



Helping Teachers Help Their Students: A Human-AI Hybrid Approach

Ranilson Paiva^(✉) and Ig Ibert Bittencourt

Instituto de Computação - IC, Universidade Federal de Alagoas - UFAL, Av. Lourival
Melo Mota, S/N - Cidade Universitária, Maceió, Alagoas 57072-970, Brazil
{[ranilsonpaiva](mailto:ranilsonpaiva@ic.ufal.br),[ig.ibert](mailto:ig.ibert@ic.ufal.br)}@ic.ufal.br
<http://www.ufal.edu.br/unidadeacademica/ic>

Abstract. There is a global interest in artificial intelligence to support online learning, but little increase in support for online professors, teachers and tutors (instructors). Over time, more students join online learning, but instructors have no equivalent increase in support to manage their online classes, leaving students under-served. This is evidenced by the number of students who dropout or fail online courses, blaming the “lack of support” from instructors. Interactions in such courses generate considerable quantity and diversity of data, allowing the extraction of pedagogically relevant information. However, instructors do not master the techniques and technologies needed to do it, and it is not practical to train them to do so. In this work, we propose an authoring tool (called T-Partner) that implements a process we created to deal with educational data. The objective is to support instructors making informed pedagogical decisions to manage their online course. T-Partner promotes the cooperation between artificial and human intelligences, however we do not know the appropriate balance between these “intelligences”. We then created two versions of the T-Partner to help instructors to: (1) find relevant pedagogical situations occurring within their online courses; (2) understand these situations; (3) create interventions (study plans, for example) to address these situations; (4) monitor and evaluate the impact of these interventions. We evaluated if both versions allowed instructors to make pedagogical decisions and their perceptions regarding this support to decision-making. The results show that both versions brought benefits to pedagogical decision-making, and were positively perceived by the participants.

Keywords: Pedagogical Decision-Making · Data-informed decisions · Authoring tools · On-line learning environments

1 Introduction

Technology can influence the processes and outcomes of education, and many countries are investing in technological support for teaching and learning [20]. Online education is one such example and the number of courses offered online

increase constantly and worldwide [8, 19]. Besides that, some countries are passing laws to regulate online learning, while others are investing considerable amounts to stimulate its use¹. These facts evidence that online education is a viable approach to propagate and democratize education, and there is demand for it.

However, professors, teachers and tutors (we will refer to them as instructors) face challenges with online education. One such challenge is to make (course) decisions using educational data. Doing so requires instructors to quickly and continually deal with these data [21], which can be diverse and in considerable volume. Training them to analyze these data would require lots of time, effort and resources, with uncertain results [7, 11]. This highlights the importance to provide instructors the necessary technological support [5, 10]. These information indicate the need to: (1) help instructors extract relevant information from educational data; (2) Provide them means to create personalized interventions to address issues discovered and; (3) Check the success of these interventions [14, 15].

Complimentary, there is a new research branch in the AIED field: the creation of artificial intelligence to **collaborate** with human intelligence [2, 4]. This should position professors, teachers and tutors as the main decision-makers [3, 20] in the online “classroom”. However, there are no scientific works regarding how much of each of these “intelligences” (artificial and human), should be used in this collaboration. Based on these information, we ask the following research question: how can we balance artificial and human intelligence in order to help instructors manage their online courses while they are occurring? Researching about this problem, we found the definition for authoring tools, which is a tool to help users (professors, teachers and tutors), allowing them to create, sequence and publish content (to students), without requiring advanced technical knowledge or training [6, 12].

In this work, we evaluate two versions of an authoring tool (we called it Teachers’ Partner or T-Partner). The two versions were named: lightweight and heavyweight, and they were created to present different combinations of artificial and human intelligences in order to assist online instructors manage their courses. In the lightweight version, users make simple choices, with the artificial intelligence having more control over the decisions. In the heavyweight version, users are required to make more choices, giving them more control over the decisions. The trade-off is: simplicity vs. control.

T-Partner was designed to assist online instructors to: (1) search for relevant pedagogical situations in the learning environment (using educational data); (2) Use these data to generate visualizations of patterns and trends in order to understand what is happening with their students/courses; (3) Create personalized study plans (interventions) and deliver them to the target students; (4) Check whether these study plans helped students or not.

