SRL) plays an important role. The SRL skills need to be cultivated and can be supported by properly designed Personal Learning Environments (PLEs) [4]. In the BOOST project we addressed the issues of informal workplace learning considering the demands both of managers and employees by providing tailored PLEs.

## 3 BOOST Methodology and Technology

The challenge was to integrate the sound methodology from the BeCome project (http://become.dedi.velay.greta.fr/) and the widget-based technology from the ROLE project (http://www.role-project.eu/). 4 phases of learning processes were supported. In *Planning* business goals in the company (with competences) are specified and employees to address them are selected. In *Tutoring* learning resources are assigned to target competences. In *Learning* access to learning resources and search facilities is provided. *Reflection* means monitoring of the learning progress of the company, as well as of individual employees. The created hierarchy has *Business Goals* at the top. Each of them refers to relevant *Learning Indicators* (competences) and for those *Learning Resources* (materials, tools, and peers) are recommended. We distinguished 2 user roles. *Manager* specifies business goals with learning indicators and assigns them to employees. This role covers also assignment of learning resources to learning indicators and monitoring of employees' learning progress. *Employees* can view their learning tasks, learn by accessing the resources, and reflect on their progress.

The BOOST platform [5] is a widget-based Web application, developed with ROLE Software Development Kit (https://github.com/rwth-acis/ROLE-SDK). Users can easily adjust the arrangement and functionality of their learning environments according to their needs and preferences. The software enables inter-widget communication and is open source. After login users enter the *Start* area, where the preferred language can be chosen and managers can assign roles to users. In the *Management* area managers specify business goals and assign them learning indicators with priorities. Then they can assign learning goals with target proficiency levels and deadlines to employees. The overall and individual progress of all employees can be monitored there. The main difference for employees in this area consists in having access only to their own data, which was a crucial requirement from our users. Managers do their tutoring and employees their learning tasks in the *Learning* area, which shows learning resources assigned to learning indicators (and business goals), displays the selected learning resource for learning, and allows searching for learning materials in predefined repositories.

## 4 BOOST Evaluation

The methodology and technology developed in the BOOST project was later on evaluated in the piloting phase. First we had to recruit suitable companies for testing the BOOST methodology and platform. The target group consisted of small enterprises with less than 20 employees. The BOOST partners contacted the enterprises that were available for our piloting and for each of them an individual plan was developed, depending on their preferences and constraints. The BOOST piloting phase started on November 2014 with preliminary actions, including the development of engagement material and the recruitment process, and ended in August 2015 with the evaluation of the piloting results. The duration of the individual cases varied from just a few days to 3 months. We have performed both quantitative and qualitative evaluation. The results of the quantitative one have been reported in [6]. Here we focus on the qualitative part of our evaluation.

Our evaluation shows that 88 % of managers found the BOOST approach of linking learning to their business goals as very or quite useful. The support in understanding and implementing the BOOST methodology and tools was perceived by managers mostly as good or excellent. The managers rated the usability, user friendliness, and graphical presentation of the BOOST online tool prevailingly as good or adequate. 88 % of managers found the results of BOOST piloting as quite or very useful, contributing to increase the employees' skills according to the company business goals. 80 % of managers expressed their interest in using the BOOST methodology and tools in the future.

88 % of the participating employees found the BOOST approach quite or very useful in increasing their personal skills. Employees rated their support for training as good and their rating of the BOOST online tool tended from good towards adequate in terms of usability, user friendliness, and graphical presentation. All in all, they found the results quite useful. 87 % of the participating employees found the results of the BOOST pilots quite or very useful and most of them thought the system would contribute to the development of their competences towards the company goals.

The piloting reports generated qualitative and quantitative evaluation data. Generally, the evaluation shows, that BOOST addressed a very relevant problem. Participating enterprises and their employees highly valuated the relevance and overall helpfulness of the BOOST approach. Some results also pointed to issues raised, such as stability issues, search results offered in the platform, its dependencies on human factors (such as the quality of the assigned learning tasks), included sources for search, interactivity restrictions, reporting restrictions. Some participants also offered proposals for the further improvement of the platform, including a new user interface design, translation issues, communication functionalities, and mobile versions (https://requirements-bazaar.org/#!/projects/8).

Among the lessons learned from the project are insights, that the problem of addressing small enterprises with tailored VET offers is more complex than previously thought. Efforts to increase their participation in VET need to be further increased in order to reach the goals set out on a scalable level. BOOST represented an important step in this direction, but this relatively small project needs to be complemented by further research and development activities, by the uptake of methods and tools, and by support on various societal levels.

## 5 Conclusion

The BOOST experience showed that there is a potential to use personal learning environments to support workplace learning in small companies. For reaching this goal, we managed to create some methodological innovations and supporting implementations using open-source technologies. One of the basic requirements was a user friendly solution both for companies and for employees, in order to motivate them to use it. The evaluation showed some clear benefits in easy organization of workplace learning and progress monitoring. At the same time important suggestions have been made how to further improve this process, especially to consider additional requirements, including team learning, automatic assessment, various levels of privacy and rights, as well as mobile learning and modern interfaces. Moreover, the piloting also clearly revealed that it is very difficult to involve a target group as diverse as small enterprises in the evaluation process, as their resources are very limited and valuable. In summary, the BOOST project (http://www.boost-project.eu/) represents an important step towards the better inclusion of MSEs and their employees in VET programs in order to consolidate and strengthen their economic role for European societies. Our workplace learning research continues in the follow-up projects Learning Layers (http://learning-layers.eu/) and WEKIT (http://www.wekit.eu/). They deal with scalability issues in informal learning and wearable experiences for knowledge intensive training respectively.

**Acknowledgments.** The presented research work was partially funded by the 7th Framework Programme large-scale integrated project Learning Layers (grant no: 318209) and by the H2020 project WEKIT (grant no: 687669). We appreciate very much the contributions of all the BOOST partners as well as of the external evaluator.

## References

1. European Commission. Rethinking Education: Investing in skills for better socio-economic outcomes. COM, 669 (2012)
2. Johnson, S.: Lifelong learning and SMEs: issues for research and policy. J. Small Bus. Enterp. Dev. **9**(2), 285–295 (2002)
3. Attwell, G., Deitmer, L.: Developing work based personal learning environments in small and medium enterprises. In: The PLE Conference, Melbourne (2012)
4. Nussbaumer, A., Kravčík, M., Renzel, D., Klamma, R., Berthold, M., Albert, D.: A Framework for Facilitating Self-Regulation in Responsive Open Learning Environments (2014). arXiv preprint: arXiv:1407.5891

5. Kravčík, M., Neulinger, K., Klamma, R.: Boosting informal workplace learning in small enterprises. In: Proceedings of the 4th Workshop on Awareness and Reflection in Technology Enhanced Learning *(ARTEL),* Conjunction with the 9th European Conference on Technology Enhanced Learning (EC-TEL), vol. 1238, pp. 73–75. CEUR (2014)
6. Kravčík, M., Neulinger, K., Klamma, R.: Data analysis of workplace learning with BOOST. In: Proceedings of the Workshop on Learning Analytics for Workplace and Professional Learning (LA for Work), Conjuction with the 6th International Learning Analytics and Knowledge Conference, 25–29 April 2016, Edinburgh, UK (2016)

# Supporting Teaching Teams in Personalizing MOOCs Course Paths

Marie Lefevre[1(✉)], Nathalie Guin[1], Jean-Charles Marty[2], and Florian Clerc[1]

[1] Université de Lyon, CNRS - Université Lyon 1, LIRIS, UMR5205, 69622 Villeurbanne, France
{marie.lefevre,nathalie.guin,florian.clerc}@liris.cnrs.fr
[2] Université de Lyon, CNRS - Université de Savoie, LIRIS, UMR5205, 69622 Villeurbanne, France
jean-charles.marty@liris.cnrs.fr

**Abstract.** One challenge that the MOOCs must face in order to ensure their durability is to provide learners with personalized trails. This paper proposes a model allowing the implementation of personalization in MOOCs. Its purpose is to enable teachers and MOOCs designers to express their educational objectives in order to obtain an adaptation of the courses to everyone.

**Keywords:** MOOC · Personalization · Pedagogic strategy · Adaptive learning

## 1 Introduction

One of the main issues relating to MOOCS is due to the diversity of learners who join a MOOC. Learners necessarily have different expectations, initial knowledge or ways of learning. However, there is currently only one course offered to learners, and this course does not necessarily suit all of them. This issue is at the heart of current research on MOOCs through the analysis of learners' behavior. As the number of learners in a MOOC is too important to rely on tutors, many believe that the personalization of learning, especially using learner profiles, is the most effective solution.

Several studies with the goal of personalizing MOOCs have emerged within the past three years. These works provide automatic personalization processes, without involving the MOOC teaching team. Our approach is to give to the MOOC teaching team the possibility to define personalization strategies that will be implemented in the platform. Therefore, we propose to exploit the PERSUA2 model [1], originally proposed to personalize educational activities involving a single learner, especially those using ITSs. In this model, the teacher's role is to define a personalization strategy, as a set of pedagogical rules specifying which activities should be offered to a learner, based on the characteristics contained in his/her profile. Activities available in an ITS and the parameters enabling to choose or configure them are described in a model respecting the AKEPI meta-model [2]. The teacher also defines a context of use, which describes the situation in which learners will carry out the activities. For each learner, the system implementing PERSUA2 can thus build activities that meet his/her characteristics (learner profile) according to the teacher's wishes (pedagogical strategy) and in the


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 605–609, 2016.
DOI: 10.1007/978-3-319-45153-4_73

context of a given session (context of use). As our aim is to use this model to personalize MOOCs, we studied its limits in this new context.

## 2 PERSUA2$_{MOOC}$: A Model for Personalizing MOOCs

**From ITSs to MOOCs: A Necessary Adaptation of the PERSUA2 Model.** In the PERSUA2 model, the teacher is the only actor involved in the personalization process, since s/he first has to instantiate the different models used (learner profile, context, activities), and s/he must then define the personalization rules. However, the design of a MOOC is a more complex process and involves many people. We believe each of these actors can play an important role in the personalization process. The roles that we have identified are: (1) the designers of the personalization module: this is the role we (the reserachers) have taken, designing and implementing the model allowing personalization; (2) the platform administrators: they are the people who manage the MOOC platform; (3) the educational team of the MOOC, which provides the content of the MOOC; (4) the learners.

As designers of the personalization module, we have specified a generic model for personalizing MOOCs, based on the PERSUA2 meta-model allowing the description of pedagogical strategies, and based on the AKEPI meta-model allowing the description of activities. This model specifies how to describe learner profiles, teaching strategies, context and activities within the MOOC. Our models of learner profiles and activities are not intended to be final and used necessarily as they are within a MOOC. They describe the general structure and the types of information they should contain. However, each MOOC platform having its own specificities (e.g. different features, traces…), administrators could be able to modify the elements contained in these models so that they best fit their system, and allow to describe the learner and the MOOC platform in a relevant way. Similarly, each MOOC is unique by its contents and objectives. The teaching team could then modify models of learner profiles and activities, in order to describe precisely the activities for a particular MOOC and the information to be obtained on learners when they perform these activities. Finally, the learner has also a role to play, as we will see later.

**Operating Process of the PERSUA2$_{MOOC}$ Model.** The different parts of the PERSUA2$_{MOOC}$ model are used within an automated process, in order to provide recommendations to each learner (see Fig. 1). Five elements constitute the input of the process. Two of them will be used to characterize the learner, and are calculated automatically:

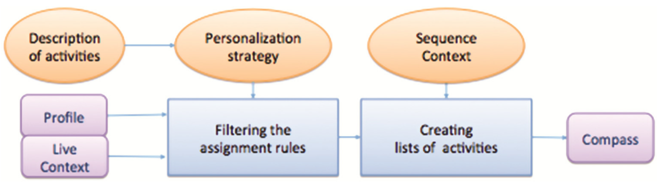



**Fig. 1.** Operating process of the PERSUA2$_{MOOC}$ model.

the profile, and the live context of use. The educational team defines the other elements: the pedagogical strategy, the description of activities and the sequence context of use. The two main steps of this process allow to obtain automatically lists of personalized activities for each learner, by using these five inputs. These activities are ultimately proposed to the learner using a "compass".

In the PERSUA2 model, the teacher specifies all parts of the learner profile he/she wishes to use to characterize learners. In order to facilitate the work of the educational teams of MOOCs, we propose in $\text{PERSUA2}_{\text{MOOC}}$ to structure the learner profile into 5 categories. The *resourcesInteractions* section contains quantitative information about the use of the MOOC resources by the learner. Thus, the educational team can include in this section indicators to know, for a given resource (a video for example), how many times the student has visited it, or the total time dedicated to this resource. The *moocInteractions* section provides a more global vision, and concerns interactions with the MOOC platform in general. It offers quantitative indicators to know for example how the learner organizes its work: days and times when he/she is most active, larger period of absence, etc. Regarding the *behavior* section, it contains essentially qualitative indicators enabling to obtain more advanced information about the learner behavior, as his/her way of learning or his/her participation on the forum. Indicators of the *knowledge* section characterize the knowledge and skills of the learner in the MOOC s/he is following. The educational team defines these indicators according to their course. All these sections contain indicators that will be calculated from the traces collected on the MOOC platform. The *learnerInformation* section contains information that cannot be derived from the learner's traces, such as demographics, or his/her learning objectives by participating in the MOOC. These indicators will be filled in through questions asked directly to the learner.

In the PERSUA2 model, the learner profile is the only structure containing information about the learner. Even if elements of the profile can be very diverse, they are only updated after the achievement of an activity, in order to reflect a "stable" view of knowledge, skills and behavior of the learner. However, in MOOCs, other relevant data are important to provide the learner with activities adapted to him/her. We then added a **live context** consisting of two parts. The *learnerLiveContext* part concerns everything that characterizes his/her learning context: *e.g.* the equipment s/he uses to connect to the platform, the available bandwidth. The *environmentContext* part describes some properties of the platform and the MOOC at a particular time when the learner logs, such as the number of learners connected to the MOOC, or the number of teachers available to answer questions.

As in the PERSUA2 model, a **personalization strategy** is a set of "IF-THEN-ELSE" rules. The conditions of these rules are constraints on the values of the elements of the learner profile. The consequences are lists of activities (constrained by some parameters), which should be proposed to the learner if he/she satisfies (or not) these conditions. The educational team of each MOOC will define these rules.

The educational team also defines a **sequence context** specifying global constraints on the sequence: minimum and maximum number of activities, (theoretical) time required to achieve the sequence, etc. Compared to the PERSUA2 model, a new element is added to the sequence context: the ability to restrict the use of some activities to some sequences. For each new sequence of the MOOC, the team will decide what are the personalization strategy and the sequence context that should be used by the system in order to personalize the MOOC. The pedagogical strategy may be global for the MOOC and associated each time with a different sequence context, or conversely, each sequence may have its own educational strategy and context.

For each learner, a first process determines which rules of the pedagogical strategy should apply. The **algorithm** used thus takes as input a pedagogical rule, the profile of each learner and the live context, and evaluates the IF part of the rule (by analyzing the constraints that constitute it and the values contained in the profile and the live context). This clears whether the condition is true for the learner, and thus whether the THEN or ELSE part of the rule should be applied for the learner. Finally, based on these rules, lists of activities are generated for each learner, using directly the THEN or ELSE parts of the rules, and taking into account the global constraints of the sequence context (e.g. the scheduled working time).

We have identified another need for adaptation, concerning the outputs of the PERSUA2 operating process. Indeed, the purpose of this model in the context of ITSs is to directly configure these systems, in order to lead the learner to perform activities that are obtained from the personalization strategy of their teacher. Yet the MOOCs are part of a different philosophy: every learner has access to all resources of a course freely and without restriction. A personalization solution that would require the learner to consult some resources, making others not available, would certainly be badly perceived. Therefore, we believe that any personalization solution within the context of MOOCs should use recommendations, and not constraints: you must tell the student what courses and what activities appear to be the most suitable for him/her, but without preventing him/her to consult other resources. This is implemented in our model by editing a **compass**, which is a list of links to resources and activities that the student is invited to consult and achieve first.

## 3 Conclusion

This model was instantiated for the FOVEA MOOC [3], and its operating process implemented as a web application. We were able to experiment all the components of our model with the authors of the MOOC, and check that their instantiation was possible, enabling finally to define a complete educational strategy and to generate lists of personalized activities for each learner. We also checked that our model enables to describe the activities proposed in the MOOC platforms Coursera, edX and Udacity.

Our approach places the teaching team at the center of the process of customization, enabling the adaptation of the MOOC to each learner, this personalization inte-grating all the functionalities offered on a MOOC platform. An important perspective of this work will be to provide the teaching team with feedback of learners' activities, in order to judge the effectiveness of their pedagogical strategy.

## References

1. Lefevre, M., Jean-Daubias, S., Guin, N.: An approach for unified personalization of learning. In: International Workshop on PALE – UMAP, pp. 5–10 (2012)
2. Lefevre, M., Jean-Daubias, S., Guin, N.: Supporting acquisition of knowledge to personalize interactive learning environments through a meta-model. In: ICCE (2009)
3. FOVEA (2014). http://anatomie3d.univ-lyon1.fr/

# Increasing Pupils' Motivation on Elementary School with Help of Social Networks and Mobile Technologies

Václav Maněna, Roman Dostál, Štěpán Hubálovský(✉), and Marie Hubálovská

Univerzity of Hradec Kralove, Rokitanskeho 62, Hradec Kralove, Czech Republic
{vaclav.manena,roman.dostal,stepan.hubalovsky, marie.hubalovska}@uhk.cz

**Abstract.** The authors focus on the way of using a combination of social networking and mobile technology at an elementary school in order to increase students´ motivation. The paper summarizes the results of a survey, which was attended by respondents from elementary schools in Hradec Králové, Czech Republic. The paper also presents methods of using mobile technologies and social networks in elementary school in order to involve students into education and increase their motivation.

**Keywords:** M-learning · Mobile technologies · Social networks · Mobile phone · Tablet

## 1 Introduction

The popularity of social networking and mobile technology is growing. The situation in the Czech Republic is similar as in other European countries as well as in the USA [1]. Social networking and mobile technology are essentially ubiquitous and greatly affect the lives of people across all age categories. Not so long ago, the question whether social networks and mobile technology can be included in teaching at primary and secondary schools has been solved by educational circles. Today, such a question does not make sense, because the mentioned technology are already implemented in schools. The first pupils have their own technologies. Penetration of mobile technology in education is also supported by the Ministry of Education. One example is the challenge ESF no. 51, which aimed to provide schools tablets and touch-enabled devices. The world is changing, and schools simply cannot ignore it. The current situation regarding the extension and how to use social networking and mobile technology among pupils at primary schools in the Czech Republic has to be known. The issue of the use of social networking and mobile technology for increasing the motivation is described by a number of experts over the world, see e.g. [2]. A combination of mobile technology and social networks is normal for children and young people. According to the results of a study [3] Mobile phones are used by 43 % of children aged 3–18 years and Facebook is used by 41 % of teenagers.

The most popular social network in the Czech Republic is Facebook. The results of the research "Czech children and Facebook 2015" [4] are summarized in [5]. Besides other the following facts are presented: "*90 % of Czech children over 13 years have*


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 610–613, 2016.
DOI: 10.1007/978-3-319-45153-4_74

*a Facebook account. Alarming is that more than half of Internet users under 13 years of age has Facebook as well. It contradicts the rules of this social networks. Overall, the Facebook account has 81 % of Czech children, 16 % has two to three accounts at once, and 12 % admitted that they have set up a fake account.*"

Learning using mobile technology (known as m-learning) is currently worldwide increasing. The potential of using of these technologies in education grow up with improvement and availability of the devices [6]. It is clear that social networking and mobile technology occupy an important role in the lives of pupils of elementary schools. The actual situation of the basic schools in Hradec Kralove is mapped in our pilot research described below.

# 2 Pilot Research

Social networks and mobile technology are important factors in the lives of contemporary schoolchildren and young people in general. The combination of social networks and mobile technology provides significant potential for use in education at all. From above mentioned reasons, we focused in our research on the use of social networks and mobile technologies by pupils of elementary as well as secondary schools.

## 2.1 The Research Goals and Methodology

The aim of the research was to identify the use of social networks and mobile technologies by pupils of basic schools. The research question is "How mobile technology and social networks are used by pupils". Sub-objectives of the research are as follows:

- Which types of mobile technology pupils used?
- What social networks pupils used?
- How often pupils use the social networks?
- Where pupils access to social networks?
- What devices pupils use at school to access social networks?

Based on the above mentioned goals the technique of non-standardized questionnaire with closed answers were used. The overall response of questionnaire was 83 %.

## 2.2 The Research Sample

The research sample consisted of 312 respondents – pupils of primary school in Hradec Kralove. The response was gained by an anonymous electronic questionnaire. The gender distribution of the respondents is 136 men and 176 women.

The age distribution of respondents was intentionally chosen so that the group of pupils under the age of 13 years is covered too. This group of pupils is interesting for two reasons. First and foremost, they are the users who use the mobile technology at school as well as outside the school environment. Another reason is the fact that most of social networks sites have rules of minimum of 13 years of age. Like the authors of the research [1], also we find that 6 % of children under 13 years of age are using social networks, even though it is contrary with the rules of use. Although we observed ratio smaller than that published by authors of the above mentioned research [1], it is a significant percentage too. Furthermore, we assume that the popularity of social networks for children under 13 years of ages will increase. This fact has been responded e.g. by Google. Although most services are not allowed to users under 13 years of age [7], if the school is using the Google Apps for Education, the administrator can enable the use of these services (e.g. Gmail or Google+) for younger pupils. These accounts, however, can only be used by pupils within the domain of school.

The authors [4] found that approximately a third of children spend more than three hours a day on the social networking site Facebook. Similar results have been reached in our research – 10 % of respondents said that social networks generally spend 5–6 h a day and nearly 10 % of respondents spend on these networks more than 6 h a day.

The most popular social networks are Facebook (94 %) and Instagram (55 %), which we had expected. We were surprised by the relatively high proportion of the social network Twitter (20 %), which is higher than Google+ (14 %).