We evaluated both versions of T-Partner by asking instructors to complete the four tasks listed above, for a specific scenario, which was to evaluate how the students’ interactions affect their performance. The results show the participants

¹ Available at: <https://bit.ly/2zM3JKp> and <https://bit.ly/2FX7saW>.

(instructors) were able to properly complete the proposed tasks and had positive perceptions about the T-Partner, considering it easy to use, helpful and interesting. The results also show that the participants preferred the heavyweight version, suggesting that the balance between artificial and human intelligence should be designed in favor of human control on the decision-making process.

2 Proposal

In this section, we present the Process (Pedagogical Decision-Making Process), and the two versions (lightweight and heavyweight) of the authoring tool (T-Partner) created.

2.1 Pedagogical Decision-Making Process (PDMP)

The T-Partner follows a process where: (1) educational data is analyzed in search for pedagogical situations; (2) Relevant issues are presented to the instructors as easy-to-understand and interactive visualization; (3) The educational resources (videos, texts, questions, etc.) are organized (domain, curriculum and knowledge component), allowing instructors to devise pedagogical interventions for the pedagogical situations found and; (4) The instructor defines the criteria to measure if the interventions were effective or not (Fig. 1).

The Pedagogical Decision-Making Process (PDMP) is a cyclical process and its objective is to guide instructors (from online learning environments) to: (1) discover issues/situations, with pedagogical value, occurring in their online courses; (2) Understand these situations; (3) Make decisions to address them; (4) Monitor and evaluate the impact of the decisions made. The PDMP has two phases: the construction phase and the execution phase. In the construction phase, human and artificial intelligences collaborate to specify (1) which, among some defined pedagogical situations, they want to search for in the learning environment; (2) What decision they want to make, considering the learning environment's capabilities, to address a pedagogical situation found; (3) How they want to measure the effectiveness of the decision made [16, 17]. In the execution phase, the successful definitions made in the construction phase are automatically repeated, if the same pedagogical situation is found again.

In previous works, we used the PDMP to: (1) evaluate the effectiveness of gamification elements in an OLE [13]; (2) Measure differences between male and female students' interactions in an online learning environment (OLE) [18]; (3) Improve students' interactions in an OLE [16]; (4) Recommend topics learners should study to improve their writing performance [1], among other uses².

2.2 T-Partner

In order to avoid the error-prone and repetitive task of manually following the PDMP, we created an authoring tool named T-Partner (Teachers' Partner). The T-Partner needs to be integrated to a learning environment in order to access

² Detect and recommend actions to disengaged learners; Recommend educational resources to practice a specific math topic etc. These works are not published, yet.

Pedagogical Decision-Making Process

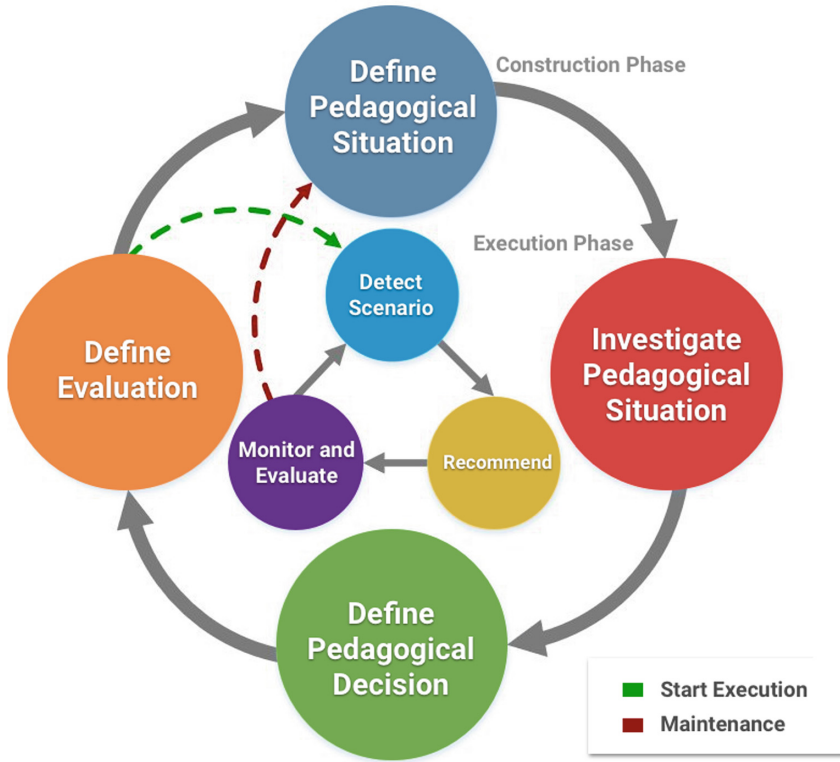


Fig. 1. The Pedagogical Decision-Making Process.

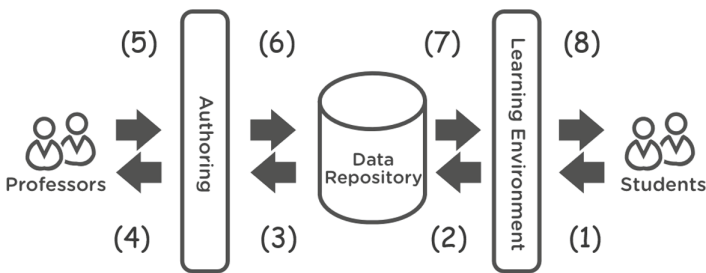


Fig. 2. T-Partner communication with an online learning environment.

its educational data³. Basically: (1) learners interact with the online learning

³ Data about learners, educational resources, interaction data (user-user, user-content and user-environment interactions), and other data to support instructors' pedagogical decision making.

environment (OLE); (2) These interactions generate (educational) data that are stored in the OLE's data repository; (3) These data are retrieved and processed by the T-Partner; (4) The results are used to inform instructors about pedagogical situations occurring in the OLE; (5) Instructors use this information to make pedagogical decisions; (6) These decisions use the educational resources available in the OLE; (7) The decisions should consider the OLE's interface capabilities; (8) The decisions are sent to the targeted learners; (9) The T-Partner measures the effectiveness of the decisions (Fig. 2).

We created two versions of the T-Partner: (1) *Light Weight*: This version is an easy to use, but more limited, version of the tool for users with little experience with computers, allowing them to make pedagogical decisions easier and faster, but more constrained. It can also be used as an entry version for training instructors; (2) *Heavy Weight*: This version has more features, allowing finer-grained decisions, but it may slow down the process and be more complex/demanding to the user⁴.

2.3 T-Partner Implementing the PDMP

In this subsection we describe how T-Partner implements the PDMP (Subsect. 2.1). It is important to mention that instructors should first define a domain and a curriculum. For example: linear functions (curriculum) in the math domain.

Step 1: Define the Pedagogical Situation. In this step, the instructors choose among the available pedagogical situations, defining some parameters in order to personalize data collection. After that, they must specify how T-Partner must classify each parameter as inadequate, insufficient and adequate, named **classes of results**. For example: an instructor wants to evaluate the impact of the students' interactions with some educational resources, for a particular subject, in the previous 15 days. The instructor chooses the domain (math), the curriculum (linear functions), the group (group 1). (S)he selects the resources (s)he wishes to measure the impact on the students' performance (the students' accesses to the course, their gamification level, the number of badges they received and the number of video classes they watched). (S)he also defines the period of time the analysis must consider (the last 15 days). Next, the instructor classifies the amount of interactions for each chosen resources as inadequate, insufficient or adequate. Considering the number of accesses, the instructor classified it as follows: (1) below 30% the average, is considered inadequate; (2) between 30% and 59% the average, is considered insufficient; (3) above 59% the average, is considered appropriate. The instructor classified level, badges and videos with the same values as the number of accesses. The T-Partner searches the data, following the parameters defined, and classifies the resulting data according to the classification values provided by the instructor.

⁴ Due to the restricted number of pages, it was not possible to add pictures of T-Partner in this paper. However, we created a website where we present images and detailed descriptions of both versions. It is available at: <http://tpartner.ranilsonpaiva.com>.