The most popular devices for access to social networks are mobile phones (90 %) followed by notebooks (61 %). Significant is the proportion of tablets (29 %) and desktop computers (37 %). Mobile technologies are used by pupils in conjunction with the social networks already more than traditional desktops and laptops. In our research, we did not distinguish notebooks and convertible devices because respondents often fail to recognize these two categories. The current convertible devices can be classified as mobile technology, thus the overall portion of mobile devices has increased.

Most of the pupils (89 %) connect to social networks via mobile technologies at schools, over 37 % of pupils are connected through computers in computer labs.

## 3 Conclusion

The results of the pilot research confirmed that mobile technology and social networks are used by pupils extensively not only in a leisure, but also in school. The combination of mobile devices (laptops, tablets), and social networks can be logically be used as suitable tool for making learning attractive and can caused increase of pupil's motivation. The research investigation indicates that the most popular social networks are among primary schools pupils Facebook and Instagram. So we have obtain the similar conclusions as the authors of national study [4]. In the next stage of our research we will focus on ways of use the combination of mobile devices and social networks in elementary schools and grammar schools. We will focus on the following options of use:

- Documentation of excursions, trips and projects. Pictures will be labeled by pupils with predefined hashtags (Facebook, Instagram).
- Photographic record of experiment in a school laboratory or classroom. We will focus mainly on labor practices in workshops and laboratories (Facebook, Instagram).
- Project learning outside – pupils will be tasked with creating pictures of buildings of a certain architecture style in their place (Facebook, Instagram).
- Preparing the project and communication within the project using a Facebook group.

**Acknowledgement.** The paper has been supported by Specific Research Projects of Faculty of Science and Faculty of Education, University of Hradec Kralove.

## References

1. Manena, V., Rybenska, K., Špilka, R.: Research of mobile technologies used by students in primary school as a basis for the creation of teaching materials. In: International Conference on AdvancedEducational Technology and InformationEngineering, AETIE 2015, Beijiing, People's Republic of China, pp. 330–335 (2015). ISBN: 978-1-60595-245-1
2. Trajkovic, V., Vasileva, M., Karbeva, S., Videnovic, M.: Increasing students' motivation by using social networks in and out of the classrooms. ICT Innovations 2012: Web Proceedings (2012). https://www.academia.edu/7628799/Increasing_students_motivation_by_using_social_networks_in_and_out_of_the_classrooms. ISSN: 1857-7288
3. Ang, Ch.: How to use smartphones in the classroom: up-to-datestatistics. e-Learning (2015). http://www.ispringsolutions.com, http://www.ispringsolutions.com/blog/how-to-use-smartphones-in-the-classroom-up-to-date-statistics
4. Centrum prevence rizikové virtuální komunikace, Pedagogická fakulta Univerzity Palackého v Olomouci. České děti a Facebook (2015). http://www.e-bezpeci.cz/facebook2015
5. Potůček, J.: Třetina českých dětí tráví na Facebooku víc než tři hodiny denně. Internet a PC (2015). www.novinky.cz, http://www.novinky.cz/internet-a-pc/bezpecnost/386965-tretina-ceskych-deti-travi-na-facebooku-vic-nez-tri-hodiny-denne.html
6. Burgerová, J., Maněnová, M., Adamkovičová, M.: New perspectives on communication and co-operation in e-learning. Praha: ExtraSYSTEM (2013). ISBN: 978-80-87570-16-6
7. GOOGLE. Age requirements on Google Accounts (2016). https://support.google.com/accounts/answer/1350409?hl=en

# Understanding Collective Behavior of Learning Design Communities

Konstantinos Michos(✉) and Davinia Hernández-Leo

ICT Department, Universitat Pompeu Fabra, Barcelona, Spain
{kostas.michos,davinia.hernandez}@upf.edu

**Abstract.** Social computing enables collective actions and social interaction with rich exchange of information. In the context of educators' networks where they create and share learning design artifacts, little is known about their collective behavior. Learning design tooling focuses on supporting educators (learning designers) in making explicit their design ideas and encourages the development of "learning design communities". Building on social elements, this paper aims to identify the level of engagement and interactions in three communities using an Integrated Learning Design Environment (ILDE). The results show a relationship between the exploration of different artifacts and creation of content in all the three communities confirming that browsing influence the community's outcomes. Different patterns of interaction suggest specific impact of language and length of support for users.

**Keywords:** Learning design · Communities of educators · Collective behavior · Social network analysis

## 1 Introduction

The current discussion on teaching and learning with the use of Information and Communication Technologies suggests the reformulation of teaching practices and alignment of ongoing pedagogies with the changes, advantages and effective adoption of emerging technologies. In this direction, the notion of "openness" in teaching with Web 2.0 environments and the movement from individual to collective practices when teachers are designing learning scenarios constitute new paradigms of knowledge exchange. Learning Design is the field that studies the art and science of *designing* meaningful and effective scenarios *for learning* and proposes tools to support the design process by enabling their explicit representation in sharable formats [1, 2]. The artifacts reflecting the designed learning scenarios are generally called *learning designs*.

Social computing enables collective action and online social interaction with rich multimedia exchanges and evolution of aggregate knowledge [3]. Significantly, social network environments are highly based on user participation and contribution behavior to benefit from collective intelligence. Existing research has studied participation behavior in diverse types of social networks [4], including teacher's communities [5, 6]. However, in the context of educators' networks whose aim is creating the


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 614–617, 2016.
DOI: 10.1007/978-3-319-45153-4_75

best possible learning designs for their particular contexts, very few studies provide results between different communities on the collective usage and contribution behavior of the users.

In this paper we focus on the online activities undertaken by three groups of educators using three separate installations of the ILDE community environment [7]. ILDE supports the development of "learning design" communities in which members are able to share and co-create multiple types of learning designs. The research question investigates and compares the usage and contribution behavior of the three learning design communities (a multilingual training community-ILDE-MOOC1, a monolingual training community-ILDE-MOOC2 and an open learning design community-ILDE-Demo). The analysis focuses on identifying common patterns and differences in four user's actions: creation, modification, exploration of learning designs and comments. Data used is extracted from log files automatically collected by ILDE. Correlation analysis examines the relationship between exploration of content and contribution behavior and social network analysis aims to identify the network structure of these communities.

## 2 Results

In each community we observed the number of learning designs viewed by user (passive participation) considering the users with at least one view and their overall creation, number of modified learning designs and comments (active participation). The aim was to identify the levels of engagement and analyze if exploration of different artifacts was related with explicit user's actions. In all the communities there was a positive relationship between viewing and modification and between viewing and creation of learning designs (see Table 1).

**Table 1.** Descriptive statistics and Spearman's correlation matrix in the three communities

| | ILDE-MOOC1(n = 315) | | ILDE-MOOC2(n = 359) | | ILDE-Demo(n = 289) | |
|---|---|---|---|---|---|---|
| | M(SD) | 1 | M(SD) | 1 | M(SD) | 1 |
| 1. Views | 33.79(44.69) | | 25.81(40.37) | | 8.36(17.04) | |
| 2. Edits | 4.79(5.09) | .827* | 3.34(4.15) | .753* | 1.36(4.79) | .434* |
| 3. LdS | 5.62(5.13) | .818* | 7.43(6.36) | .553* | 3.15(8.03) | .426* |

*p < .01, LdS (Learning design Solution, in ILDE/LdShake terminology) = Total created learning designs per user, Views = Total number of LdS viewed per user, Edits = Total number of LdS edited by user.

Although in the open-environment (ILDE-Demo) this was identified in a lower level since the other two communities were running within a MOOC training course [8], this relation was present. These results propose that users do check examples of learning designs when they create new artifacts and that learning designers in a community platform can influence each other on the way they design. To further explore the interaction patterns between different users in the communities using the ILDE environment and identify how users influence each other we followed a social network analysis approach. We constructed in each community two directed, weighted networks based on the following relationships: a views network which was representing that one user (node x) viewed the learning design (edge) of another user (node y), a comments network which was representing that one user (node x) commented the learning design (edge) of another user (node y). Table 2 presents network statistics of the observed networks in the three different communities.

**Table 2.** Statistics of the different networks

| | Views network | | | Comments network | | |
|---|---|---|---|---|---|---|
| | MOOC1 | MOOC2 | Demo | MOOC1 | MOOC2 | Demo |
| Nodes | 310 | 264 | 229 | 154 | 191 | 22 |
| Edges | 5729 | 1134 | 1050 | 376 | 481 | 36 |
| Degree | 101.31 | 29.27 | 16.17 | 2.98 | 3.49 | 2.22 |
| Modularity | .12 | .35 | .35 | .42 | .64 | .43 |

We can see in the views network that in the monolingual community (MOOC1) more users (nodes) compared to the multilingual community (MOOC2) browsed the designs of others (edges). In the multilingual community (MOOC2), participants concentrated in browsing mostly designs created in the language they understand best and thus created more clusters (higher modularity) while in the first MOOC all participants explored designs (only in English) created by the whole community. In contrast, in the comments network of the monolingual community (MOOC1) fewer users commented the learning designs of others. This suggests that the familiarity of users with the language can influence the commenting behavior and the frequency of messages between them. Additional differences like domain of expertise or familiarity with technology may also influence their interactions. In the open community (Demo) the network was developed through a three year period of time, and users periodically contributed with creation of learning designs and comments to them. Views network shows that fewer users, than in the others communities, explored learning designs created by others. However, despite the use of ILDE was self-organized or free use in this case, we observe an arguably relevant interest of users in browsing designs in the community. In terms of communication, the community showed a similar behavior (less clusters) as the first MOOC because the interaction occurred in English. Although comments were few, the fact that some users knew each other and had a common goal (e.g., project members designing training workshops) created a dense network and purposeful interactions.

## 3 Conclusion

Sharable formats of learning designs serve as representation of designers' thinking about effective learning in their contexts and as means of communication between educational practitioners. Our results suggest that visibility for popular users and designs, monitoring of users' participation and identification of high quality artifacts in such communities may add additional value in the way users explore and contribute. Scaling sharing of teaching practices in community environments enables the identification of patterns shedding light about how teachers are designing being inspired by other educators' ideas and based on diverse pedagogical approaches. In this paper we touched one aspect of collective behavior analysis in the usage of a social online platform for learning design in three particular communities. Further studies should consider properties of the designs (learning design representations and tools used, qualitative analysis of its content) and whether created designs have been created from scratch or refine copies of reused designs available in the community.

**Acknowledgments.** This research is partly funded by RecerCaixa and the Spanish Ministry of Economy and Competitiveness under RESET (TIN2014-53199-C3-3-R) and the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502). DHL is a Serra Hunter Fellow.

## References

1. Mor, Y., Craft, B., Hernández-Leo, D.: The art and science of learning design: editoral. Res. Learn. Technol. **21**, 1–8 (2013)
2. Lockyer, L., Bennett, S., Agostinho, S., Harper, B.: Handbook of Research on Learning Design and Learning Objects: Issues, Applications, and Technologies (2 volumes). IGI Global, Hershey (2009)
3. Parameswaran, M., Whinston, A.B.: Social computing: an overview. Commun. Assoc. Inf. Syst. **19**(1), 37, 762–780 (2007)
4. Jiang, J., Wilson, C., Wang, X., Sha, W., Huang, P., Dai, Y., Zhao, B.Y.: Understanding latent interactions in online social networks. ACM Trans. Web (TWEB) **7**(4), 18 (2013)
5. Recker, M.M., Yuan, M., Ye, L.: CrowdTeaching: supporting teachers as designers in collective intelligence communities. Int. Rev. Res. Open Distrib. Learn. **15**(4), 138–160 (2014)
6. Pynoo, B., van Braak, J.: Predicting teachers' generative and receptive use of an educational portal by intention, attitude and self-reported use. Comput. Hum. Behav. **34**, 315–322 (2014)
7. Hernández-Leo, D., Asensio-Pérez, J.I., Derntl, M., Prieto, L.P., Chacón, J.: ILDE: community environment for conceptualizing, authoring and deploying learning activities. In: Proceedings of 9th European Conference on Technology Enhanced Learning, EC-TEL 2014, September 2014, Graz, Austria, pp. 490–493 (2014)
8. Garreta-Domingo, M., Hernández-Leo, D., Mor, Y., Sloep, P.: Teachers' perceptions about the HANDSON MOOC: a learning design studio case. In: Proceedings of 10th European Conference on Technology Enhanced Learning, EC-TEL 2015, September 2015, Toledo, Spain, pp. 420–427 (2015)

# A Value Model for MOOCs

Yishay Mor[1(✉)], Marco Kalz[2], and Jonatan Castano-Munoz[3]

[1] PAU Education, Barcelona, Spain
{yishay.mor,muriel.garreta}@paueducation.com
[2] Open University Netherlands, Heerlen, Netherlands
marco.kalz@ou.nl
[3] Institute for Prospective Technological Studies, Seville, Spain
Jonatan.CASTANO-MUNOZ@ec.europa.eu

**Abstract.** Massive Open Online Courses (MOOCs) are changing the educational field, challenging traditional institutional strategies and recognition schemes and opening up new opportunities for learners and educators both from within and outside formal education. However, while the potential benefits and risks of the MOOCs have been discussed by scientists and policy makers, the corresponding empirical data is scarce. What's more, the evidence that is available is usually restricted to a single course or single provider.

MOOCKnowledge (http://moocknowledge.eu/), funded by the European Commission's Institute for Prospective Technological Studies (IPTS), aims to facilitate a shared understanding of the value and efficacy of MOOCs by developing a set of analysis tools and applying them to a wide range of MOOCs.

The most powerful outcome of the project would be the possibility to correlate different dimensions of MOOC production, execution, and learners experience. For example, identifying links between financial investment, learning design, and learner outcomes. To do this, we must first develop a conceptual model of the factors which determine or contribute to the value of a MOOC.

**Keywords:** MOOCs · Learning design · Evaluation · Cost · Value

## 1 Introduction

Massive Open Online Courses (MOOCs) are changing the educational field, challenging traditional institutional strategies and recognition schemes and opening up new opportunities for learners and educators both from within and outside formal education. However, while the potential benefits and risks of the MOOCs have been discussed by scientists and policy makers, the corresponding empirical data is scarce. What's more, the evidence that is available is usually restricted to a single course or single provider.

MOOCKnowledge (http://moocknowledge.eu/), funded by the European Commission's Institute for Prospective Technological Studies (IPTS), aims to


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 618–621, 2016.
DOI: 10.1007/978-3-319-45153-4_76

facilitate a shared understanding of the value and efficacy of MOOCs by developing a set of analysis tools and applying them to a wide range of MOOCs. We have already developed a three-survey (pre- post- and follow-up) tool, which compares learner's expectations and intentions to their perceptions and the observable evidence of their actual benefits from the MOOC. We are in the process of developing a design analysis tool, which will include a set of rubrics to evaluate a MOOC's design - from it's overall structure to the details of specific media assets.

The most powerful outcome of the project would be the possibility to correlate different dimensions of MOOC production, execution, and learners experience. For example, identifying links between financial investment, learning design, and learner outcomes. To do this, we must first develop a conceptual model of the factors which determine or contribute to the value of a MOOC.

This paper presents our current version of this model, and invites the community to engage with it. The model was developed through a combination of desk research and expert review.

The mindmap of the model is available at:

https://atlas.mindmup.com/2016/03/f8dfb450cc3101338f4d19e3b2bc43d4/mooc_value/index.html

A version of this paper open for commenting is available at:

https://docs.google.com/document/d/1oVfZ2WGLklJNfRissjdOkkbK8A7yNZvPSJrZaKb780o/edit.

## 2 Method

The Model is being developed through iterations of desk research and expert review. We started by looking at the typical parameters used to list/catalogue MOOCs. We then expanded it to include factors that are often neglected, such as the institutional/individual motivations for creating a MOOC. This model was presented to experts at the RIDE conference and online, and was updated based on their feedback.

This process of calibrating literature, common practice and expert review is ongoing. Our presentation at EC TEL will be another major iteration.

## 3 The Model

The model currently has nine sections (Fig. 1): meta-data, cost, drivers, benefits, risks, regulatory framework, learner profile, efficacy, and figures. This model is not a taxonomy, it is simply a guide for identifying the factors that play a potential role in determining the value of a MOOC, and a starting point for exploring correlations and dependencies between these.

**Meta-data.** Parameters typically used to index or catalogue a MOOC.
The Meta-data parameters are:

**Topic** e.g. Java programming, web design, art history

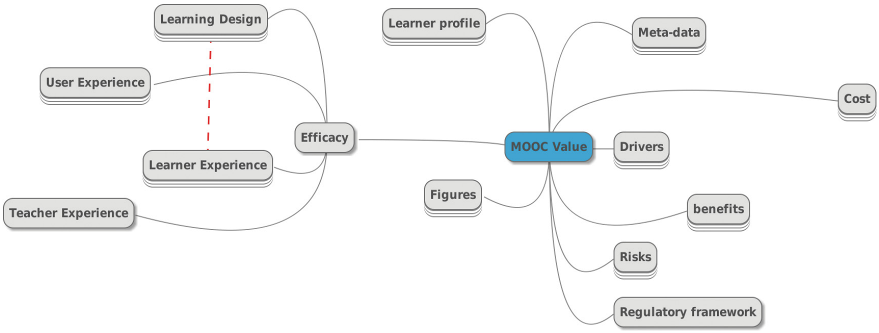



**Fig. 1.** Overview of the model

**Level/type** educational institution type (K12, Higher education, professional development) and level (Introductory, Intermediate, Advanced)
**Title** course title
**Timing** start date and length (in weeks)
**prerequisites**
**Institution** (and faculty) providing the MOOC
**Delivery mode** Scheduled/self-paced
**Platform** e.g. Coursera, EdX, FutureLearn
**Language** e.g. English, Spanish, Arabic
**Effort** required by the student, in hours per week
**Certification** types of certificates offered (including ECTS)
**Target audience** profile of expected participants
**Size** expected number of students, including possible caps on size.

**Cost.** The various factors that determine the cost of designing, developing and delivering a MOOC.
The cost factors we identified are:

**Design and planning** Research Design Prototyping
**Production** content production, including text, media (graphics, animations, games, and video), markup and media integration on the platform, assignments and assessments, and content maintenance (updating the content from time to time).
**Quality Assurance**
**Marketing**
**Hosting** either on an established platform or on a self-hosted/externally hosted VLE.
**Presentation** the actual "running" cost, including the time of faculty, facilitators/moderators, and tech support
**Assessment** in particular procturing and marking
**Certification** mainly the platform fees
**Evaluation** from audit of the MOOC design pre-presentation to the analysis of the feedback and analytics.

**Drivers.** Drivers are the factors that motivate institutions and individuals to offer MOOCs.

**Benefits.** Benefits are the actual positive outcomes that a MOOC may have for the individuals attending them, the institutions and individuals providing them, and society as a whole.

**Risks.** By contrast to benefits, risks enumerate the possible negative consequences of MOOCs.

**Regulatory Framework.** MOOCs (as all educational instruments) are governed by national and international regulatory frameworks, which enable and delimit their potential impact and dictate some of the practices of their providers and participants.

**Learner Profile.** The Learner Profile includes the characteristics of the MOOC participants that can be inferred from questionnaires or observations.

**Efficacy and Learning Design.** Efficacy refers to the predicted capacity of the MOOC to achieve its aims.

## 4 Summary

We have presented a proposed model for the value of a MOOC. Although this model admittedly still requires refinement and validation, we believe it is nevertheless of value for whoever is considering developing a MOOC, or needs to make policy decisions regarding MOOCs.

The most significant value of this model will be as a research tool for exploring the interaction and dependencies between the different dimensions. For example, to answer questions such as:

– what is more cost-effective (in terms of learner benefits) - investment in video quality or in the quality of assignments and assessment?
– are certain media types more appealing to specific learner profiles?
– what are the hidden costs, benefits and risks that need to be considered when evaluating a proposal for producing a new MOOC?

We plan to collect data along these dimensions and make it available under an open licence, to facilitate research of such questions and others.

# Framework for Learner Assessment in Learning Games

Mathieu Muratet[1,2(✉)], Amel Yessad[1], and Thibault Carron[1,3]

[1] Sorbonne Universités, UPMC Univ. Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu, 75005 Paris, France
{mathieu.muratet,amel.yessad,thibault.carron}@lip6.fr
[2] INS HEA, 58-60 Avenue des Landes, 92150 Suresnes, France
[3] Université Savoie Mont Blanc, 73376 Le Bourget-du-Lac, France

**Abstract.** Learner assessment in learning games (LG) is an interesting research area for both academia and industry. The play traces resulting from the learner's activity in LGs with large state spaces and a large amount of free interactions, are hard to analyze and to interpret by teachers. In this paper, we present a framework to assist the building of an expert's solving process that is the base of the algorithm that analyzes player's traces and generates pedagogical labels about the learner's behavior.

**Keywords:** Learning game · Behavioral model · Petri net

## 1 Introduction and Positioning

Learner assessment is considered as a key issue in Technology-Enhanced Learning (TEL). Learners are not all alike, and it is interesting to assess the behavior of each learner who uses the system in order to implement adapted scenarios and provide feedback. Our approach aims (1) to be seamlessly integrated with a LG, rather than presented as a separate artificial assessment disconnected from the nature of the task and (2) to compare a learner's behavior with experts' solving model.

Our research focuses on learning games which simulate process (physical, industrial, business, etc.). In this kind of complex systems with a large amount of freedom in interaction, it is hard to model game actions and experts' solving processes in order to understand and to analyze students' activity. Thus, our objective is to assist designers in building a model of the experts' solving and to compare it with the learner's solving in order to generate a description of learner's behavior, readable by teachers and designers.

Several research had already considered the issue of automatic assessment of learners by analyzing play traces. Thus, in [1], the authors propose a methodology for extracting conceptual features from student's log data using a two-dimensional context-free grammar. This contribution is focused on puzzle games


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 622–626, 2016.
DOI: 10.1007/978-3-319-45153-4_77

like *RumbleBlocks*[1] or *Refraction*[2]. Other research used Petri nets to describe experts' solving of "Case study" games and proposed an algorithm to label learners' actions [2]. However, this algorithm is adapted to unique type of games (case studies) and is not at all suitable for learning games with large state spaces and a large amount of freedom in interaction.

Our approach shares the same objective with these approaches but aims to propose a scalable and generalizable framework giving more accurate pedagogical information about the learner's behavior. The pedagogical labels defined are based on the comparison between the learner's behavior and the expert's solving of a game level. Figure 4 depicts the global architecture of the assessment framework. In this paper, we focus on the workflow that assists designers to built the model of experts' solving.

## 2 Assistive Workflow to Build the Expert's Solving Process

A key point in our methodology is to model the experts' solving process by a executable model and to assess the learners actions by comparing them with this model. Like [2], we choose to use Petri nets which is a powerful modeling formalism in computer science, system engineering, and many other disciplines (see [3] for details on Petri nets). Petri net combines a well-defined mathematical theory with a graphical representation of the system's behavior. The theoretical aspect of Petri nets allows precise modeling and analysis of system behavior [4]. However, modeling a complex simulation game with Petri nets is a difficult task both for game designers and experts. The main difficult task is to assure consistency between Petri net modeling and game simulation. In our framework, we propose an assistive workflow to semi-automate the Petri net building.

### 2.1 Example: The Frozen Door

We illustrate our contribution with a simple example of a frozen door. Figure 1 depicts a simple Petri net of a door that the player can open or close (in the initial marking, the door is closed). If the door is connected to other game objects like a key, then this Petri net is extended in order to match with the simulation (cf. Fig. 2). In this second Petri net, the door is locked and the key is required to open it. We also added a boiler into this game level that the user has to turn on in order to solve the level. In the initial marking of this Petri net, the door is closed, the boiler is turned off and the key is in the inventory.

In order to implement the automatic learner assessment, we construct two Petri nets semi-automatically. The first Petri net is called "Full Petri net" (FullPn) and includes all actions that learners can perform in the game. A FullPn

[1] RumbleBlocks: http://rumbleblocks.etc.cmu.edu/, accessed April 4, 2016.

[2] Refraction: http://games.cs.washington.edu/refraction/refraction.html, accessed April 4, 2016.

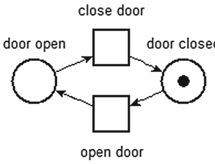



**Fig. 1.** Petri net of a door that the player can open or close.

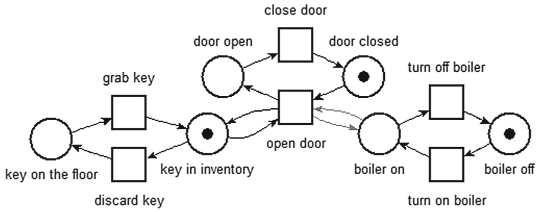



**Fig. 2.** Full Petri net of a frozen door. Only gray arcs are manually added. The other places, transitions and arcs of this figure are built automatically.

models game simulation and its marking depicts the state of the simulation. The second Petri net, called "Filtered Petri net" (FilteredPn), is a part of the FullPn and includes only actions used by experts to solve the current level. It embeds the expert's action sequences allowing to solve the game level.

The building of the FullPn is a challenging task due to the high number of actions that the learner can perform in each state of the game. This building process has to be automatic or at least semi-automatic. In our work, the semi-automatic building of Petri nets is based on the definition of game objects and their properties. Each game object is described by the actions the user can perform on. For instance, in the role playing game we used to test our framework, the object "the door" can be opened or closed, the object "the key" can be grabbed or discarded and the object "the boiler" can be turned on or turned off. The objects and their properties are described in a user friendly editor called Tiled[3]. We have implemented a complex XSLT transformation to build a Petri net from Tiled game object descriptions (for instance, Fig. 2 is the result of this transformation for a simple level, only the gray arcs were added manually). We can summarize the benefits of this transformation process by the following points: (1) the transformation process is weak-dependent on the game level because once the game objects are described in Tiled, the transformation is not changed and the game object can be reused in several levels; (2) the effort of developing the transformation is performed once, while we can use it many times, at each game level; (3) the transformation generates less errors than the manual building of Petri nets; and (4) the Petri net building has to be validated/completed by LG designers, but the validation task is less time-consuming and less complicated than building a Petri net from scratch.

Once we have built the FullPn, we filter it by removing transitions that are not used by experts, in order to build the FilteredPn. In the example of the frozen door (cf. Fig. 2), the objective is that the player opens the frozen door. The expert's solving consists in turning on the boiler and opening the door with the key. Formally, it corresponds to fire, in sequence, the transition "turn on boiler" and then the transition "open door". Figure 3 represents the FilteredPn

[3] Tiled: http://www.mapeditor.org/, accessed April 4, 2016.

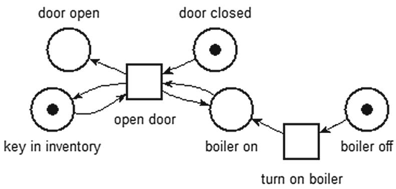



**Fig. 3.** Filtered Petri net of the frozen door. Only transitions (actions) used by the expert are kept from the Full Petri net, here, turning on the boiler and opening the door.

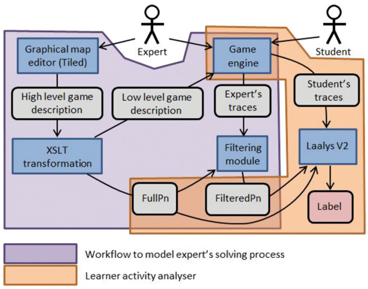



**Fig. 4.** Global architecture of the assessment framework.

that results from the filtering of the FullPn of Fig. 2. Once the FilteredPn is built, we compute its reachability graph that serves us to analyze the learners' actions.

### 2.2 Workflow Overview

As depicted in Fig. 4, the designers start by using a user-friendly graphical tool to build a game level. From this high level game description, we use an XSLT transformation to build two files: (1) a low level game description that is compatible with the game engine and (2) a Petri net that describes the game simulation (the FullPn). An expert can play this new level (several times if several solutions are available) and the game engine traces the expert's actions. These traces are used to filter the FullPn and build the FilteredPn. We notice that a non-expert's trace could be used to filter the FullPn, for example an original and correct solving made by a learner positively assessed by teacher can be added to the expert's traces to enlarge the FilteredPn.

Once the FullPn and the FilteredPn are generated, they can be validated or completed by expert/designer manually in order to include constraints not configurable with the graphical editing tool. Then, this validated FilteredPn is used by the labeling algorithm to label learners' actions pedagogically.

## 3 Conclusion

The work presented in this paper deals with the assessment of learners' behavior in learning games. This paper focuses on a workflow to help the designers to model expert's solving with Petri nets. We illustrated the methodology with the simple and pedagogical example. This framework was used to design 18 levels of the LG "Les Cristaux d'Ehere" and produced full and filtered Petri nets of these levels automatically. On average, the Petri nets produced by this way are composed of 22 places, 19 transitions and 59 arcs. The most complex Petri net produces more than 127 000 game states.

## References

1. Harpstead, E., MacLellan, C.J., Koedinger, K.R., Aleven, V., Dow, S.P., Myers, B.A.: Investigating the solution space of an open-ended educational game using conceptual feature extraction. In: Proceedings of the 6th International Conference on Educational Data Mining, Memphis, Tennessee, USA, 6–9 July 2013, pp. 51–58 (2013)
2. Yessad, A., Thomas, P., Capdevila, B., Labat, J.-M.: Using the petri nets for the learner assessment in serious games. In: Luo, X., Spaniol, M., Wang, L., Li, Q., Nejdl, W., Zhang, W. (eds.) ICWL 2010. LNCS, vol. 6483, pp. 339–348. Springer, Heidelberg (2010)
3. Peterson, J.L.: Petri Net Theory and Modeling of Systems. Prentice Hall, Reading (1981)
4. Wang, J.: Petri Nets for Dynamic Event-Driven System Modeling. Computer & Information Science Series. Chapman & Hall/CRC (2007)

# A Bayesian Network for the Cognitive Diagnosis of Deductive Reasoning

Ange Tato(✉), Roger Nkambou, Janie Brisson, Clauvice Kenfack, Serge Robert, and Pamela Kissok

Université du Québec à Montréal, Montréal, Canada
angetato@gmail.com, nkambou.roger@uqam.ca

**Abstract.** In our previous works, we presented Logic-Muse as an ITS that helps improve logical reasoning skills in multiple contexts. All its three main components (the learner, tutor and expert models) have been developed while relying on the help of experts and on important work in the field of reasoning and computer science. The main purpose of this paper is to present and assess the Bayesian Network (that allows real time diagnosis and modeling of the learner's state of knowledge) implemented in the learner component. We demonstrate the prediction and the adaptive capabilities for our learner model by using data mining techniques on data from 71 students. We believe this work will help the research community in building and assessing a BN in an ITS that teach logical reasoning.

**Keywords:** Bayesian network · Deductive reasoning · Learner model · Intelligent tutoring system

## 1 Introduction

The work presented here is part of the development of an Intelligent Turoring System (ITS) Logic-Muse [6] which aim is to help learners improve their reasoning skills in the context of classical propositional logic. All its three main components have been developed while relying on the help of experts and on important work in the field of reasoning and computer science. Modeling students' knowledge is a fundamental part of intelligent tutoring systems. A learner's state of knowledge is subject to change and competence should be assigned with some degree of certainty, so the learner model can only be an approximation of his actual condition. It is thus important to support the diagnosis with a formalism that allows uncertain inferences about a learner. Bayesian Networks (BN) are quite adequate for the task: they allow to infer the probability of mastering a skill from a specific response pattern [1, 2]. We thus created a BN that allows real time diagnosis and modeling of the learner's knowledge state. Learner modeling is valid only if it accurately reflects the learner's progress longitudinally. Evaluation of the inference mechanism addresses the evaluation of the validity of user properties inferred from the input data previously collected. In order to ensure the effectiveness of the learner model, we performed a formative validation.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 627–631, 2016.
DOI: 10.1007/978-3-319-45153-4_78

This paper aims, firstly, at providing some relevant information about the BN such as all the details about the choice of the a priori probabilities, the structure of the network and the nodes representing measured skills. We will also present a preliminary evaluation of the Network using some relevant data-mining techniques. The preliminary results showed that the learner model implemented in Logic-Muse is able to model and predict learner's knowledge with an accuracy of about 90 %.

## 2 The Learner Component of Logic-Muse

The learner model allows an ITS to adapt the interaction to its user's specific needs. One of the biggest challenges in designing ITS is the effective assessment and representation of the student's knowledge state and specific needs in the problem domain based on uncertainty information. It is thus important to support the diagnosis with a formalism that allows uncertain inferences about a learner. We use a BN to represent the user's knowledge as accurately as possible. It was built from the domain knowledge, where causal relationships between nodes (reasoning skills) as well as prior probabilities are provided by the experts.

### 2.1 The Bayesian Network for the Cognitive Diagnosis

BN is represented as a directed acyclic graph (DAG) with nodes for uncertain variables and edges for directed relationships between the variables. In the BN built, the nodes are directly connected to the reasoning activities. The skills involved in the BN are those put forward by the mental models theory to reason in conformity to the logical rules. This includes the inhibition of exceptions to the premises, the generation of counterexamples to the conclusion and the ability to manage all the relevant models for the concrete, contrary to fact and abstract informal [5]. To develop our BN, we have considered that cognitive parameters and diagnosis can be modeled by random variables. We have considered two types of nodes. The nodes measuring the learner' s knowledge or skills, and those containing the evidence, which represent answers to exercises.

Because deductive reasoning is what to be learn, it represents the global node of the BN. According to [5], there is 3 steps or "know-how" to make a conditional reasoning and then succeed to all type of exercises (MPP (Modus Ponendo Ponens), MTT (Modus Tollendo Tollens, DA (Denied the Antecedent), AC (Affirmation of the Consequent)).

**Inhibition of P and not Q:** It is to inhibit the "disabler" or restrictive condition that is to inhibit any conditions which lead to think that P & not Q is true. In example: If it rains I take my umbrella. It is raining", the logical conclusion is to say" I take my umbrella. The disabler would be for example, "we cannot be sure that I will take my umbrella because it can be broken". We forget that P implies Q.

**Generation of not P and Q:** Is to generate alternatives as and thus avoid fallacies and succeed on AC and DA exercises type.

**Three Mental Models Management:** P and Q, not P and not Q, not P and Q. These 3 models are needed to completely understand the deductive reasoning.

These 3 steps represent skills nodes in the BN that are directly connected to inferences (MPP, MTT, AC, DA are also skills nodes) and different contexts implemented in order to make the reasoning exercise more or less difficult [5]. There are 3 reasoning contexts; the causal context (or familiar): reasoning on real life sentences; the contrary-to-fact: reasoning on sentences that are not feasible according to our knowledge of the world: "If I throw ketchup on a shirt then it will be clean."; the abstract: reasoning on abstract terms: "If a person morp, it will become plede". We denoted 28 skills. The number of items nodes is the size of our item bank. The structure and the prior probabilities of our BN was built with the help of human experts in psychology of reasoning.

The system's estimate that a student has acquired a skill is continually updated every time the student gives a first response to a step in the problem. The system then recomputes the probability that the student knew the skill before the answer, using the evidence from the answer. Exercises are chosen according to these probabilities. Further-more, a CDM-Based (Cognitive Diagnosis Models) psychometric model [3, 7] is built using the item bank, a Q-Matrix (items/skills), as well as data from all student responses to items. The resulting model is part of the learner model as well and allows for initial predictions of learner strengths and weaknesses regarding the reasoning skills given his/her performance on items. More concretely, we predict the probability a learner mastering the overall competence via their pre-test results. For this, we use the "posterior" matrix obtained through the CDM. We seek a learner's response pattern, the line of the "posterior" matrix containing the same pattern or a similar pattern. The joint probability matching this pattern, calculated based on the probabilities associated with each skill is used as the a priori likelihood of mastering the root node of the BN.

### 2.2 Evaluation of the Bayesian Network

To assess the predictive ability of our BN and its ability to best represent the current skills of a learner, we opted for an incremental cross validation. Evaluation of the inference mechanism addresses the evaluation of the validity of user properties inferred from the input data previously collected. The reliability assessment is done by using a prediction procedure [4] and an incremental cross validation in which one tries to predict learner 'answers and skills using the BN. These assessments use data collected from 71 students. These data consist of answers from the 71 students to a test containing 48 deductive reasoning problems prepared by our team.

**Data Preparation.** The very first step was to preprocess the raw data obtained from the 71 students. For each of the 48 questions, students had to choose between 3 answers (the valid one, the invalid typical one and the invalid atypical one). We generated a binary context that has 71 rows and 48 columns. The 3 choices were encoded as "1" for the valid answer, "0" for the invalid typical answer and the invalid atypical answer.

Student models that focus on knowledge assessment may be evaluated by comparing their predictions of the student's knowledge to actual student performance. Thus, to assess the predictive ability of our BN we opted for an incremental cross validation.

The training data increase one by one and the test data decrease one by one. For each of the 71 students, we have compared the real answer of each question with the one predicted by the network. For example, for a student, we extracted the likelihood of correctly answering question 1 and then we com-pared it with his actual answer. After that, we introduced his real answer to the network and we extracted the likelihood of the second question, which we compared with his answer to that question. We noticed that, after an average of 10 to 15 questions answered, the BN is able to predict the behavior of a learner with an accuracy of 95 %. Some errors can be due to the guess (giving a correct answer, despite not knowing the skill) and slip (knowing a skill, but giving a wrong answer) parameters. We summarize by saying that the system gives a good representation of the learner's knowledge. However, we must improve the prior probabilities. Currently, an incorrect answer to a question is represented by a probability below 0.6; it would be ideal if this limit could vary according to a specific skill.

## 3 Conclusion and Future Work

We presented a BN (which represent the learner model) and theoretical elements that led us to such a structure. A contribution of Logic-Muse' student model is that it supports prediction of student knowledge and behavior in a learning session of logical reasoning. We obtained a very high accuracy rate of the prediction model compared to what is usual. Such encouraging results show that our prediction model is valid as well as reliable. We have proven its effectiveness on 71 students. The BN is able to predict learner knowledge and make a faithful representation of the learner's knowledge state. The prior probabilities in the network will be refined according to the results obtained from this first evaluation. Since we have planned to deploy Logic-Muse in a Logic course in autumn 2016, we will conduct the summative evaluation (regarding the added value of such a system in the learning of logical reasoning) at that time. We believe this work will help the research community in building and assessing a BN in an ITS that teach logical reasoning.

## References

1. Conati, C., Gertner, A., Vanlehn, K.: Using Bayesian networks to manage uncertainty in student modeling. User Model. User-Adap. Inter. **12**(4), 371–417 (2002)
2. Conati, C., Cerri, Stephano A.: Bayesian student modeling. In: Nkambou, R., Bourdeau, J., Mizoguchi, R. (eds.) Advances in Intelligent Tutoring Systems. Studies in Computational Intelligence, vol. 308, pp. 281–299. Springer, Heidelberg (2010)
3. De La Torre, J.: A cognitive diagnosis model for cognitively based multiple-choice options. Appl. Psychol. Measur. **33**(3), 163–183 (2009)
4. Lesta, L., Yacef, K.: An intelligent teaching assistant system for logic. In: Cerri, S.A., Gouardéres, G., Paraguaçu, F. (eds.) ITS 2002. LNCS, vol. 2363, pp. 421–431. Springer, Heidelberg (2002)
5. Markovits, H.: On the road toward formal reasoning: Reasoning with factual causal and contrary-to-fact causal premises during early adolescence. J. Exp. Child Psychol. **128**, 37–51 (2014)

6. Nkambou, R., Brisson, J., Kenfack, C., Robert, S., Kissok, P., Tato, A.: Towards an intelligent tutoring system for logical reasoning in multiple contexts. In: Conole, G., Klobucar, T., Rensing, C., Konert, J., Lavoué, E. (eds.) EC-TEL 2015. LNCS, vol. 9307, pp. 460–466. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24258-3_40
7. Robitzsch, A., et al.: CDM: Cognitive diagnosis modeling. R Package version, 3 (2014)

# Finding the Needle in a Haystack: Who are the Most Central Authors Within a Domain?

Ionut Cristian Paraschiv[1], Mihai Dascalu[1,2(✉)], Danielle S. McNamara[2], and Stefan Trausan-Matu[1]

[1] Computer Science Department, University Politehnica of Bucharest, Bucharest, Romania
ionut.paraschiv@cti.pub.ro, {mihai.dascalu,stefan.trausan}@cs.pub.ro

[2] Institute for the Science of Teaching and Learning, Arizona State University, Tempe, USA
dsmcnama@asu.edu

**Abstract.** The speed at which new scientific papers are published has increased dramatically, while the process of tracking the most recent publications having a high impact has become more and more cumbersome. In order to support learners and researchers in retrieving relevant articles and identifying the most central researchers within a domain, we propose a novel 2-mode multilayered graph derived from Cohesion Network Analysis (CNA). The resulting extended CNA graph integrates both authors and papers, as well as three principal link types: co-authorship, co-citation, and semantic similarity among the contents of the papers. Our rankings do not rely on the number of published documents, but on their global impact based on links between authors, citations, and semantic relatedness to similar articles. As a preliminary validation, we have built a network based on the 2013 LAK dataset in order to reveal the most central authors within the emerging Learning Analytics domain.

**Keywords:** Learning analytics · 2-mode multilayered graph · Co-authorship · Co-citation · Semantic similarity

## 1 Introduction

With the growing flow of information and emerging new inter-disciplinary research topics, it is becoming increasingly difficult to find and follow relevant publications and authors. Each research sub-domain (e.g., Learning Analytics or Educational Data Mining) usually starts from a few authors who introduce broad research questions or trending topics around which a community gradually evolves. Usually, the initial authors become central members in the research network, being cited in new publications. The research question that arises regards how can we identify the most important authors and publications within a sub-domain, and what are the metrics that can be effectively applied in order to obtain a relevant global view of the underlying research? In our previous research studies [1, 2], we have built a learning analytics engine capable of annotating a dataset of articles using their semantic context, and displaying them within a network of papers that highlights their semantic relations.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 632–635, 2016.
DOI: 10.1007/978-3-319-45153-4_79

In addition, our work has made extensive use of Cohesion Network Analysis (CNA) [3], a cohesion centered representation of discourse in which semantic similarity links between different text segments are combined into a multi-layered cohesion graph [4]. This graph provides valuable insights of local cohesion expressed in the semantic relatedness between adjacent or transition sentences, meanwhile transcending towards global cohesion when evaluating inter-paragraph cohesion flow. Having this background, we propose a new approach, an extended CNA *2-mode multilayered graph*, capable of facilitating the identification of the most important authors and publications from a research domain by applying various Social Network Analysis (SNA) metrics [5]. As an initial validation, we have used the model to identify the top central authors and articles from the LAK (Learning Analytics and Knowledge) Dataset [6], which includes publications from the Learning Analytics domain (652 LAK and EDM conference papers, 45 journal papers, and 1214 distinct authors) in RDF format (https://www.w3.org/TR/REC-rdf-syntax/), containing unique URIs for all authors, articles and citations.

## 2 The Extended CNA 2-Mode Multilayered Graph

Our model combines three different approaches to evaluate the importance of both authors and articles within a domain: *Co-citation Analysis*, *Co-authorship Networks* and *Semantic Similarity*. These three types of links are used to build a *2-mode multilayered graph* on which graph theory measures [7] are applied to identify the most central nodes, (i.e. authors, papers) from the input dataset. The generated graph represents an integrated view of articles and authors, where each layer contains links with scores computed using different approaches. By jointly indexing the two different sets of nodes contained in our 2-mode graph, co-occurrence patterns emerge [8], suitable for generating an overview of the domain.

*Co-authorship links* [9] represent the first layer of our extended CNA graph in which two papers are related if they have at least one common author. Usually, the same author is interested in similar topics, so we can assume that papers with at least one common author are related. At the second layer, *co-citations* are enforced, having as roots one of the first techniques developed to annotate a dataset of articles [2]. The idea is that two papers are related if they contain at least one common citation, meaning that they should have semantic resemblance. The increase in the number of common citations between two articles usually denotes a higher degree of similarity and a tighter coupling among them. Third, the *semantic similarity* layer shifts the focus towards the actual content of the papers by evaluating the degree of their relatedness. Our integrated framework, *ReaderBench* [3, 4], integrates the automated building process of the CNA cohesion graph in which multiple semantic models are combined: (a) cosine similarity in Latent Semantic Analysis (LSA) vector spaces, (b) Jensen-Shannon dissimilarity between Latent Dirichlet Allocation (LDA) topic distributions, and (c) semantic distances (e.g., path length, Wu-Palmer, Leacock-Chodorow) in lexicalized ontologies – WordNet [4].In addition, we take the analysis further by applying SNA metrics [5] to identify patterns and meaningful relations between nodes, in conjunction with the evaluation of each node's centrality. First, *degree centrality* quantifies the importance of each node as the sum of the scores of all links connected to it. Second, *closeness* reflects the centrality of each node as the average sum of all shortest paths between the current node and all other nodes in the graph; closeness can therefore be considered a measure of speed in terms of spreading the information within the network [10]. Third, *betweenness* evaluates the number of times a given node acts as a bridge along all shortest paths between pairs of any two other nodes. In contrast to closeness, betweenness can be perceived as a measure of control for the linkage among other nodes [10].

## 3 Exploring the LAK Dataset

Our CNA *2-mode multilayered graph* was applied to the 2013 LAK dataset [6] that contains machine readable information in which each resource (author, article or citation) is uniquely identified. Table 1 depicts the top 10 authors in terms of betweenness centrality. The top 5 authors are *"Ryan Baker", "Neil Heffernan", "Joseph Beck", "Kenneth Koedinger"* and *"Jack Mostow"*, authors with a high impact in the broader Computer Education domain, as well as the Learning Analytics domain, having a total of 102 unique published papers and more than 33,000 collective citations according to Google Scholar. The top ten authors collectively reach more than 80,000 citations and 141 unique papers in the dataset. Of particular interest is "Jose Gonzales-Brenes" who does not have many citations ($n$ = 125), but is a co-author in 5 out of 8 papers with "*Jack Mostow*" (ranked 5) and in one with "*Peter Brusilovsky*" (ranked only 25 in this data set, but with more than 20,000 citations worldwide). Therefore, Gonzales-Brenes is tightly connected to two highly influential researchers and creates a bridge between the two research communities.

**Table 1.** Top 10 authors from Learning Analytics ordered by their betweenness centrality.

| Author | M1 | M2 | M3 | P | CC | NP |
|---|---|---|---|---|---|---|
| Ryan Baker | 43,191 | 0.9 | 2,817 | 36 | 5,968 | 2 |
| Neil Heffernan | 23,823 | 0.8 | 2,317 | 25 | 3,645 | 2 |
| Joseph Beck | 18,906 | 0.8 | 2,110 | 18 | 2,958 | 1 |
| Kenneth Koedinger | 17,938 | 0.8 | 2,274 | 23 | 17,317 | 1 |
| Jack Mostow | 15,689 | 0.8 | 1,943 | 16 | 3,773 | 0 |
| Arthur Graesser | 14,573 | 0.7 | 1,788 | 16 | 34,539 | 1 |
| Zachary Pardos | 12,920 | 0.8 | 2,149 | 13 | 857 | 0 |
| Jose Gonzalez-Brenes | 12,448 | 0.8 | 1,848 | 8 | 125 | 0 |
| Sebastian Ventura | 11,200 | 0.8 | 1,832 | 14 | 6,035 | 0 |
| Cristobal Romero | 10,312 | 0.8 | 1,810 | 15 | 5,077 | 0 |

* SNA Metrics: *M1* = Betweenness centrality; *M2* = Closeness centrality; *M3* = Degree; *P* = Number of published articles; *CC* = Citation count; *NP* = Number of papers from Top 10.

## 4 Conclusions

In this paper, we have introduced a *2-mode multilayered graph*, an extension of our Cohesion Network Analysis, which a represents a combination of multiple complementary perspectives in order to build a mixed *article-author* graph. In the context of hundreds, even thousands of publications within each research field every year, our approach provides valuable support in retrieving relevant resources, helping learners to *find the needle in the haystack*.

Our method can be further extended with additional SNA metrics, enhanced visualization tools, and the ability to check the evolution of a domain. Currently, the views are highly cluttered because of the large number of nodes - a potential solution would presume the creation of hierarchical clusters that group similar nodes. With such future modifications, we expect the CNA 2-mode multilayered approach to have a significant impact on information retrieval.

**Acknowledgement.** This work is partially funded by the 644187 H2020 RAGE (Realising an Applied Gaming Eco-System) http://www.rageproject.eu/project.

## References

1. Paraschiv, I.C., Dascalu, M., Dessus, P., Trausan-Matu, S., McNamara, D.S.: A paper recommendation system with readerbench: the graphical visualization of semantically related papers and concepts. In: Li, Y., et al. (eds.) State-of-the-Art and Future Directions of Smart Learning. LNET, pp. 443–449. Springer, Germany (2015)
2. Paraschiv, I.C., Dascalu, M., Trausan-Matu, S., Dessus, P.: Analyzing the semantic relatedness of paper abstracts - an application to the educational research field. In: DS-CSCL-2015/CSCS20, pp. 759–764. IEEE, Bucharest (2015)
3. Dascalu, M., Trausan-Matu, S., McNamara, D.S., Dessus, P.: ReaderBench – automated evaluation of collaboration based on cohesion and dialogism. Int. J. Comput. Supported Collaborative Learn. **10**(4), 395–423 (2015)
4. Dascalu, M.: Analyzing discourse and text complexity for learning and collaborating. Studies in Computational Intelligence, vol. 534. Springer, Cham (2014)
5. Scott, J.: Social Network Analysis. SAGE Publications Ltd., Thousand Oaks (2012)
6. Arora, R., Ravindran, B.: Latent Dirichlet Allocation based multi-document summarization. In: 2nd Workshop on Analytics for Noisy Unstructured Text Data, pp. 91–97. ACM, Singapore (2008)
7. Biggs, N., Lloyd, E., Wilson, R.: Graph Theory, 1736-1936. Oxford University Press, Oxford (1986)
8. Borgatti, S.: 2-mode concepts in social network analysis. In: Meyers, R.A. (ed.) Encyclopedia of Complexity and System Science, pp. 8279–8291. Springer, New York (2009)
9. Newman, M.E.J.: Coauthorship networks and patterns of scientific collaboration. In: Mapping Knowledge Domains. Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering, Irvine (2003)
10. Newman, M.E.J.: A measure of betweenness centrality based on random walks. Soc. Netw. **27**, 39–54 (2005)

# Bio-inspired Computational Algorithms in Educational and Serious Games: Some Examples

Michela Ponticorvo(✉), Andrea Di Ferdinando, Davide Marocco, and Orazio Miglino

Department of Humanistic Studies, University of Naples "Federico II", Naples, Italy
michela.ponticorvo@unina.it

**Abstract.** Bio-inspired computational algorithms can be effectively employed to develop games for learning. Game design, which we propose to describe according to a multi-level framework where the external level is distinguished from the game engine and the tutoring level, can host different bio-inspired computational algorithms.

Some examples of educational games employing bio-inspired algorithms at different levels are reported: BreedBot in which bio-inspired computational algorithms are used at game level and Infanzia Digi.tales project where these techniques are used at tutoring level.

**Keywords:** Technology Enhanced Learning · Serious games · Educational games · Bio-inspired computational models · Game design

## 1 Introduction

In recent years an epochal turn has been observed in education coming from a twofold pathway. On one side, a growing effort has been devoted to the use of new technologies, in particular ICT (information and communication technologies), as educational tools. Technology-Enhanced learning (TEL) has intercepted this tendency by promoting new educational practices, new communities and new ways of communication [1]. On the other side, a lot of interest has arisen about the use of game for learning. This interest is witnessed by the numerous research branches that emerged, game-based learning [8], edutainment [2], gamification of learning [6], just to cite some. In particular many games have been developed under the label Educational Games and Serious games. Educational games include card, board and videogames. Playing a game always requires to learn something, at least game content and dynamics and in educational games this aspect can be exploited to convey specific contents. Serious Games (SG) are games that educate, train, and inform [7], sharing the same educational mission. The design process is crucial to fully express educational potential of digital games and, in the domain of digital SG, computational models can be exploited for this goal. Between the


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 636–639, 2016.
DOI: 10.1007/978-3-319-45153-4_80

computational models that can be chosen, bio-inspired computational models are extremely fit for educational purpose if the goal is to teach biological, psychological and social matters, because they allow to convey knowledge about dynamic and complex system, emergence, evolution and development better than other computational models.

## 2 Serious and Educational Game Design According to a Multi-level Framework

In this section we will describe the SG design process according to a multi-level framework where we can distinguish two concentric levels, the shell and core level and a ubiquitous one, the evaluation and tutoring level [3], represented in Fig. 1. The shell and the core level are present in every kind of game, and, more in general in almost every cultural product. The shell level represents the visible content that is immediately accessible to the player. It frames the game engine, the game dynamics that are hold in the core level. The third level, the evaluation and tutoring level, even if it is present in many entertainment games, is characterizing for Educational and Serious games, as it allows, on the teachers side, to understand if and how the player/learner has acquired the concepts conveyed by the Educational game.

The shell level represents what the player sees, the setting she is immersed in. Here we find what we call the game narrative. Digital games, as many other cultural products, are expressed through a narrative metaphor that carries out the crucial role to give sense to the game. In designing the shell level we have to define the context: who are the agents, what actions they can display, what interactions are possible between them.

The shell level, based on narrative, holds an hidden level with a specific operation, the game engine, what we call the core level. The game engine, a term that is commonly used in the context of videogames creation and development, allows to implement core functionalities related to game dynamics, for example related

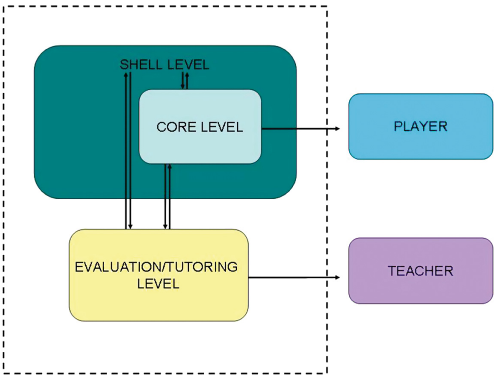



**Fig. 1.** Multi-level framework for educational games

to physics, animation, artificial intelligence, etc. These levels are in dynamic interaction and have strong effects one on the other: the narrative provides a frame where the hidden content resides. In educational context, the shell level is necessary in providing a semantic context to educational activities whereas the core level defines the skills or the abilities to be transferred.

If our goal is to build educational tools and materials which are related to biology, psychology and sociology or if we want to transmit different subjects adopting a point of view that takes into account aspects related to emergence, complex and dynamic systems, evolution and development, we can resort to a wide class of bio-inspired algorithms. Bio-inspired computing exploits the study of natural phenomena to apply it to machine learning: from evolution to genetic algorithms, from natural complex systems to cellular automata, from the nervous systems to artificial neural networks.

In educational and serious games, a relevant role is played by the evaluation and tutoring level. The evaluation and tutoring layer complements the core and shell layers. This level analyze players game performances relatively to the specified training objectives, and provides the players and the trainer, whose role is indeed relevant in educational context, with important information and data about the learning process. At this level we find learning analytics, which are the measurement, collection, analysis and reporting of data about learners to improve the whole learning process.

## 3 Bio-inspired Computational Models in Educational and Serious Games: Some Examples

The first examples we want to cite are about the use of bio-computational algorithms to teach evolutionary dynamics. In this case, the serious game becomes a virtual laboratory where the user can directly manipulate the relevant variables involved in the game, thus determining the game evolution in an immediate manner. At the same time, this direct manipulation takes place in a protected environment where failures or error do not determine a menacing outcome. An interesting example of this kind of games we have worked on is Breedbot[1] and its sequels Bestbot, and Brainfarm [4,5]. These are integrated software/hardware platforms that allow players, even without any particular computer skill, to breed, within customizable virtual worlds, artificial organisms that can be downloaded onto real robots.

Breeding is implemented through a user-guided genetic algorithm, where the user determine robots evolution acting as a breeder. In these games there are the following bio-inspired computational models: robots are embodied agents whose artificial intelligence is implemented with artificial neural networks and their evolution/development carries out adopting evolutionary algorithms.

Bio-inspired computational model can enter the evaluation/tutoring level, as shown by INF@NZIA DIGI.tales project. This level foresees a smart interaction with the user/player. This smartness resides in adapting, inferring, profiling and anticipation, functions that mimic human teachers actions. This level provides

[1] The interested readers can contact authors for additional materials.

an appropriate and timely feedback to player action, it adapts to player special needs according to her actual performance and the desired educational goals, it tracks player performance in terms of achievements and improvements. Up to now, this smart interaction has been mediated by the use of Intelligent Tutoring Systems (ITS).

Bio-inspired algorithms can be useful at this level too: artificial neural networks can be applied to teaching and learning processes, as they can capture interesting regularities that help profiling the student/player/user, modelling student/teacher interaction is a smart way. Learners and teachers can be conceived as cognitive agents, starting from the regularities extracted by Educational data mining.

## 4 Conclusions

Bio-inspired computational methods can be applied effectively in designing Serious and Educational games because they are fit to teach some arguments such as biology, psychology, sociology with an isomorphic approach; they open the way to some aspects which are indeed relevant, but are often neglected in educational contexts such as physical embodiment, autonomy, social interaction, evolution and development; they help reproducing ecological dynamics in the abstract world of digital games.

**Acknowledgments.** The INF@NZIA DIGI.tales has been funded by Italian Ministry for Education, University and Re-search under PON-Smart Cities for Social Inclusion programme. Authors would like to thank Onofrio Gigliotta for Breedbot, Bestbot and Brainfarm materials.

## References

1. Balacheff, N., Ludvigsen, S., De Jong, T., Lazonder, A., Barnes, S.A., Montandon, L.: Technology-Enhanced Learning. Springer, Heidelberg (2009)
2. Charsky, D.: From edutainment to serious games: a change in the use of game characteristics. Games Cult. **5**, 177–198 (2010)
3. Dell'Aquila, E., Di Ferdinando, A., Marocco, D., Miglino, O., Ponticorvo, M., Schembri, M.: New perspective. In: Educational Games for Soft-Skill Training. Springer, Heidelberg (2017, in press). ISBN 978-3-319-06311-9
4. Miglino, O., Gigliotta, O., Ponticorvo, M., Nolfi, S.: Breedbot: an edutainment robotics system to link digital and real world. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part II. LNCS (LNAI), vol. 4693, pp. 74–81. Springer, Heidelberg (2007)
5. Miglino, O., Gigliotta, O., Ponticorvo, M., Nolfi, S.: Breedbot: an evolutionary robotics application in digital content. Electron. Libr. **26**(3), 363–373 (2008)
6. Kapp, K.M.: The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education. Wiley, Chichester (2012)
7. Michael, D.R., Chen, S.L.: Serious Games: Games That Educate, Train, and Inform. Muska and Lipman/Premier-Trade, New York (2005)
8. Tobias, S., Fletcher, J.D., Wind, A.P.: Game-based learning. In: Handbook of Research on Educational Communications and Technology, pp. 485–503. Springer, New York (2014)

# Learning Experiences Using Tablets with Children and People with Autism Spectrum Disorder

David Roldán-Álvarez[1], Ana Márquez-Fernández[2], Estefanía Martín[2(✉)], and Cristian Guzmán[2]

[1] Universidad Autónoma de Madrid, 28049 Madrid, Spain
david.roldan@uam.es
[2] Universidad Rey Juan Carlos, 28933 Móstoles, Madrid, Spain
anamarqfer@gmail.com, estefania.martin@urjc.es, c.guzmanl@alumnos.urjc.es

**Abstract.** Learning technologies offer new opportunities to children and people with disabilities to develop their autonomy and independence. In recent years, and thanks to the emergence of touch devices, lots of efforts have been put into creating content and applications for these kinds of surfaces. In addition, current literature shows the benefits of using technology for improving the learning process of these students. This paper presents two learning experiences where pre-school aged children and students with special needs performed learning activities using tablets. The results obtained shed light on these types of devices, which are suitable for early ages and special education since they do not need intermediate devices such as the keyboard or the mouse.

**Keywords:** Learning · Tablets · Children · Kindergarten · Special needs

## 1 Introduction

In the educational environment and little by little, books have been complemented in classrooms with technological devices [1]. ICT provide excellent tools to help people with special needs to gain independence when doing their daily activities [2] and also to work new concepts with very young children. Technology can help improve their confidence and motivation since it promotes errorless learning. It allows teachers to offer them personalized assessments and to adapt the rhythm of learning [3].

In the last decade, touch devices have emerged as an alternative to traditional interfaces, providing users a new way of interacting without intermediate elements such as mouse or keyboard. Through the use of natural gestures to interact with touch devices, users can express themselves in a physical way, enhancing communication and comprehension [4]. It has been proven that touch devices help users to focus on the contents and solve problems more quickly while they have fun [5]. Combining a touch interaction with appropriate multimedia content, users feel that they are controlling the information and the way they interact, which helps them gain deeper knowledge of the topic presented [6].


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 640–643, 2016.
DOI: 10.1007/978-3-319-45153-4_81

Some examples can be found in the literature about the use of touch devices to improve student social behavior and their gain of knowledge [7, 8]. These studies show how touch devices promote social interaction among the participants while they perform the activities and how their knowledge gain is enhanced when compared to those students who performed the activities in a more traditional way.

This paper presents two learning experiences with pre-school aged children and people with special needs where we measured students' learning and the implications of using touch technology in their learning process. The activities were designed with DEDOS-Editor [9] and the students solved them on tablets using DEDOS-Player.

## 2 Learning Experiences

As mentioned in the previous section, tablets are suitable devices for children and for people with special needs since they eliminate the need for intermediate devices. The experiences presented in this section are focused on tablets. Their goal was to measure the effects of using tablets during the learning process of students (Fig. 1).

**Fig. 1.** Students performing learning activities with Android tablets

### 2.1 Pre-primary Education Experience

We performed a learning experiment with 20 students aged between 5 and 6 years old in order to study if the use of technology influenced their learning processes. During this experiment, students were divided in two groups of 10 students each. The first group performed the activities with tablets while the second group performed the same activities with paper. The topic of the activities was the environment, which is part of the of pre-primary education curriculum. Specifically, activities about wild animals, domestic animals and farm animals. The study was performed during three months. Before starting and after finishing the learning experience, the students did tests (pre-test and post-test) so we could check more precisely their previous knowledge on this topic and to see if there was significance learning when the experiment ended. Both the activities in the tests and the activities they performed using paper and tablets where multiple choice, pair-matching and math activities.

The results obtained shed light on the importance of technology in classrooms. On one hand, students who performed the activities with tablets scored 14,18 out of 24 points in the pre-test and improved their results in the post-test, obtaining 21,42 out of 24 points. On the other hand, students who performed the activities with paper obtained 17,21 out of 24 points in the pre-test and 15,21 out of 24 point in the post-test. As the results show, the use of technology has led the students to gain meaningful learning (Wilcoxon test, $W = 0$ $Z = 2.8$ $p < 0.00$) while we could not confirm if there was a meaningful learning among the students who did not use tablets (Wilcoxon test, $W = 41$ $Z = 1.38$ $p = 0.19$).

### 2.2 Special Needs Experiment

17 students between 12 and 20 years old were involved in the second learning experiment. All of them have cognitive disabilities and 5 of them have also been diagnosed with Autism Spectrum Disorder (ASD). Students had to complete two assignments. The first one was composed of 24 activities about musical instruments where the students were asked about the name or the type of a certain instrument. The second project contained 17 activities where students were asked about questions related to the "Theory of Mind" (recognition of facial expressions and understanding of feelings) and daily life activities. Each student completed the assignments at least once a week during three weeks total. The duration of the sessions was from 7 to 15 min, depending on the time the participant took to solve all the activities. We did not set a maximum time to solve both projects since we did not want the students to nervous.

During the learning experiment, students performed the activities more independently as the study progressed. For instance, in the first session students gave 241 correct answers in a total of 312 activities while they gave 265 correct answers in the last session. In addition, the number of supports they needed decreased along the duration of the study. Students required help from their teachers or the observers; the student who needed the most help was assisted 11 times in the first session and in the last session he/she only was only helped 4 time. Moreover, they were aware of the mistakes they made, self-correcting when asked again about the same topic. For example, when asked about the type of instrument of a piano, the first time 9 out of 13 participants where wrong, while in the last session only 1 participant chose the wrong answer. Students were motivated throughout the entire interaction and they wanted to perform as well as they could.

## 3 Conclusion

As shown in the learning experiments, we strongly believe that the use of technology influences students when performing educational activities. Moreover, the use of technology motivates them and, combined with traditional methods and other learning sources, we could achieve good academic results. The portability and accessibility provided by tablets make them an interesting tool to be used in classrooms both for students. By promoting smooth and direct interaction with tablets we facilitate the students to engage with the activities they have to solve, reducing their frustration and increasing their willingness to interact with the application.

The limitations of these two studies were the number of participants. Therefore, it would be interesting to repeat them with more students. However, the results obtained are promising since students were focused on the learning concepts and were motivated to manipulate the elements using their own hands.

**Acknowledgements.** The research presented in this paper has been funded by the Spanish Ministry of Economy and Competitiveness under grant agreement TIN2013-44586-R, "*e-Training y e-Coaching para la integración socio-laboral*" and by Comunidad de Madrid under project S2013/ICE-2715.

## References

1. Tondeur, J., Van Keer, H., van Braak, J., Valcke, M.: ICT integration in the classroom: challenging the potential of a school policy. Comput. Educ. **51**(1), 212–223 (2008)
2. Seegers, M.: Special technological possibilities for students with special needs. Learn. Lead. Technol. **29**(3), 32–39 (2001)
3. Duffy, L., Wishart, J.: The stability and transferability of errorless learning in children with Down syndrome. Down Syndr. Res. Pract. **2**(2), 51–58 (1994)
4. Cantón, P., González, L., Mariscal, G., Ruiz, C.: Applying new interaction paradigms to the education of children with special educational needs. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012, Part I. LNCS, vol. 7382, pp. 65–72. Springer, Heidelberg (2012). doi:10.1007/978-3-642-31522-0_10
5. Africano, D., Berg, S., Lindbergh, K., Lundholm, P., Nilbrink, F.: Designing tangible interfaces for children's collaboration. In: Extended Abstracts on Human Factors in Computing Systems. ACM, New York, pp. 853–868 (2004). doi:10.1145/985921.985945
6. Roldán-Álvarez, D., Márquez-Fernández, A., Rosado-Martín, S., Martín, E., Haya, P.A., García-Herranz, M.: Benefits of combining multitouch tabletops and turn-based collaborative learning activities for people with cognitive disabilities and people with ASD. In: IEEE 14th International Conference on Advanced Learning Technologies, pp. 566–570. IEEE (2014)
7. Ortega-Tudela, J., Gomez-Ariza, C.: Computer assisted teaching and mathematical learning in Down syndrome children. J. Comput. Assist. Learn. **22**, 298–307 (2006). doi:10.1111/j.1365-2729.2006.00179.x
8. Lingnau, A., Zentel, P., Cress, U.: Fostering collaborative problem solving for pupils with cognitive disabilities. In: Chinn, C.A., Erkens, G., Puntambekar, S. (eds.) Proceedings of the Computer Supported Collaborative Learning Conference. International Society of the Learning Sciences, Rutgers University, New Brunswick, pp. 450–452 (2007)
9. Roldán-Álvarez, D., Martín, E., García-Herranz, M., Haya, P.A.: Mind the gap: impact on learnability of user interface design of authoring tools for teachers. Int. J. Hum. Comput. Stud. (2016, in press). doi:10.1016/j.ijhcs.2016.04.011

# Introducing the U.S. Cyberlearning Community

Jeremy Roschelle, Shuchi Grover, and Marianne Bakia(✉)

SRI International, Menlo Park, CA, USA
{jeremy.roschelle,shuchi.grover,marianne.bakia}@sri.com

**Abstract.** The term "Cyberlearning" is used in the United States to describe a community of researchers, largely funded by the US National Science Foundation, who are exploring the integration of computer science research with learning sciences research. The Cyberlearning community is parallel to the EC-TEL community and the purpose of this poster is to foster mutual engagement between the communities. The paper describes the origin of the term, the conception of the field, the kinds of research being conducted, and some of the exemplary projects. The paper will also introduce the Center for Innovative Research in Cyberlearning (CIRCL), which is the hub of the knowledge network (research community) for cyberlearning and hosts a useful collection of resources.

**Keywords:** Innovation · Learning · Technology

## 1 Introduction

Researchers in the United States have begun using the term "Cyberlearning" to describe a portfolio of early-stage, conceptual projects. The projects collectively aim to tightly intertwine emerging technology with recent progress in the learning sciences to enable a broader diversity of people to learn advanced content. A group of US-based researchers engaged in this work intends to participate in the EC-TEL meeting in order to exchange ideas with like-minded European researchers; this paper is intended to lead to a poster at EC-TEL which would encourage interchange.

The term "cyberlearning" was coined in a 2008 report [1], which identified that advancing network technologies could enable ambitious designs for learning to break out of conventional school-based learning structures. The report advocated for 7 priorities: (1) advance seamless cyberlearning across formal and informal settings, (2) seize the opportunity for remote and virtual laboratories, (3) investigate virtual worlds and mixed-reality environments, (4) institute programs and policies to promote open educational resources, (5) harness the scientific-data deluge, (6) harness the learning-data deluge, and (7) recognize cyberlearning as a pervasive NSF-wide strategy.

Cyberlearning was defined (somewhat vaguely) as "learning that is mediated by networked computing." The referent was to "cyberinfrastructure" – a term in use in the United States and which is parallel to the European "e-science." The term was not intended to relate to "cyber-crime" or "cyber-security." The report task force urged researcher to go beyond behind the typical classroom computers and to address mobility, sensors, augmented reality, big data, and other new affordances of technology.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 644–647, 2016.
DOI: 10.1007/978-3-319-45153-4_82

The term led to a National Science Foundation funding program called "Cyberlearning: Transforming Education" (CTE) in 2011. CTE [2] further refined the definition of cyberlearning to take it beyond simply using educational technology tools and emphasized "integrating advances in technology with advances in what is known about how people learn" – that is, a strong emphasis on learning sciences research in conjunction with a focus on emerging technologies. In addition, CTE added a focus on "populations not served well by current educational practices," to address issues of equity and diversity and was very deliberately defined to span informal and formal learning environments.

A research summit (see http://circlcenter.org/events/summit-2012/) was held in 2012 and helped to launch the nascent field. With regard to the emphasis on equity, Todd Rose gave a talk that has now become a book; the theme was needing to move beyond the implicit notion of a typical, normal, or average student to fully embrace the diversity of how people learn [3]. Many presentations shared emerging forms of technology, such as expansion of making to include digital fabrics, tangibles and ink-based circuitry. With regard to learning, many presentations focused on how learners' *identities* changed as they participated in new experiences. The summit helped to define cyberlearning as tackling new ways of working with the diversity of students; exploring the new activities with forms of user experience; and as focused on newer theoretical constructs such as embodied learning and development of identity.

Since 2012, the Cyberlearning portfolio has grown to include over 250 projects. Prominent themes of Cyberlearning projects include mobile learning, bridging informal and formal learning, making and creating, citizen science, collaborative learning, embodied learning, data visualization, games and virtual worlds, augmented reality/immersive environments, virtual and remote labs, learning analytics and adaptive learning. This portfolio is already having an important impact in the United States – for example, it has been featured in the U.S. National Educational Technology Plan [6] to illustrate to educators how technology is moving beyond school installations of educational technology. In addition, following up on a recommendation in the task force report [1], the Center for Innovative Research in Cyberlearning (CIRCL, http://circlcenter.org) was created to serve as a community hub, similar to a European knowledge network like Kaleidoscope, Prolearn, or the other TEL-related coordination efforts. CIRCL acknowledges that research in the Cyberlearning portfolio has many parallels in European TEL work and thus is organizing a group of Cyberlearning researchers to attend EC-TEL to engage in scientific exchange.

## 2 Explorations of Immersion and Augmented Reality

Here we describe one fertile area that would be ripe for mutual exploration with European colleagues: immersive, augmented, and virtual reality projects. Individual projects in the Cyberlearning portfolio are exploring how technology can lead to experiences where students either feel more immersed in a context for scientific investigation or use technology to otherwise augment their actual context for learning.

In RoomQuake [4], students become immersed in a classroom-sized simulation of an earthquake. As the sounds of an earthquake play on speakers, the students can take readings on “seismographs” at different locations in the room, inspect an emerging fault line, and stretch twine to identify the epicenter. No real seismographs are used, rather tablet computers are used to simulate the measurement instruments and to reveal imaginary cracks in an otherwise normal classroom wall. Nonetheless, the experience is intense enough that students feel transported out of their classroom and begin working together like scientists in the field. Students must decide what to measure and how to analyze data in order to solve a challenging problem. In other classroom-scale immersive simulations, students travel inside a rocket to the moon or uncover an (imaginary) invasion of insects making a habitat in the walls of the classroom.

In contrast, in the “In Touch with Molecules” project [5] students manipulate a physical ball-and-stick model of a molecule such as hemoglobin, while a camera senses the model and visualizes it with related scientific phenomena, such as the energy field around the molecule. Students simultaneously see the molecule that they are physically moving and a visualization of the molecule on a screen, with colorful dynamic energy fields. Students’ embodied and tangible engagement with a physical model is thereby connected to more abstract, conceptual models, supporting students’ growth of understanding.

Whereas the first two examples take place in a school, the Connected Worlds [6] exhibit re-uses a large space remaining from the 1964 New York World’s Fair. Participants enter this space, which is now part of the New York Hall of Science (a science museum), and find a series of large screens simulating a set of connected ecological niches, each with fanciful simulated flora and fauna. The simulated work responds to how people move and gesture near the screens. For example, one full body gesture can cause a new tree to sprout. In addition, participants can move foam “rocks” and thus redirect the water supply to different ecological niches. As a consequence of these changing water available trajectories, life forms may die off, become more profuse, or migrate across the screens representing the ecological niches.

Other forms of augmented, immersive or virtual realities are also explored in Cyberlearning projects. In one project, students wear personal activity sensors and data flows to an online video game about health. Remote scientific laboratories are another type of virtual experience explored both in cyberlearning and European-based TEL projects. Multimodal input using a variety of sensors that can capture speech, body movement, touch, and other forms of expression and related emerging analytics techniques to interpret that data feature across many projects. In addition, projects explore how computer-generated output can be embedded in the real environment (as robots) or virtual environment (as avatars) in forms that do not seem as computer-like.

## 3 Discussion of Themes of Learning, Computation, and Equity

We anticipate that by sharing examples of cyberlearning research, and through learning about related EC-TEL research by participating at the conference, researchers from the United States and Europe will be able to engage on topics of mutual interest. For example, we have already had several successful exchanges between the US-based and

Israel-based researchers regarding virtual reality and augmented reality learning, and this has led to fertile discussion about "empathy," activity design, and desired platform capabilities. Three broad areas for discussion are:

1. **Diversity and Equity.** How can learning activities designed with emerging technologies enable new forms of participation and engagement that draw a broader population into opportunities for important learning?
2. **Forms of Interaction and Forms of Data.** What are the computational challenges in allowing activity developers to design new forms of interactive learning using these emerging capabilities (e.g. immersive, augmented, and virtual features)? How can we collect and work with the rich, multi-modal data that results?
3. **Frontiers for Learning Research.** What are the new research questions about learning that become important and addressable in these environments? What existing learning sciences methods and theories continue to be applicable, and how can research inform development of new theory or methodology development and growth?

**Acknowledgement.** This material is based upon work supported by the U.S. National Science Foundation under grants IIS-1233722, IIS-1441631, and IIS-1556486. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



## References

1. Borgman, C.L., Abelson, H., Dirks, L., Johnson, R., Koedinger, K.L., Linn, M.C., Lynch, C.A., Oblinger, D.G., Pea, R.D., Salen, K., Smith, M.S., Szaly, A.: Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge. NSF, Washington, D.C. (2008)
2. National Science Foundation: Cyberlearning: Transforming Education (2011). https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503581
3. Rose, T.: The End of Average. HarperOne, New York (2016)
4. Moher, T., Wiley, J., Jaeger, A., Silva, B.L., Novellis, F., Kilb, D.: Spatial and temporal embedding for science inquiry: an empirical study of student learning. In: Proceedings of the 9th International Conference of the Learning Sciences, vol. 1, pp. 826–833. International Society of the Learning Sciences (2010)
5. Davenport, J.: In touch with molecules (no date). http://molecules.wested.org/home/index.php
6. New York Hall of Science: Connected Worlds (no date). http://nysci.org/connected-worlds/

# Future Research Directions for Innovating Pedagogy

Jeremy Roschelle[1(✉)], Louise Yarnall[1], Mike Sharples[2], and Patrick McAndrew[2]

[1] SRI International, Menlo Park, CA, USA
{jeremy.roschelle,louise.yarnall}@sri.com
[2] The Open University, Milton Keynes, UK
{mike.sharples,iet-director}@open.ac.uk

**Abstract.** A series of reports on Innovating Pedagogy were launched in 2012 to look at the trends that show how practitioners may engage in innovation in pedagogy. This paper looks at the latest set of trends, and highlights four 2015 trends that seem particularly rich for researchers to explore in the next five years.

**Keywords:** Pedagogy · Educational technology · Innovation · Learning · Instruction

## 1 Introduction

Innovation is often associated with advances in technology, but approaches that make a profound change to education are usually based not on technology but on innovations in pedagogy for a technology-enabled and mobile world. Since the Innovating Pedagogy annual series was launched in 2012, over 30 different trends have been examined. This paper highlights four for research. Since December, the 2015 report has garnered more than 66,000 downloads from 128 countries. Fourteen researchers from The Open University (UK) and SRI International (US) contributed to the latest report.

### 1.1 2015 Practitioner Trends

The image in Fig. 1, produced by TeachOnline, summarizes the 2015 pedagogical trends for practitioners at a glance. For more detail, the reader may review the full report at www.open.ac.uk/innovating.

## 2 Four Promising Trends for Research

To reflect on the prospective future for learning and teaching in school and beyond, we selected four 2015 pedagogical trends that advance long desired pedagogical goals through the use of new technology: Incidental Learning, Context-based Learning, Embodied Learning, and Analytics of Emotions. Future research should focus on how


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 648–651, 2016.
DOI: 10.1007/978-3-319-45153-4_83

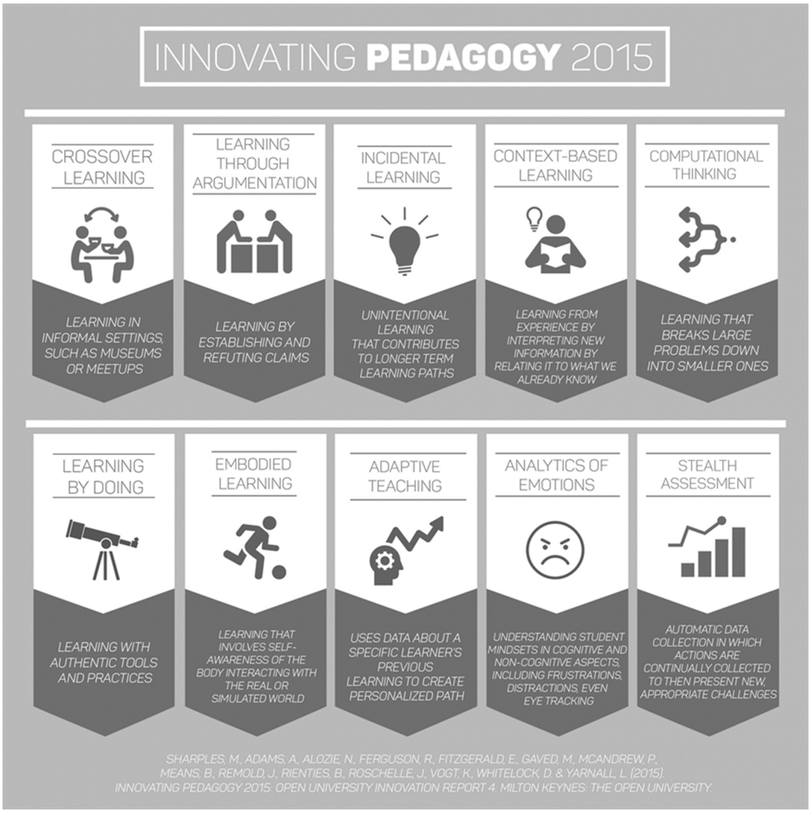



**Fig. 1.** The 2015 Innovating Pedagogy top 10 trends. Image credit: Stephen Valdivia of TeachOnline, the Arizona State University Instructional Design Community

all four involve intelligent technologies in delivering the most human and powerful features of pedagogy – mentoring, timely information presentation, and responsiveness to the learner's physical and emotional processes.

### 2.1 Overview

Education pioneer John Dewey wrote, "Such happiness as life is capable of comes from the full participation of all our powers in the endeavor to wrest from each changing situations of experience its own full and unique meaning," [1, p. 25]. To begin to wrest meaningful learning from technology-rich situations, we look to the four themes below.

**Incidental Learning.** Incidental Learning captures learning ephemera for productive use. It brings up pedagogies and technologies for noticing, reflecting, and connecting the unplanned learning that we experience daily. A mobile app may permit a learner to record a feeling or impression after an experience in the workplace, and then later refer back to recall, index, and share it. Learners may receive a text "nudge" to help maintain focus on an extended task, thus supporting memory, motivation, planning, revision, and mentoring. Future research drawing on behavioral economics and cognitive behavioral therapy may explore how learners may use technology to record instances of incidental learning so they can reflect on them and obtain social support around them. Theories of social-emotional learning, such as self-determination theory [2], growth mindsets [3], and self-regulation [4], offer a useful starting point for investigation.

**Context-Based Learning.** Context is both something we are immersed in and something we create. As technology becomes more embedded in life through the so-called Internet of Things, the opportunities for learning in context can be expected to increase. Contextual learning technology may bring realistic simulation into a classroom or project an instructional overlay onto the world through augmented reality in a mobile device. It capitalizes on the human capacity to see similarities and differences when the same process is applied in different settings and conditions. Future research may explore how such technologies can improve knowledge transfer by linking knowledge learned in school with knowledge gained from informal contexts [5]. Research may also examine how learners create context through interaction using technologies [6]. Also useful are theories of knowledge representation, symbol systems, and distributed cognition, particularly for designing augmented reality overlays projected on to an environment. One of the core challenges is to help learners regulate their access to these opportunities and to help educators be aware of when to switch distributed networks on and off as needed.

**Embodied Learning.** Embodied learning considers how the learner is engaged as a whole person with the learning process. As someone performs a task, new technology can focus attention and help link knowledge to activity, moving the learning from abstract to concrete action that embeds learning deeply. Such embodied learning is not a new concept [7], but technology supports how we can measure performance, as reflected in the use of Fitbits and health apps. Embodied learning presents analytics of both individual and collective activity, permitting comparisons for performance improvement. Future research may focus on designing for embodied experiences, technological transformation, and forms of feedback. This research can cross from neuroscience and educational design to new technologies.

**Analytics of Emotions.** Analytics of Emotions research identifies the emotions relevant to learning and develops the sensing technologies that can track and respond to learner emotions during online learning. A theme of research since the mid 1990s [8], it is now being extended into classrooms and informal settings. Early work focused on inferring a learner's motivational states from logs of online learning, but more recent studies track states with eye tracking, facial recognition and posture analysis. Such studies aim to help learners understand when they are struggling and need to seek help. This research builds understanding of how emotional constructs interact with attention, memory and understanding.

## 3 Conclusion

We have discussed the pedagogy of emerging innovation and we invite the research community to consider learning environments that anticipate incidental learning, support an interdependence of content and context, engage the integration of body and mind, and are responsive to learners' emotional states.

**Acknowledgements.** This material is based in part upon work supported by the National Science Foundation under Grant No. IIS-1233722. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.



## References

1. Dewey, J., Hickman, L., Alexander, T.M.: The Essential Dewey: Ethics, Logic, Psychology, vol. 1. Indiana University Press, Bloomington (1998)
2. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. **55**(1), 68–78 (2000)
3. Dweck, C.: Mindset: The New Psychology of Success. Random House, New York (2006)
4. Gollwitzer, P.M., Oettingen, G.: Planning promotes goal striving. Cogn. Physiol. Neurol. Dimens. **2**, 162–185 (2011)
5. Salomon, G., Perkins, D.N.: Rocky roads to transfer: rethinking mechanism of a neglected phenomenon. Educ. Psychol. **24**(2), 113–142 (1989)
6. Dourish, P.: What do we talk about when we talk about context. Pers. Ubiquitous Comput. **8**(1), 19–30 (2004)
7. Lakoff, G., Johnson, M.: Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought. Basic Books, New York (1999)
8. del Solato, T., Du Boulay, B.: Implementation of motivational tactics in tutoring systems. J. Interact. Learn. Res. **6**(4), 337–378 (1995)

# Platform Oriented Semantic Description of Pattern-Based Learning Scenarios

Zeyneb Tadjine(✉), Lahcen Oubahssi, Claudine Piau-Toffolon, and Sébastien Iksal

LUNAM University, University of Maine, EA 4023, LIUM, 72085 Le Mans, France
{zeyneb.tadjine,lahcen.oubahssi,claudine.piau-toffolon,
sebastien.iksal}@univ-lemans.fr

**Abstract.** In our research work, we address the issue of representing the learning scenario's concepts, in a learning platform. In this context, we have proposed a process for operationalizing pattern-based learning scenarios. We present the first two steps dealing with the new challenge of modeling deployable e-learning scenarios using Semantic Web technologies. It is primarily an ontology-based description of learning scenarios, which helps reducing the gap between human-readable and machine-readable vocabulary. We highlight the effectiveness of orienting teachers-designers, non-platform experts, toward creating adaptable and deployable learning scenarios. We defend that an assisted and platform-oriented design, allows the teachers to have a better pedagogical use of the embedded tools and features of learning platforms.

**Keywords:** Learning design · Learning platforms · Operationalization · Learning scenarios · Patterns

## 1 Aim and Motivation

Learning Management Systems (LMS) are more and more used by teachers, their use is not anymore restricted to content repository for distant learning [1, 2]. Nevertheless, we note that teachers find some difficulties using the LMS, especially when they are not platforms experts. The challenge is to easily master the process from the design to the operationalization of learning scenarios, for that, we believe that the operationalization of learning scenarios on LMSs is more than a technology-related question. Different research issues around instructional design are to be addressed in order to provide pedagogical expressiveness of the different elements within a learning scenario, while the design respects sufficiently the structure to describe the learning scenario [3, 4]. In our research, we seek to provide solutions to the problem of the automatic deployment of learning scenarios. We propose a process of the operationalization of pattern based learning scenarios [4]. The aim is to offer the use of a pattern formalism to create and edit learning scenarios, allowing the learning scenario design to be open enough to express teachers' concerns on one side and on the other side, structured enough to be machine interpretable for deployment purposes. In this paper we intend to mainly present a semantic model to help creating learning scenarios as a part of our process.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 652–655, 2016.
DOI: 10.1007/978-3-319-45153-4_84

## 2 Structuring and Indexing Platform-Based Learning Scenarios

In order to define a clear idea of the way to address the challenge of properly design platform-oriented learning scenarios, we investigated the benefits, as well as the issues, regarding using a pattern-based LD tool by teachers-designers [2]. We studied the learning scenario from two viewpoints: (i) Starting from teachers intentions going down to its representation on an LMS, (ii) using a pattern-based design. This study allowed us to identify the assets (patterns formalism requirements and ontological modeling) leading us to automate the deployment of learning scenarios. As a result we have settled on semantically modeling and mapping the double vision: human intentions and platform representations for guaranteeing to teachers-designers a design tool able to assist them in deploying their learning scenarios with less effort of manual adaptations. We also proposed a classification of the different approaches dealing with learning design [5–7], more specifically, those using ontologies as a semantic base to improve the learning process [8–11]. Although, all the effort made in developing systems to support the learning design process, literature has shown they had not yet reached a sufficient spread among teachers. We have noticed that most of the proposed design languages and tools do not preserve the semantic meaning of teachers' intention while transposing it on a LMS.

After that, we started collecting and structuring the available information and concepts related to the field of education [7, 12–15]. We were concerned only by the learning scenario's concepts necessary to its deployment, justified by the fact that we focus our research on platform-oriented learning scenarios. This step is very important, since it is a key solution to index learning platforms pedagogical language into a general semantic description of a platform-oriented learning scenario. Studying the existing learning design repositories and theories where instructional scenarios can be modeled, we defined a five levels structure of the learning scenario, which represents the structuring step of our process. We believe that the right set of abstractions will give more benefits to easily map the human design language to the machine interpretable one. We had to make sure that the technological tools will easily support our proposed model. For that, the other point was to study an example of a deployed learning scenario. It consists in peer assessment of a synthesis. The course covers most of the features that Moodle 2.4 includes. Next, we explain an extract of concepts of our structure. We managed to introduce the most relevant concepts to deployment goals. The first level formalizes the notion of "*learning scenario*" in terms of structure and content, based on the different definitions researchers assigned to learning scenario [14]. A scenario describes roles, activities and also knowledge resources, tools and services necessary to the implementation of each activity. From all this emerges the most used concepts that summarizes the essence of a deployable learning scenario: the learning scenario structure is what defines in a design any sequential ordering of activities, it is mainly inspired from [14] research work, it is defined by a set of three concepts: "*Structuration unit*", "*Activity sequence*" and finally, "*Elementary activity*". This model was implemented as an ontology-based e-learning scenario model, using the Protégé tool[1]. Besides increasing

[1] http://protege.stanford.edu/.

the level of sharing content between teachers-designers, the ontological description will help us to ensure the support of the technological aspect for a learning scenario. Ontology will help teachers-designers to formalize pattern-based scenarios with the editing tool conformed to the conceptual framework we proposed.

The concept of "*Elementary activity*" is assigned to a category from bloom's taxonomy [13], the categorisation will help the teacher to better create pedagogically reusable pattern-based learning scenarios, as well as it will help us to index it according to the most suitable platform tool. Any learning scenario has some necessary conditions and rules to be executed as teachers-designers intended, and since our learning scenarios are designed to be platform oriented in terms of design and deployment, we must take into account both the platforms and the pedagogical point of view. For that, we defined two sets of constraints. The first one concerns the human reasoning of the right conditions to manage the learning scenario, as for example the fact of restricting the access of an activity to the learners on the base of the previous activities results. The second set of constraints concerns the machine readable part of the scenario, although the previous ones are also machine interpretable, but they mostly relate to a pedagogical use, while the platform oriented set is fully built on computed learning environments. As we studied Moodle 2.4 platform, we retained the constraints adding a pedagogical dimension to the deployed scenario. We take the visibility constraint as an example, this added value allows the teacher to hide any activity for the learner until a time he judged suitable for his goals: it could be according to a score of a given evaluation, or a certain duration in time etc. We complete pedagogical goals and all others concepts describing evaluation in a learning scenario by all missing information needed to operationalize an evaluation based scenario. The agent of evaluation could be the teacher, the students, in case of a peer assessment, and even the learning platform itself in case of an auto evaluation. We also note that an evaluation activity is a set of some evaluation tools, helping the teacher to assess students according to their needs: graded assessment, auto-evaluation, paper exam, quiz, with or without feedback etc.

After identifying our structure, helping teachers towards a platform-oriented learning design, we must ensure the mechanism to automatically transform their pedagogical intentions into modules and content on the targeted learning platform. Next, we show through an example the way we deducted our manual ontology alignments between the semantic descriptions of our pattern-based learning scenario and Moodle's pedagogical embedded language. We started by transforming the metamodel into a semantic description; this is a very important phase because it is the first step toward a platform semantic description as a form of an ontology. In order to align our two semantic descriptions, we studied the example presented earlier about peer assessment in Moodle, adding to that our collaboration with a pedagogical designer, to come up with the right mappings of Moodle's tools and features. Starting from the most frequently functions required by teachers-designers, we grouped the set of offered tools as follows: collaborative work tools (glossary, journal, wiki, workshop etc.), synchronous and asynchronous communicative tools work tools (forum, chat, and survey), learning tools (lesson) and evaluative tools (assignments, workshop, quiz, etc.). We believe that this work has to be refined with teachers' experiences using learning platforms, thus we highlight again the importance of using a semantic description because it is extensible, and allows indexing and adding of more features as the technological updates are evolving around distance learning.

## References


1. El Mawas, N., Oubahssi, L., Laforcade, P.: Making explicit the Moodle instructional design language. In: 2015 IEEE 15th International Conference on Advanced Learning Technologies (ICALT), pp. 185–189. IEEE, July 2015
2. Oubahssi, L., Piau-Toffolon, C., Clayer, J.-P., El Mawas, N.: Design and operationalization of learning situations based on patterns for a public in professional development. In: Rensing, C., de Freitas, S., Ley, T., Muñoz-Merino, P.J. (eds.) EC-TEL 2014. LNCS, vol. 8719, pp. 580–581. Springer, Heidelberg (2014)
3. Prieto, L.P., Asensio-Pérez, J.I., Dimitriadis, Y., Gómez-Sánchez, E., Muñoz-Cristóbal, J.A.: GLUE!-PS: a multi-language architecture and data model to deploy TEL designs to multiple learning environments. In: Kloos, C.D., Gillet, D., García, R.M.C., Wild, F., Wolpers, M. (eds.) EC-TEL 2011. LNCS, vol. 6964, pp. 285–298. Springer, Heidelberg (2011)
4. Tadjine, Z., Oubahssi, L., Piau-Toffolon, C., Iksal, S.: A process using ontology to automate the operationalization of pattern-based learning scenarios. In: Zvacek, S., Restivo, M.T., Uhomoibhi, J., Helfert, M. (eds.) Computer Supported Education, pp. 444–461. Springer, Heidelberg (2015)
5. Laurillard, D.: Teaching as a Design Science: Building Pedagogical Patterns for Learning and Technology. Routledge, New York (2013)
6. Koper, R.: Current research in learning design. Educ. Technol. Soc. **9**(1), 13–22 (2006)
7. Koper, R.: Modeling Units of Study from a Pedagogical Perspective: The Pedagogical Meta-model Behind EML. Open Universiteit Nederland En ligne, Heerlen (2001)
8. Chimalakonda, S., Nori, K. V.: A patterns-based approach for modeling instructional design and TEL systems. In: 2014 IEEE 14th International Conference on Advanced Learning Technologies (ICALT), pp. 54–56. IEEE, July 2014
9. Fensel, D.: Ontologies, pp. 11–18. Springer, Heidelberg (2001)
10. Amorim, R.R., Lama, M., Sánchez, E., Riera, A., Vila, X.A.: A learning design ontology based on the IMS specification. Educ. Technol. Soc. **9**(1), 38–57 (2006)
11. Mizoguchi, R., Bourdeau, J.: Using ontological engineering to overcome common AI-ED problems. J. Artif. Intell. Educ. **11**, 107–121 (2000)
12. Paquette, G.: A competency-based ontology for learning design repositories. Int. J. Adv. Comput. Sci. Appl. **5**(1), 55–62 (2014)
13. Krathwohl, D.R.: A revision of Bloom's taxonomy: an overview. Theor. Pract. **41**(4), 212–218 (2002)
14. Lejeune, A., Pernin, J.P.: A taxonomy for scenario-based engineering. In: CELDA, pp. 249–256, December 2004
15. Churchill, D.: Towards a useful classification of learning objects. Educ. Tech. Res. Dev. **55**(5), 479–497 (2007)

# Model of Articulation Between Elements of a Pedagogical Assistance

Le Vinh Thai(✉), Stéphanie Jean-Daubias, Marie Lefevre, and Blandine Ginon

Université de Lyon, CNRS, Université Lyon 1, LIRIS, UMR 5205, 69622 Lyon, France
{le-vinh.thai,stephanie.jean-daubias,marie.lefevre,
blandine.ginon}@liris.cnrs.fr

**Abstract.** The AGATE project proposed the SEPIA system that allows an assistance designer to define assistance systems added in target applications. In Interactive Learning Environments, such assistance systems are useful to promote the acquisition of knowledge. These assistance systems consist of a set of aLDEAS rules. Our study of assistance in existing applications shows that the articulation between the rules of assistance can take many forms. We propose and implement a model of articulation between assistance rules with the five modes of articulation that we have identified. This model makes explicit and facilitates the definition of articulation between the rules of an assistance system.

**Keywords:** User assistance · Pedagogical assistance · Epiphytic approach · Mode of articulation

## 1 Introduction

More and more applications are used in different contexts: professional, personal and educational. However, because of handling difficulties, users can under-exploit application or abandon it, and lose their motivation. In ILEs (Interactive Learning Environments), learners use various applications to acquire knowledge, but technical difficulties can compromise this acquisition. Additionally, some applications don't meet the teachers' pedagogical goals. Adding an assistance system is considered as a solution for both technical and pedagogical problems of an existing application. Such pedagogical assistance systems consists in varied and complex assistance actions (explanation message, error detection, etc.). They can have different modes to sequence assistance events which describe the articulation between the assistance elements. For instance, successive assistance gives *one message after another* in order to guide learners.

The SEPIA system [1] allows assistance designers (teachers in the pedagogical context of this paper) to add an assistance system in an existing ILE by creating and executing aLDEAS rules [2]. SEPIA is a full solution to create rich assistance systems. However, the definition of the articulation between the assistance elements is still implicit and difficult in our system. So, this paper presents the evolution that we proposed and implemented into SEPIA to overcome these limitations.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 656–659, 2016.
DOI: 10.1007/978-3-319-45153-4_85

## 2 SEPIA System

The **AGATE project** (Approach for Genericity in Assistance To complEx tasks) aims at proposing generic models and unified tools to enable the setup of assistance systems in various existing applications, that we call target-applications, through a generic and epiphytic approach [2]. Within this project, the **SEPIA system** [1] implements this approach in two tools: an assistance editor and an assistance engine. The assistance editor allows assistance designers to define an assistance system; while the assistance engine executes this assistance system to provide assistance to final users in the target application.

The **aLDEAS language** (a Language to Define Epi-Assistance Systems) [2] is proposed in order to connect these both tools. The assistance systems are defined by a set of aLDEAS rules. An aLDEAS **rule** begins with an event wait called **trigger event**. When this event occurs, the **assistance actions** are immediately launched (see upper path in Fig. 1), or constrained by a **condition** (see lower path in Fig. 1). This condition takes the form of a consultation associated with different alternatives, each associated with one or more actions. The rule can be terminated by an **end event** that ends all actions launched by this rule. For instance, Fig. 1 shows one of the rules that define an assistance system. This rule waits until a click on the button ‘help’ in order to verify the answer of the learner and to provide an error message when this answer is not correct (text written by learner is not equal to 1). This message is closed after 10 s.

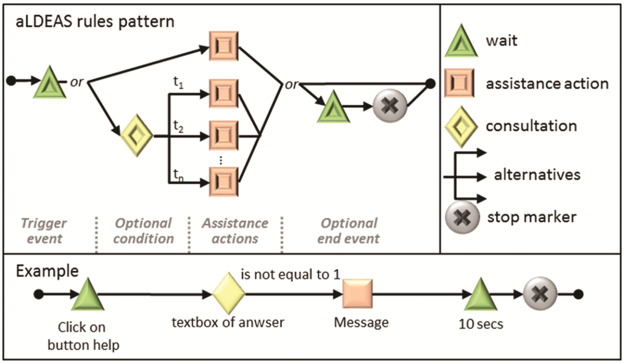



**Fig. 1.** aLDEAS rules pattern [2]

## 3 Modes of Articulation Between Assistance Elements

In ILE, pedagogical assistance can be found in some applications. This assistance can be executed according to different modes in sequencing assistance events. These modes describe articulation between the different assistance elements. Through a study of numerous applications, we identified five modes of articulation between assistance elements: independent, simultaneous, successive, progressive and interactive [3]. In the *independent mode*, an assistance element is given independently from another. In the *successive mode*, the assistance elements are given one after the other. In the *simultaneous mode*, all assistance elements are given at the same time. In the *progressive mode*, the given assistance elements are more and more detailed and concrete. In the *interactive mode*, the given the assistance elements depends of information such as the application state, the user profile or the choice of user.

## 4 Model of Articulation Between ALDEAS Rules

If aLDEAS language and its implementation in SEPIA already allow the definition of articulation between assistance elements such as those presented in the Sect. 3. An assistance system is currently always defined in SEPIA by a set of same level aLDEAS rules. In the aLDEAS rules pattern (Fig. 1), trigger event, end event and trigger condition are central elements to form the articulation between rules. On the one hand, we must carefully define elements in the rules in order to ensure correct articulation between them. On the other hand, we must examine them in order to understand which mode of articulation to choose. So, this articulation between rules is only implicitly expressed and is complex to define with aLDEAS.

For these reasons, we propose to complete aLDEAS language by a model of articulation between assistance rules. To simplify the representation of the model, we note that rules between which we want to make an articulation are named $R_i$ with $i \in [1, n]$, $(n \geq 2)$. The representation of our model (Fig. 2) gives an overview of the five modes of articulation that we identified from a study of existing works: independent, successive, simultaneous, progressive and interactive. In each mode of articulation, there are constraints that rules must respect to ensure the correct articulation between them (for instance, for successive mode, each rule should be launched by the end of the previous rule). These constraints are represented by the aLDEAS rules.

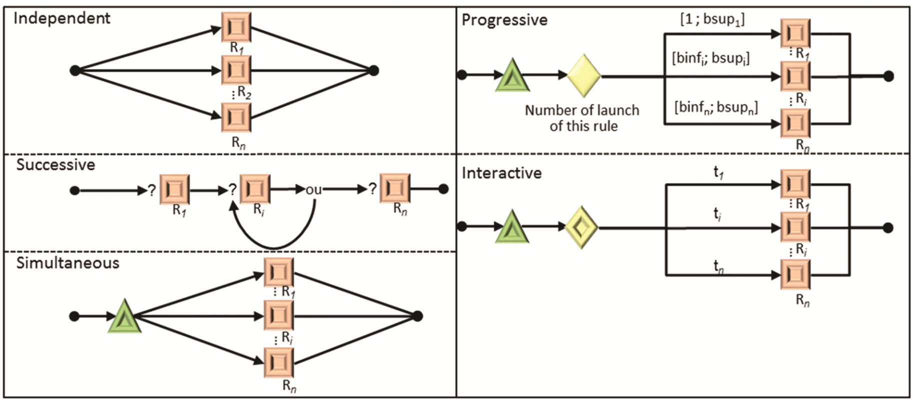



**Fig. 2.** Model of articulation between aLDEAS assistance rules

Let's take the example of an assistance system only consisting of the three steps of a tutorial. This assistance is created through three rules articulated in successive mode (defined in Fig. 2 and in more detail in Fig. 3). This mode forces the previous rules to end with any event and the next rules to start at the end of the previous rules. Thus, the three rules in this example respect these constraints of successive mode. The first rule $R_1$ waits until a user click on button "Tutorial" in order to show a welcome message that will be closed after 10 s. The rule $R_2$ that waits until the end of $R_1$ shows a message explaining a first part of the screen that will be closed after 10 s, etc.

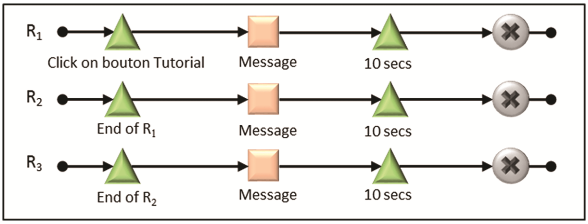



**Fig. 3.** Detail of three rules articulated in successive mode in aLDEAS

## 5 Conclusion and Future Work

In this article, we presented our model of articulation between the rules of an assistance system which completes the aLDEAS language. This model explicitly express the notion of articulation between rules of an assistance system. It offers five modes of articulation corresponding to those we have identified in our bibliographical study. We implemented this model in the SEPIA system by adding the notion of block of rules articulated in a mode. This implementation has two main advantages: it makes explicit the definition of blocks of rules within a graphical interface and it applies semi-automatically constraints on rules, which simplifies the user's work [3]. With the introduction of this model in our approach, an assistance system is defined not only by a set of rules, but also by a set of blocks that explain the articulation between these rules. We evaluated our propositions by experiments that confirmed their potentials [3].

However, an assistance system can be described by many blocks of rules articulated in different modes. Therefore, in the future, we will aim at a global graphical representation of assistance systems which will allow to show many blocs at the same time.

## References

1. Ginon, B., Jean-Daubias, S., Champin, P.-A., Lefevre, M.: Setup of epiphytic assistance systems with SEPIA. In: EKAW, pp. 1–4. Linkoping (2014)
2. Ginon, B., Jean-Daubias, S., Champin, P.-A., Lefevre, M.: aLDEAS: a language to define epiphytic assistance systems. In: Janowicz, K., Schlobach, S., Lambrix, P., Hyvönen, E. (eds.) EKAW 2014. LNCS, vol. 8876, pp. 153–164. Springer, Heidelberg (2014)
3. Thai, L.V., Jean-Daubias, S., Lefevre, M., Ginon, B.: Model of articulation between aLDEAS assistance rules. In: DCCSEDU. Roma, Italy (2016)

# Simulation-Based CALL Teacher Training

Ilaria Torre[1(✉)], Simone Torsani[2], and Marco Mercurio[2]

[1] Department of Computer Science, Bioengineering, Robotics, University of Genoa, Genoa, Italy
ilaria.torre@unige.it

[2] Department of Modern Language and Culture, University of Genoa, Genoa, Italy
{simone.torsani,marco.mercurio}@unige.it

**Abstract.** Technology-enhanced language learning enables activities to be adapted to several factors including technological constraints and students' special needs. To train language teachers to design goal-based learning activities that satisfy diverse student needs, learning platform features and device constraints we designed an ontology-based simulator. The tool generates usage scenarios that trainee teachers have to solve, moreover it provides recommendations about the most suited instructional design solutions given multiple constraints. In this paper we present the approach and we describe the exploratory evaluation we performed on a demonstrator. Results confirm the validity of the approach and the ability of the tool to generate realistic scenarios.

**Keywords:** Technology-enhanced learning · Adaptable instructional design · Ontological reasoning · Recommender systems

## 1 Introduction and Background

Teacher Education is an important area of research and practice within Computer Assisted Language Learning (CALL). A major theme in CALL Teacher Education (CTE) is the complexity of exploiting technology in order to enhance language teaching. It implies choosing one or a set of tools and knowing how to configure and to use them to achieve a certain goal, given a certain context and certain learners.

Research on CTE, therefore, has focused on the development of transferable integration skills, i.e. more abstract skills that enable teachers to evaluate, choose and exploit different technologies in different contexts [1]. A brief look at CTE literature will reveal that trainers favour experiential and reflective learning (e.g. [7]) instead of simple transmission of facts and guidelines in instructor-driven setups. In the field of Second Language Teacher Education, Ellis [2] distinguishes between two general areas in training activities: experiential and awareness-raising. The former provides teachers with a direct experience of what happens in actual contexts, the latter helps them to achieve a deeper understanding of that experience. This same distinction is maintained in CTE research and practice. It is assumed that, by dealing with complex and realistic situations, teachers achieve a deeper understanding of the factors behind the work with technology and, as a consequence, develop integration skills. Simulation is an effective method to practice with complex and realistic situations [5, 6].


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 660–664, 2016.
DOI: 10.1007/978-3-319-45153-4_86

Following this line, our research focuses on the design and development of a simulator for instructional design. The tool exploits an ontology to generate scenarios in which teachers must find a viable process of solution given a number of constraints. The tool can be used for training but also to get recommendations.

There are different experimentations on simulated environments in teacher training. Foley and McAllister [3] illustrate Sim-school©, a tool which can generate different virtual student profiles teachers have to deal with to learn the complexity of curriculum design. Also Girod & Girod make use of a web-based simulation tool in which the tool generates different student profiles, which will differently respond to the teacher's choices [4]. The main contribution of our approach is to design a simulator that exploits the semantic relations among learning technologies and mental/physical operations which link a learning goal to a learning technology. This is new in CTE literature and is a challenging task for knowledge representation.

## 2 Semantic Modeling of Knowledge for Scenario Generation

In this section we briefly describe our L-max ontology used by the simulator to: (i) generate realistic scenarios with options that teachers have to set to find a solution, (ii) provide feedbacks about teachers' choices and errors. It extends and relates well-established data models and frameworks in diverse domains: language learning, device constraints, environmental conditions and learning disabilities. *L-max* stands for *Learning maximization given goals and constraints*.

The language learning part of the ontology is based on the Common European Framework of Reference for Languages (main classes: `Ability`, `Competence`, `Activity Type`). We have extended it with a set of classes and relationships that formally represent the design and development of learning activities and exercises (main classes: `MentalOperation`, `PhysicalOperation`, `CorrectionType`, `ActivityEnvironment`, `Tools`).

As concerns the device features and environmental conditions, we have integrated our ontology with the data model defined in the GPII/Cloud4all framework (gpii.net).

Finally, a part of the ontology addresses the learner profile and learning disabilities. To model this part we have exploited the ICD v. 10 classification from the World Health Organization (http://who.int/classifications/icd/) which is used by several Governments as reference model for learning diseases (main classes: `ICDClass`, `CompensationTools`, `DispensationMeasures`).

## 3 Empirical Evaluation

**Objectives and Method.** The study described in this paper concerns the early phases of development life cycle. It was undertaken with the main objectives of investigating the perceived utility of the training tool and gaining a first validation of the scenarios generated by the simulator, based on the L-max ontology.

For the evaluation, we developed a demonstrator which implements a set of learning goals and constraints, generating different scenarios. Its objective is to show the functionalities and potential uses of the simulator, not to evaluate its usability.

1. Participants taking part in the study were asked **to perform ten tasks on given scenarios**. During the tests, an observer annotated problems and difficulties experienced by the subjects and their execution time to complete the tasks.
2. Moreover, participants were invited **to complete a questionnaire** with questions presented after each task.

**Subjects.** We used two groups of 10 language teachers: the former was composed of ICT-skilled teachers and the latter of teachers without specific ICT competencies.

**Scenario Generation and Task Description.** For the evaluation, we developed a web-based demonstrator named SIMUL-TOOL. Each scenario is defined as $S_i = (L_i, C1_i, C2_i)$ that is Scenario i-th is composed of a Language Learning Goal and two Constraints. Below we describe one of the ten scenarios generated for the evaluation.

| |
|---|
| $L_1$ = designing an activity that develops the competence of: *understanding brief and simple texts*<br>$C1_1$ = developing a learning activity which *allows automatic correction*<br>$C2_1$ = the learning activity *has to be accessed by students through heterogeneous mobile devices* so that they can perform the activity anytime and anywhere. |

The *TEACHER*, on the basis of her/his language teaching competence, has to:

(a) think about an `activity type` (e.g. a quiz, a cloze test, an association activity, a thematic discussion) that can develop the specified `competence`,
(b) choose from the simulator interface the *ingredients* for developing the activity, respecting the constraints. The ingredients are the instances of the `PhysicalOperation` class of the ontology (e.g., creating an audio track for an exercise, creating a test, creating a forum, creating an account for a platform),
(c) ordering the ingredients based on their prerequisites.

*SIMUL-TOOL* performs a first evaluation, explaining whether the selected ingredients allow to develop the specified competence and whether the selected ingredients enable the delivery of an activity that respects the specified constraints. If the ingredients selected by the teacher are not correct wrt the learning goal or wrt their ordering, the simulator provides a feedback and requires the teacher to take a backward step and make new selections.

If all the selections are correct, the *TEACHER* is presented with a new set of options, based on the `Tool` class of the ontology. The available options for the demonstrator are: Moodle, Edmodo, and Hot Potatoes.

*SIMUL-TOOL* performs a new evaluation of the selected options querying the ontology to discover if there are any relations among the selected ingredients and tools chosen by the teacher and if they respect the constraints in the task set up.

After the subject has identified a proper combination and a right sequence of instructional design components that allow to develop the learner competence in *understanding brief and simple texts* and in the meanwhile satisfy the automatic correction and the multi-device delivery constraints, SIMUL-TOOL re-composes all the steps to be performed to develop the activity.

**Questionnaire.** The perceived effectiveness of the tool was assessed through a questionnaire. After each task, the teacher had to answer to these questions:

- Q1. *Do you consider the proposed task as realistic with regard to language education? (Answers are provided on a 4-point Likert scale 0-3.)*
- Q2. *Does the solution you initially thought of match with the solution proposed by the simulator? (Range 0-1).*
- Q2.a. *In case it didn't. Do you consider the proposed solution better than the one you had thought of? (Range 0-1).*
- Q2.b. *In case it didn't. Did the simulator provide enough options to develop your solution (i.e. could you implement your solution with the available options)? (Range 0-1).*

**Results. G1** = group of ICT-skilled teachers, **G2** = non-skilled ones. Concerning Q1, 70 % of the population ($n = 10 \in G1$ and $n = 10 \in G2$) considered the proposed tasks realistic (score $\geq 2$). Notice in particular that 84 % of G1 teachers considered the scenarios realistic, with a Mean = 2,4 and Standard Deviation = 0,712. This is a relevant result since G1 represents the expert group and this provides a first validation of the scenarios generated from the L-max ontology. Notice moreover that the difference between the two groups is statistically significant with p-value < 0.01.

**Table 1.** Results for Q2, Q2a and Q2b

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Q2 | Mean= | 0,59 | Mean per group | G1= | 0,76 | G2= | 0,42 |
| Q2a | Mean= | 0,72 | Mean per group | G1= | 0,64 | G2= | 0,79 |
| Q2b | Mean= | 0,42 | Mean per group | G1= | 0,33 | G2= | 0,52 |

Answers to Q2 are reported in Table 1. An average near to 60 % of the teachers thought of a solution similar to the one proposed by the tool, with G1 = 76 %. This suggests that the simulator is generally considered credible as regards both tasks and solutions. Q2a confirms this result with 72 % of teachers considering the proposed solution better than their own. However, it is interesting to underline that G1 shows an average lower than G2. We can argue that G1 teachers, being used to integrate technology in their learning activity, have thought of solutions that the demonstrator does not include yet. Q2b confirms this hypothesis since only 33 % of G1 consider the options provided by SIMUL-TOOL enough to develop their solution. This is in line with an early version of development and provides confirmation of the validity of the approach and of the underlying data model.

## References

1. Chao, C.C.: Rethinking transfer: Learning from call teacher education as consequential transition. Lang. Learn. Technol. **19**(1), 102–118 (2015)
2. Ellis, R.: Activities and procedures for teacher training. ELT J. **40**(2), 91–99 (1986)
3. Foley, J.A., McAllister, G.: Making it real: sim-school a backdrop for contextualizing teacher preparation. AACE J. **13**(2), 159–177 (2005)
4. Girod, M., Girod, G.R.: Simulation and the need for practice in teacher preparation. J. Technol. Teach. Educ. **16**(3), 307–337 (2008)
5. Hixon, E., So, H.J.: Technology's role in field experiences for preservice teacher training. Educ. Technol. Soc. **12**(4), 294–304 (2009)
6. Lateef, F.: Simulation-based learning: just like the real thing. J. Emergencies Trauma Shock **3** (4), 348–352 (2010)
7. O'Dowd, R.: Supporting in-service language educators in learning to telecollaborate. Lang. Learn. Technol. **19**(1), 63–81 (2015)

# Adaptable Learning and Learning Analytics: A Case Study in a Programming Course

Hallvard Trætteberg, Anna Mavroudi(✉), Michail Giannakos, and John Krogstie

Norwegian University of Science and Technology, Trondheim, Norway
{hal,anna.mavroudi,michailg,john.krogstie}@idi.ntnu.no

**Abstract.** The focus of this case study is the exploitation of visual learning analytics coupled with the provision of feedback and support provided to the students and their impact in provoking change at student programming habits. To this end, we discuss mechanisms of capturing and analysing the debugging habits and the quality of the design solutions provided by the students in the context of an object-oriented programming course. We instrumented the programming environment use by the students in order to track the student behavior and visualize metrics associated with it, while the students developed programs in Java.

**Keywords:** Adaptation · Learning analytics · Object-oriented programming · Student behavior · Programming habits

## 1 Introduction

It has been argued that "despite our best efforts as educators, student programmers continue to develop misguided views about their programming activities, particularly during freshman and sophomore courses" ([2], p. 26). To this end, there is a need to disengage students from the trial and error approach into practicing problem-solving strategies coupled with reflection and this is challenging also due to the fact that typical programming assignments are poor in promoting reflective mode on behalf of the student [2]. In the case discussed herein, problem-based learning was fostered in the context of a sophomore course on object-oriented programing (TDT4100). In the TDT4100 course, the students' abilities, aspirations and motivation are quite diverse. Their interest in and willingness to struggle with programming and debugging varies considerably, and our hypothesis is that this affects their habits when working on the programming assignments. Hence, offering rich feedback that allows practitioners to tailor their instruction, and providing insight into students' behavior scaffolds the teaching-learning experience. To this end, a digital environment for programming is exploited which is augmented by Learning Analytics (LA) while fostering adaptability. The purpose of this augmentation is to record, visualise and empower students reflect on their programming habits, focusing on debugging their programs.

LA can help to track the student progress over time and empower both students and tutors to make well-informed and evidence-based decisions. It has been suggested that among the factors that are driving the development of LA is the emergence of "big data" and the increase on the uptake of Virtual Learning Environments (VLEs) [3]. On the


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 665–668, 2016.
DOI: 10.1007/978-3-319-45153-4_87

other hand, a related challenge is that, although student-tracking is typically included in VLEs today, an in-depth reporting and visualisation of the built-in analytics that would help optimize opportunities for online learning and extract value from "big data" has often been rudimentary or practically non-existent [3]. To this end, our LA approach (see Sect. 2) captures and visualises all the key indicators of student behavior focusing, in particular, on their debugging habits with the aim of provoking reflection and remedial actions that can destabilize bad debugging habits.

## 2 Context and Approach

The TDT4100 course is an introductory course in object-oriented programming with Java. It is followed by approximately 600 students each year, typically in their second semester of study. To qualify for the exam, each student must earn 750 points from 10 assignments, each worth 100 points. The assignments are composed of smaller exercises, and the student typically needs to complete two or three of them each week. To address the diversity in motivation, the exercises have varying difficulty level, and the students can select among them according to their own skill and aspirations. Most of the exercises have pre-written tests (JUnit tests) for the programming snapshots (Java classes) that students are required to write. The success (or failure) of the tests gives them some feedback about their progress (or lack thereof) in order to scaffold the learning experience and make it easier to incrementally work towards the assignments' learning goals. We encourage the students to validate that their code works as expected before running the JUnit tests, but since assignment points are given based on the test results, many end up focusing too much on the tests, rather than on the stated requirements. Students use the Eclipse platform as an integrated development environment for which we provide learning resources and support. To collect data about the students' actual working habits, including how they use the tests and debug their code, we have created two Eclipse plug-ins that can log data about the use the Eclipse platform: files saved, number of errors

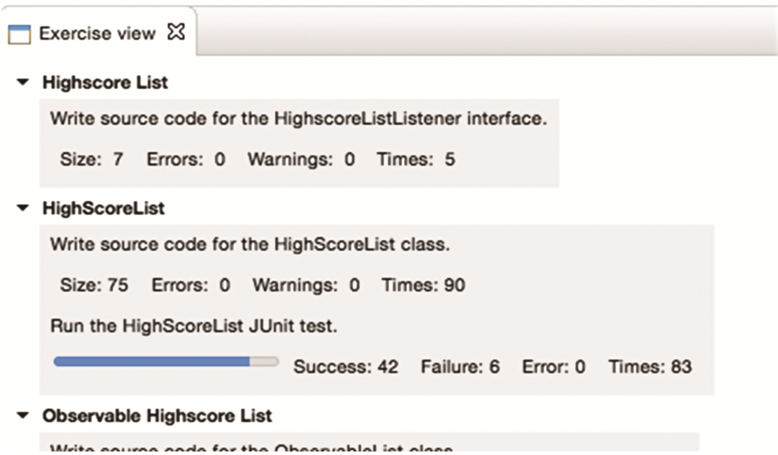



**Fig. 1.** The exercise view providing student feedback (Color figure online)

and warnings in the code, launching of their own code (the standard Java main method), the JUnit test results (success, failure or error), activation of Eclipse perspectives and views (e.g. for debugging), debugging events (e.g. stopping on break points or resuming execution), and execution of commands (e.g. stepping through code). Using the plugins, the student receives real-time feedback: an Eclipse view that shows the current student status, and a plot of the history of the student behavior in the programming environment (see Figs. 1 and 2 respectively). In Fig. 1, the blue line exemplifies the progress bar that indicates the student progress in this specific programming exercise. The history plot (Fig. 2) is indicating student code growth over time, how the success (or failure) of tests change over time and time periods of debugging.

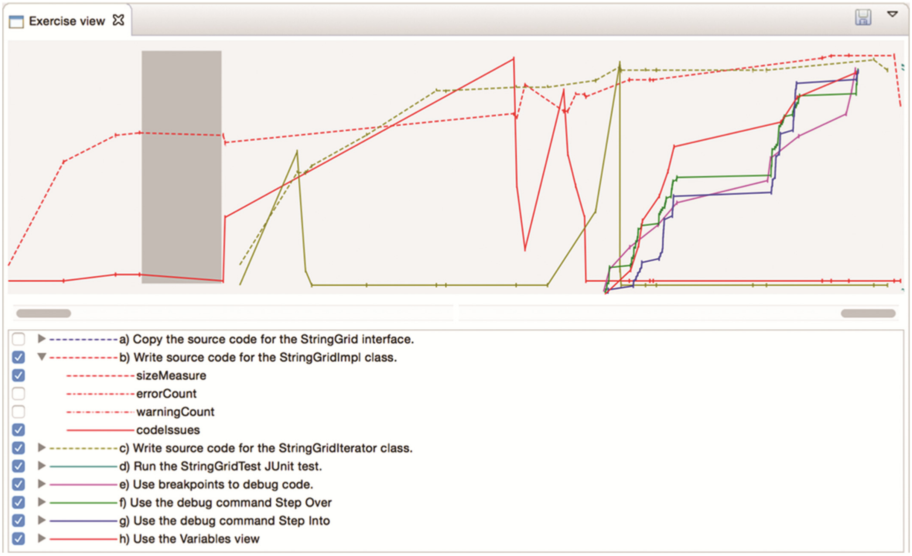



**Fig. 2.** Visual learning analytics in the programming environment

The history plot visualises how various metrics changes over time (x-axis). To reduce the horizontal extent, periods of inactivity above a certain threshold (e.g. lunch break or night) are condensed and shown with a darker, shaded background. The student can customise this visual learning analytics view and adjust it to her informational needs: to focus on certain data specific curves can be turned on or off and the student can zoom in interesting time intervals. To allow some level of exploration, the student can enter expressions over existing values that are shown as additional curves, like the ratio of test success and sum of test failure and error. The y-axis depends on the kind of data shown, hence the plot's focus is the trend, rather than absolute values; specific data points are shown when hovering over them (e.g. one can see how editing, testing and debugging activities are alternating and when measures of progress increase or decrease).

## 3 Conclusion and Discussion

Adaptive learning and learning analytics can inform each other, since they both cater for learners' variability and diversity. Yet, new methods of reporting and visualising analytics are needed which "are personalised, can be easily understood by learners and are clearly linked with ways of improving and optimising their learning" ([4], p. 314).

A recent review of the literature [5] revealed that the vast majority of the interventions that revolve around the combination of adaptive learning and learning analytics focus on student competences merely in terms of knowledge acquired. Only a small number focuses on the acquired skills, while none of the interventions in the area focuses on student learning in terms of attitude change, like change of programming habits; something that signifies the added value of this intervention.

In addition, it has been mentioned in the literature that "there is a lack of tools in programming courses that focus on amplifying learning opportunities and support learning activities" ([1], p. 24). Preliminary discussions with the tutor of the TDT4100 course revealed that the visual LA (Fig. 2) plot can indicate how each student progresses across three stages of development: (1) code authoring, where the size of the code (red dashes) grows considerably, (2) debugging, where the code is edited but does not grow much, and is run in debugging mode and (3) finalization, where bugs are discovered and fixed and tests begin to run successfully.

**Acknowledgements.** The work presented is supported by the European Research Consortium for Informatics and Mathematics (ERCIM). Contract Nr. 2015-07.

## References

1. Awasthi, P., Hsaio, I.-H.: INSIGHT: a semantic visual analytics for programming discussion forums. In: Proceedings of the First International Workshop on Visual Aspects of Learning Analytics, vol. 1518, pp. 24–31 (2015)
2. Edwards, S.H.: Using software testing to move students from trial-and-error to reflection-in-action. ACM SIGCSE Bull. **36**(1), 26–30 (2004)
3. Ferguson, R.: Learning analytics: drivers, developments and challenges. Int. J. Technol. Enhanced Learn. **4**(5–6), 304–317 (2012)
4. Lee, J., Park, O.: Adaptive instructional systems. In: Spector, J.M., Merill, M.D., van Merrienboer, J., Driscoll, M.P. (eds.) Handbook of Research for Educational Communications and Technology, pp. 469–484. Routledge, Taylor & Francis Group, New York (2007)
5. Mavroudi, A., Giannakos, M., Krogstie, J.: Insights on the interplay between adaptive learning and learning analytics. In: 16th IEEE International Conference on Advanced Learning Technologies – ICALT 2016, 25–28 July, Austin, Texas, USA (2016, accepted)

# Recommending Physics Exercises in Moodle Based on Hierarchical Competence Profiles

Beat Tödtli(✉), Monika Laner, Jouri Semenov, and Beatrice Paoli

Swiss Distance University of Applied Sciences, 8105 Regensdorf, Switzerland
{beat.toedtli,monika.laner,jouri.semenov,beatrice.paoli}@ffhs.ch

**Abstract.** We present a prototype for an adaptive navigation system which uses a dissimilarity measure between student and exercise profiles to rank and recommend exercises. Both types of profiles are structured as a hierarchical tree. We are developing a Moodle plugin that presents the top-ranked exercises as recommendations to distance learning students. A visualization of the student competence profiles provides progress feedback within the plugin.

**Keywords:** Adaptive navigation support · Progress-based adaptation · Recommender systems · Moodle

## 1 Introduction

Technology Enhanced Learning (TEL) aims to design, develop and test socio-technical innovations that enhance learning practices of individuals and organizations [1]. Learning takes place in many different settings and web systems are adapted to a heterogeneous set of user needs. Adaptive educational hypermedia overcome the "one-size-fits-all" problem by changing their characteristics according to the learner's needs and offering adaptive navigation support [2]. The increasing role of recommender systems for TEL evidences a growing interest in their development and deployment [3]. Nevertheless traditional open-source learning management systems such as Moodle lack personalization and adaptivity [4].

The novelty of this work is the presentation of a Moodle plugin prototype that provides personalized, progress-based exercise recommendations for navigation support. The selection is based on a dissimilarity measure between hierarchical exercise profiles and student competence profiles. This system is being developed in the context of an adaptive learning project at our distance learning university, where in-class guidance is provided regularly but not frequently. The system uses concept-based adaptation with a domain model for students and exercises. The concept-based approach is known to be very powerful and is able to achieve precise adaptation [5]. The prototype is currently designed for an undergraduate course in physics but will be extended to other courses and subjects.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 669–672, 2016.
DOI: 10.1007/978-3-319-45153-4_88

## 2 Architecture of the Exercise Recommender

The main learning goals of the physics course are organized in a hierarchical tree structure where all leaf nodes have the same depth. The exercises usually cover one or two physics topics and ask for a numerical answer. The nodes in the learning goal tree are weighted according to their relevance. Each parent learning goal is detailed by child learning goals and the sum of their weights equals the weight of the parent node. A teacher is directly involved in the adaptation process by specifying the weights of the leaf nodes. The exercise and student tree profiles are derived from this prioritized learning goal structure (see Fig. 1). The exercise profile node obtains topic relevance values indicating the relevance of the exercise for each of the node's learning goals. An exercise may touch several topics in the learning goal tree that may or may not be children of the same parent node. The student profile nodes obtain competence values indicating the competence of the student with respect to each node's learning goal. Successfully solved exercises increase the competence node value of the student profile in proportion to the learning goal weight. When the student fails an exercise, the competence node values are left unchanged because the student has not shown any additional competence and because topics already worked on should not be prioritized over (possibly many) topics that the student has not yet studied.

Based on this tree structure the recommender suggests exercises whose topic relevance profiles are maximally complementary to the competence profile of the student. The suggested exercises help the student to learn topics not yet

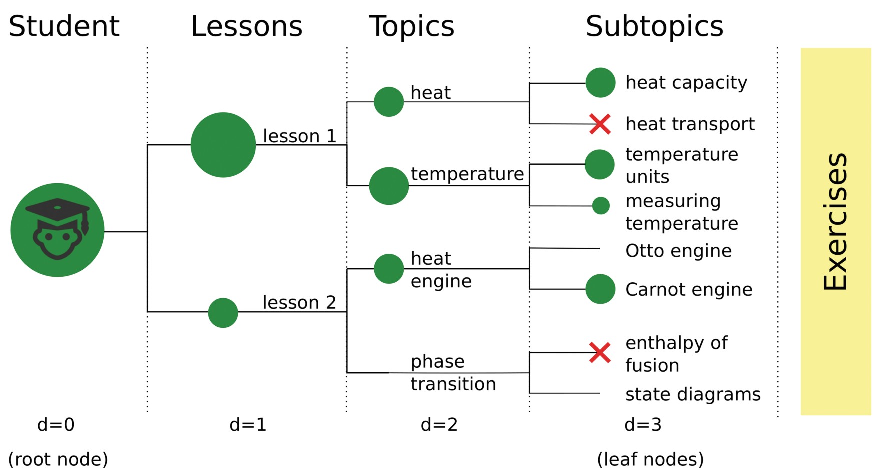



**Fig. 1.** Schematic representation of the hierarchical tree structure of the learning goals and the superimposed student profile. The successfully solved exercises contribute to the student's progress as indicated by the size of the green spheres while the failed exercise is marked with a red cross. $d$ indicates the depth of the node. (Color figure online)

mastered or attempted. In that way, the recommender guides the student to effective learning content.

Recommendations are computed using a dissimilarity measure between the student's profile tree $s$ and each exercise profile tree $e$ as follows. The profile trees are each converted into several lists $s(d)$ and $e(d)$ of node values at depth $d$ of the student and exercise profile tree, respectively. A dissimilarity measure $\mathrm{dis}(s, e, d) = -s(d) \cdot e(d)$ measures the dissimilarity of a student and exercise profile tree at depth $d$. For each exercise and student a dissimilarity vector $\mathrm{dis}(s, e) = (\mathrm{dis}(s, e, 1), \mathrm{dis}(s, e, 2), \ldots, \mathrm{dis}(s, e, d))$ is computed. For each student $s$ the vectors $\mathrm{dis}(s, .)$ of all exercises are arranged as rows in a table. The rows are first sorted with respect to the dissimilarity at the lowest depth $d = 1$ (column 1), and then with respect to increasing depths $d = 2, 3$, etc. (columns 2, 3, etc.). The exercises corresponding to the top $k$ (usually 5) entries in this list are recommended to the student. This procedure prioritizes dissimilarity at depth $d = 1$ over dissimilarity at a higher depth. Thus, exercises may be similar with respect to more general topics but dissimilar with respect to more specific ones.

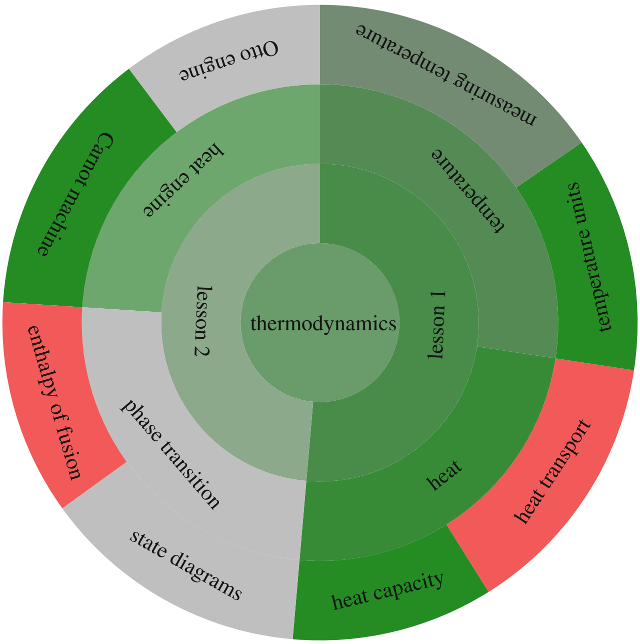



**Fig. 2.** Visualization of the learning progress of the student after solving several exercises. Unattempted topics are colored in gray. A failed exercise results in a red coloring of a topic if that topic was unattempted previously. Successfully solved exercises contribute to a positive topic score (colored in green) in proportion to their topic relevance value. The angles covered by the topic segments are proportional to the learning goal weights assigned by the teacher. (Color figure online)

## 3 Progress Visualization

A large body of research exists on the value of formative feedback for learning, and the premise that good feedback can significantly improve learning processes is widely shared [6]. The competence profile trees are well suited for progress visualization. For each topic both the learning goal relevance and the competence of the student are visualized intuitively using a sunburst diagram (see Fig. 2). The learning goal relevance is indicated by the angular size of the learning goal segment. Topics not yet studied are colored in gray and successfully studied ones using a gray-to-green scale. The scale indicates the sum of topic weights of correctly solved exercises. Failed topics that were never attempted previously are visualized in red. The intention is to provide an incentive to the student to cover all topics. It is done despite the fact that the adaptive navigation support does not distinguish between unsuccessfully attempted and never attempted topics.

## 4 Future Work

The recommender prototype is connected to a database that stores the learning goal weights, the exercise and student profiles as well as the current recommendations for each user. Currently a Moodle plugin is under development to complement the prototype with a suitable presentation framework. Within the Moodle platform it will present the recommended exercises to the student. The progress visualization is shown using the student profile data from the database.

**Acknowledgements.** We thank our colleagues from the IFeL (Institut für Fernstudien & eLearningforschung) for the cooperation in the project ALMoo (Adaptive Learning with Moodle) and fruitful discussions.

## References

1. Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., Koper, R.: Recommender systems in technology enhanced learning. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 387–415. Springer, Heidelberg (2011)
2. Brusilovsky, P.: Adaptive navigation support. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 263–290. Springer, Heidelberg (2007)
3. Drachsler, H., Verbert, K., Santos, O.C., Manouselis, N.: Panorama of recommender systems to support learning. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, 2nd edn. Springer, New York (2015)
4. Tsolis, D., et al.: An adaptive and personalized open source e-learning platform. Procedia **9**, 38–43 (2010)
5. Sosnovsky, S., Brusilovsky, P.: Evaluation of topic-based adaptation and student modeling in quizguide. User Model. User-Adap. Inter. **25**(4), 371–424 (2015)
6. Shute, V.J.: Focus on formative feedback. Rev. Educ. Res. **78**, 153–189 (2008)

# Learning Analytics for a Puzzle Game to Discover the Puzzle-Solving Tactics of Players

Mehrnoosh Vahdat[1,3](✉), Maira B. Carvalho[1,3], Mathias Funk[1], Matthias Rauterberg[1], Jun Hu[1], and Davide Anguita[2]

[1] Department of Industrial Design, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands
{m.vahdat,m.brandao.carvalho,m.funk,g.w.m.rauterberg,j.hu}@tue.nl
[2] DIBRIS - Università degli Studi di Genova, 16145 Genoa, Italy
davide.anguita@unige.it
[3] DITEN - Università degli Studi di Genova, 16145 Genoa, Italy

**Abstract.** Games can be used as effective learning tools, proved to enhance players' performance in a wide variety of cognitive tasks. In this context, Learning Analytics (LA) can be used to improve game quality and to support the achievement of learning goals. In this paper, we investigate the use of LA in digital puzzle games, which are commonly used for educational purposes. We describe our approach to explore the way players learn game skills and solve problems in an open-source puzzle game called *Lix*. We performed an initial study with 15 participants, in which we applied Process Mining and cluster analysis in a three-step analysis approach. This approach can be used as a basis for recommending interventions so as to facilitate the puzzle-solving process of players.

**Keywords:** Learning Analytics · Educational Data Mining · Serious games · Puzzle games · Technology enhanced learning · Cluster analysis · Process mining

## 1 Introduction

There is a growing field of investigation on the application of games as technology-enhanced learning tools, used to complement or enhance traditional education [1]. Learning Analytics (LA) and Educational Data Mining (EDM) can be applied in combination with game analytics to improve game quality and to support the achievement of learning goals [2,3]. Various methods of analytics in e-learning and game analytics help researchers make sense of data collected from user behavior, particularly through the use of modeling techniques [4] such as Process Mining (PM) [5,6] and cluster analysis [7].

In this paper, we propose the use of LA methods in one specific class of games: digital puzzle games. This type of game is commonly used for educational purposes [8], possibly given its typical reliance on problem-solving and on logical and mathematical intelligence [9]. We describe our approach to explore


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 673–677, 2016.
DOI: 10.1007/978-3-319-45153-4_89

the way players learn game skills and solve problems in the game, automatically extracting players' tactics and creating reference models for further analysis of other players' behavior.

We developed and tested our proposed approach in a puzzle game that offered an adequate development and testing environment, given its constrained interaction, deterministic game engine, clear success criteria, and limited dependence on external knowledge. Our goal is to define an analytics approach that can be extended to different types of learning games. Additionally, we aim to use this approach in the future to support the implementation of automatic adaptive features for educational games, such as targeted interventions, appropriate feedback, and timely hints for the player/learner.

## 2 Game Description and Data Collection

We extended an existing open-source puzzle game called *Lix* [10], which is inspired by *Lemmings*, a 1991 game by DMA Design. In *Lix*, the objective is to guide a group of simple characters to a designated exit (Fig. 1).

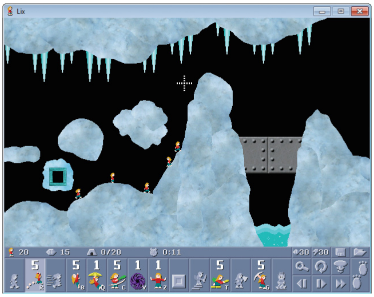


**Fig. 1.** *Lix* game interface.

To collect game data, we altered the game following a Service-Oriented Architecture approach [11]. The game performs network calls to a web service that listens to relevant game events, and records them in a database. The events recorded are of two types: *game* traces and *meaningful variable* traces [3]. The *game* traces indicate timestamps of when the player started the game, started or restarted a puzzle, paused the game and returned to the menu. The *meaningful variable* traces consist of a simple record of a timestamp, a short code describing the skill assigned to a character, an internal identifier of the character to which the skill was assigned and an internal measure of game time.

We collected preliminary game data in a study with 15 adult participants. Participants were given a brief explanation of the study and of the goal of the game. No explanations about the game user interface were given. Participants were given a pre-test questionnaire to collect demographic data and gaming experience. They were asked to play one intermediate level puzzle of the game as many times as they wanted. They were asked to think aloud for us to take notes on their reactions, tactics, persistence, etc. The data was used to develop our analytics approach, explained in the next section.

## 3 Analytics Approach

We developed a data-driven analytics approach that combines PM and cluster analysis to discover the way players learn skills, solve problems, and succeed in a specific puzzle game. In particular, our objective was to discover the clusters of the tactics applied by the players and identify a reference sequence for each cluster. We obtained the reference sequences by building the process models of tactics. These process models identify the most significant activities and transitions through PM. The reference sequences play a central role in validation of the results. By comparing a player's process to previously established successful references, we aim to detect whether the player behaves closely to them.

Our analytics approach comprises of three main steps. A preliminary step is collecting the data from the game ('A' in Fig. 2), as explained in Sect. 2. The first step is to identify the tactics adopted in the game by players through cluster analysis ('B'). In the second step, we aim to obtain the process models of the identified tactics through PM. These models represent the most significant components of the tactics which yield references that are central in validation of our results ('C'). Finally, we validate the results of cluster analysis and PM by measuring how the elements of a tactic cluster converge to their reference ('D').

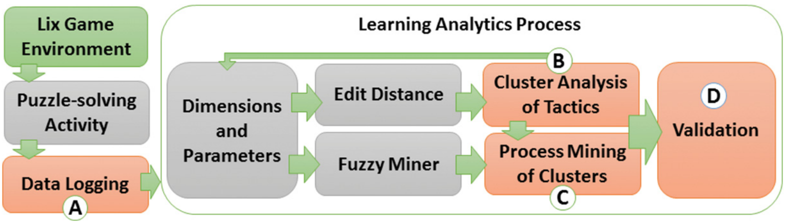



**Fig. 2.** Learning Analytics approach

## 4 Conclusions

In this study, we present a novel approach to apply LA methods on interaction data collected from an open-source puzzle game called *Lix*. This game is used in our study because of the value of puzzle games for educational purposes [8,9], and, as such, developing ways to automatically analyze players' problem-solving processes can be a valuable tool for educators and game designers alike.

We presented a three-step analytics approach that uses clustering, process mining, and validation to extract the puzzle-solving tactics from data, even without previous knowledge about the nature of the puzzle. The advantages of this approach can be explained as follows: we can identify previously unknown tactics, and not only the ones assumed by the game designer. We can avoid manually defining optimal strategies for every level of a game. Also, this approach can raise awareness of the educators about the learning progress by visualizing how close or far away from the optimal tactic any player is.

The results confirms that our LA approach was successful: two main successful tactics were discovered through cluster analysis, the process models of these tactics were successfully obtained, and yielded references for validation. Finally, the validation results indicate that we obtained meaningful clusters of different tactics, as the members of each cluster converged to their reference.

In the future, we will verify our approach by reporting the results and applying the same methodology to more data, in order to cross-validate the obtained process models. Additionally, we aim to extend our approach to other skill-based puzzle games, using it as input to automatically recognize the different stages of puzzle solving and to detect which players are most likely to quit the game. Finally, we plan to use this approach as the basis for recommending interventions that could allow the game to provide the player/learner with help on time, for instance by automatically comparing a given player's tactic to the successful tactics identified by this approach.

**Acknowledgments.** This work was supported in part by the Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments, funded by the EACEA Agency of the European Commission under EMJD ICE FPA n 2010-0012.

## References

1. Erhel, S., Jamet, E.: Digital game-based learning: impact of instructions and feedback on motivation and learning effectiveness. Comput. Educ. **67**, 156–167 (2013)
2. Bohannon, J.: Game-miners grapple with massive data. Science **330**(6000), 30–31 (2010)
3. Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., Fernández-Manjón, B.: Application of learning analytics in educational videogames. Entertainment Comput. **5**(4), 313–322 (2014)
4. Siemens, G., Baker, R.S.: Learning analytics and educational data mining: towards communication and collaboration. In: 2nd International Conference on Learning Analytics and Knowledge, pp. 252–254 (2012)
5. Trcka, N., Pechenizkiy, M., Van Der Aalst, W.: Process Mining from Educational Data. Chapman & Hall/CRC, London (2010)
6. Vahdat, M., Oneto, L., Anguita, D., Funk, M., Rauterberg, M.: A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator. In: Design for Teaching and Learning in a Networked World, pp. 352–366 (2015)
7. Bauckhage, C., Drachen, A., Sifa, R.: Clustering game behavior data. IEEE Trans. Comput. Intell. AI Games **7**(3), 266–278 (2015)

8. Liu, E.Z.F., Lin, C.H.: Developing evaluative indicators for educational computer games. Br. J. Educ. Technol. **40**(1), 174–178 (2009)
9. Becker, K.: How are games educational? Learning theories embodied in games. In: DiGRA: Changing Views - Worlds in Play (2005)
10. Naarmann, S.: Lix (2011). https://github.com/SimonN/Lix
11. Carvalho, M.B., Bellotti, F., Hu, J., Baalsrud Hauge, J., Berta, R., Gloria, A.D., Rauterberg, M.: Towards a service oriented architecture framework for educational serious games. In: IEEE 15th International Conference on Advanced Learning Technologies (ICALT), pp. 147–151 (2015)

# Recommending Dimension Weights and Scale Values in Multi-rubric Evaluations

Mikel Villamañe(✉), Ainhoa Álvarez, Mikel Larrañaga, and Begoña Ferrero

Department of Languages and Computer Systems, University of the Basque Country UPV/EHU, Leioa, Spain
{mikel.v, ainhoa.alvarez, mikel.larranaga, bego.ferrero}@ehu.eus

**Abstract.** Rubrics are scoring tools that lay out the specific expectations for an assignment. They are very appropriate tools for formative assessment as they have proved to be adequate to reduce subjectivity in the evaluation process. When the evaluation entails several tasks, a rubric for each task should be defined. However, computing the final score using rubrics is not always a simple task. On the one hand, each task has its own relevance in the final grade. On the other hand, the score of each rubric depends on the performance levels achieved in each dimension and the importance or weight of each dimension. Determining the most appropriate weight for each task, dimension and performance level is complex. This paper presents a recommender for settling those values in a multi-rubric evaluation process.

**Keywords:** Rubric-based evaluation · Evaluation factors adjustment · Recommender system

## 1 Introduction

Nowadays, using rubrics to evaluate students' work is becoming very popular. This interest is due, in part, to the fact that rubrics are scoring tools that lay out the specific expectations for an assignment, encouraging consistent grading and increasing objectivity in the evaluation [1, 2]. Building any rubric entails defining 4 elements [3].

1. *Task description*: A brief description of the assignment that will be evaluated.
2. *Scale*: Levels of performance for the task.
3. *Dimensions*: Breakdown of the parts involved in the task. The relevance of each dimension can be represented by its weight on the final rubric score.
4. *Description for each performance level*: Specific expectations for each dimension.

Rubrics are primarily thought for formative assessment; however, sometimes a final grade for the course must also be obtained. The final grade for a course involving several tasks, can be calculated using a weighted average of the grades obtained by the student in each task (Eq. 1). In this equation, $M_i$ represents the grade obtained by the student in task *i* and $W_i$ the weight of task *i* on the final grade.


K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 678–681, 2016.
DOI: 10.1007/978-3-319-45153-4_90

$$FinalGrade = \frac{\sum (W_i * M_i)}{\sum W_i} \tag{1}$$

$$RubricGrade = \frac{\sum W_i(g_i - \min)}{\sum W_i(\max_i - \min_i)} \tag{2}$$

Using rubrics, Eq. (2) is widely used to compute the numeric grade when dimensions weighted differently are considered. In this equation $g_i$ represents the grade assigned to dimension $i$, $min_i$ and $max_i$ the minimum and maximum grade achievable in dimension $i$ respectively, and, finally, $W_i$ the weight of dimension $i$ in the final rubric grade. This equation computes the rubric score as a percentage.

Equations (1) and (2) were integrated into a multi-rubric evaluation system and tested to obtain the final grade for Final Year Projects (FYPs) in Computer Science. For this experiment, a set of experts defined a procedure to adequately evaluate FYPs [4], and identified 6 deliverables or tasks, each of them with different relevance in the project's final grade calculation [6]. They also defined a rubric to assess each of the tasks, its performance levels and the weight for each dimension.

To test the results obtained with the given equations, the evaluation board provided the estimated grade for each deliverable and the final grade of the project before filling out the corresponding rubrics. The grades given by the evaluators were used as gold-standard as their evaluations were assumed to satisfy the evaluation policies for FYPs. Using the given equations and the weights estimated by the set of experts 1.5 RMSE (over 10) was obtained. This aligns with other studies such as that of Salinas and Erochko [5]. This shows that determining which are the weights and the values for the performance levels that better reflect the evaluation policy is not an easy task. However, there exists no system in the literature that helps adjusting those values.

This paper presents a system that recommends the influence that each element of a multi-rubric evaluation should have.

## 2 Weight Adjustment System

The *Weight Adjustment System* can recommend the weights and influence of the different elements that take part in a rubric-based evaluation. This system proposes the influence each task should have in the final grade -the *Task Adjusted Model* (TAM)-, and it also computes the weights of dimensions and the most appropriate scale-level values for each one of the involved rubrics (*Rubric Element Model* or REM).

To obtain each of those models, a dataset with information regarding the evaluation process is required. Two of the elements considered to build the dataset must be supplied by the lecturers: the grade the students should have in each task and the final grade the students should obtain in the course. The last element in the dataset is the grade obtained for each task using the corresponding rubric. Combining these three elements allows constructing the datasets for the TAM and REM obtaining processes.

### 2.1 Settling the Task Adjusted Model (TAM)

One of the elements to be adjusted in the evaluation process is the influence each task should have in the final grade, i.e., the *Task Adjusted Model* (TAM). Determining the TAM is a regression problem, where the target variable is the final grade of the student in the course and the features are the grades in the different items (course tasks) that according to experts should influence the final grade. The objective is to obtain a percentage denoting to which extent each task affects the final grade.

The process to obtain the TAM, takes the dataset including the grades for the involved tasks for each student and the final grade obtained by the student in the course. The course tasks always reflect aspects that the course being evaluated should satisfy, whilst a negative weight would mean that an undesirable or wrong task is being evaluated. Therefore, the Lawson-Hanson Non-negative least-squares technique (NNLS) [7] is used to determine the most appropriate weight for each task [6].

### 2.2 Settling the Rubric Element Model (REM)

Settling the REM implies obtaining the optimal *Scale-level Model* -i.e., the most appropriate value for each performance level- and the weight of each dimension. Again, adjusting those values is a regression problem where all weights should be non-negative. The target variable of the process is the final grade and the features are the performance levels selected for each dimension of the rubrics.

The process to obtain the REM has 4 phases: Configuration, preprocess, training and model selection. The data required to execute them include: the rubrics filled out detailing the performance level selected for each dimension and the estimated overall grade for each task without using the rubrics.

In the configuration phase, each performance level is assigned a range of possible numeric values. This correspondence is determined by the experts according to the numeric grades the rubric should obtain. For example, in our University, students are graded in the [0, 10] range. Therefore, in this case the experts could assign the 10 value to the *Excellent* level, the [9, 9.99] range to the *Advanced* level, and so on.

In the preprocess phase, a set of *Scale-level Models* ($SM_i$) are generated. To build up those models, all the possible combinations for the distinct potential numeric values of the performance levels must be considered. In order to obtain those combinations, the values from the top value to the lower value of the corresponding range are generated with 0.1 increments between each pair of values. Next, for each *Scale-level Model*, the initial dataset is adjusted with the specific values of that model generating different datasets ($DS_i$). Following, the original dataset is randomly partitioned into two groups that determine which students will be used for the *training sets* ($TS_i$) and which for the *validation sets* ($VS_i$). Then, each $DS_i$ is split into its corresponding *training set* and *validation set* according to the partition made with the initial dataset.

Once the data is prepared, the training phase is carried out with each one of the obtained *training sets* ($TS_i$). Again, the Lawson-Hanson Non-negative least-squares (NNLS) technique [7] is used to train the corresponding *Rubric Element Models*, enforcing non-negative values for the weights of the dimensions. In this training step, the features are the performance level assigned to each dimension of the rubrics and the predicted variables are the estimated grades of the corresponding tasks.

Finally, in the model selection step, the trained evaluation models obtained in the previous step ($REM_i$) are tested on the corresponding *validation set* ($VS_i$) to determine the best performing model to be recommended. To this end, the final grades of the tasks are computed using the *Task Adjusted Model* (TAM, Sect. 3.1) and the grades assigned by each *Rubric Element Model* ($REM_i$) to the rubrics. The Root-Mean-Squared Error (RMSE) is used to compare the computed final grades with the course's grades given by the lecturers to choose the $REM_i$ with the lowest $RMSE_i$.

## 3 Conclusions and Future Work

Rubrics are assessment tools that are becoming very popular. They have very positive characteristics but they also show some problems. One of the most challenging tasks is the adjustment of their different weights and values required to obtain a numerical grade. This problem gets bigger when the course evaluation is based in multiple tasks, each of them with a different weight in the final grade calculation. To overcome this problem, a system which automatically adjusts those values has been presented.

Adjusting those values also allows identifying those tasks or dimensions that have been identified by the experts but that are not statistically significant for obtaining the final grades (those with a weight of 0). This is, it allows improving the course evaluation process removing non-significant tasks or dimensions.

In the near future, the *Weight Adjustment System* is planned to be used in some other courses in the Computer Science grade.

**Acknowledgements.** This work is supported by the Basque Government (IT722-13) and the University of Basque Country UPV/EHU (UFI11/45).

## References

1. Hafner, J., Hafner, P.: Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. Int. J. Sci. Educ. **25**, 1509–1528 (2003)
2. Jonsson, A., Svingby, G.: The use of scoring rubrics: reliability, validity and educational consequences. Educ. Res. Rev. **2**, 130–144 (2007)
3. Moskal, B.M.: Scoring rubrics: what, when and how? Pract. Assess. Res. Eval. **7**, 88–96 (2000)
4. Villamañe, M., Ferrero, B., Álvarez, A., Larrañaga, M., Arruarte, A., Elorriaga, J.A.: Dealing with common problems in engineering degrees' Final Year Projects. In: Proceedings of International Conference on IEEE Frontiers in Education (FIE), pp. 2663–2670. IEEE Computer Society, Madrid (2014)
5. Salinas, J.J., Erochko, J.: Using weighted scoring rubrics in engineering assessment. In: Proceedings of the Canadian Engineering Education Association (2015)
6. Villamañe, M., Larrañaga, M., Álvarez, A., Ferrero, B.: Adjusting the weights of assessment elements in the evaluation of Final Year Projects. In: Proceedings of the International Conference on Educational Data Mining. pp. 596–597. Madrid, Spain (2015)
7. Lawson, C., Hanson, R.: Solving least squares problems. Soc. Ind. Appl. Math. (1995)

# Author Index