Step 2: Investigate Pedagogical Situation. In this step, the T-Partner groups the students according to the way instructors classified the resources. In this part of the tool, data is processed using an algorithm associated with the pedagogical situation chosen in step 1 (for example: if the instructor chose to evaluate the students' interactions impact on their performance, the algorithm used is a Decision Tree). Before data processing, instructors select how to pre-process the data (imputation, remove registries with missing values, remove outliers etc.), and how they wish to visualize the data processing result. The resulting visualization uses different colors to represent different result classes: inadequate - red, insufficient - yellow and adequate - green. The aim is to provide instructors with information extracted from the educational data, in order to aid their decision process.

Step 3: Define Pedagogical Decisions. In this step, instructors create a personalized intervention (for example: a study plan) for each class of results (inadequate, insufficient and adequate). For each intervention, instructors must give it a name and define: (1) the activities learners should do (texts, videos, questions etc., depending on what is available in the OLE); (2) The amount for each activity; (3) The order the activities should be arranged; (4) The desired modifiers for the activity (for multiple choice questions, a modifier can be its difficulty); (5) The target class of results; (6) The amount of time learners have to complete the task; and (7) The pedagogical approach learners should follow to finish the task (for example: do it individually, do it in group, peer-evaluate colleagues answers, receive points or badges for doing it in case of gamified learning environments etc. It depends on what the OLE offers). For example, an intervention can be, for **linear function** in the **math domain**, to **read one text, watch one video-class, answer one easy multiple-choice question, answer three difficult multiple-choice questions (in this order)**. This must be sent to students in the **insufficient** result class, who must do it **individually** and in the **next 10 days**.

Step 4: Define Assessment. In this step, instructors set the desired percentage for *adherence*⁵ and the *desired outcome* from those who followed the recommended intervention. This is done for each class of results. For example: an instructor defines, for the inadequate result class, 50% adherence and an 20% increase in the students' performance (number of correct answered divided by the number of questions answered).

3 Experimentation

In this experiment, we invited professors, teachers and tutors to evaluate the T-Partner. This experiment was available on-line for a 30 days period. After this period, we collected the participation data, cleaned them, removing test data,

⁵ The amount of learners that completed the intervention recommended by the instructor.

empty and incomplete records. Next, we performed the data analysis following the guidelines proposed by [9].

Part 1 - Using T-Partner to Solve a Pedagogical Issue. Participants were randomly assigned to one of the two versions of the T-Partner⁶. They had to read a description of a real education scenario, guiding them to perform the following tasks: (1) evaluate the performance of the students based on their interactions in the OLE; (2) Create a study plan for each class of results; (3) Define the criteria of a successful intervention.

Based on the scenario, we asked participants to: (1) choose the issue they wanted to search for in the learning environment. The available options were: evaluate students' failing probability; evaluate students' dropping out probability and evaluate students' interactions with the educational resources; (2) Choose one of the pre-processing techniques available. Some available options were: remove empty and null registries and apply imputation technique. Next, choose the way they wanted to visualize results⁷; (3) Create a study plan, to address the issue, for each class of results and define how long students had to complete it; (4) Define the adherence and the desired outcome.

Participants could make different choices (decisions). Some were appropriate, some were not. We calculated a **score**, for each step, which was the sum of the tasks completed appropriately, divided by the total amount of tasks, according to the formula:

$$SCORE = \frac{\sum_{i=1}^n e_i}{MAX} \quad (1)$$

Part 2 - The Participants' Perceptions About the T-Partner. We asked the participants' perceptions, regarding the following metrics: (1) Perceived utility (PU) - if participants considered the tool useful to manage their courses; (2) Perceived ease of use (PEU) - if participants considered the tool easy to use; (3) Attitude towards use (ATU) - if participants had a positive attitude towards using the tool; (4) Intention to use (IU) - if participants would use the tool if it was available in their workplace; (5) Visualizations used (VIZ) - if the visualizations used were informative; (6) Color scheme used (COL) - if the colors used (red, yellow and green) to represent the classes of results, helped participants understand the situation learners were facing; and (7) vocabulary used (VOC) - if the vocabulary used was appropriate. The first 4 metrics were **based** on the Technology Acceptance Model [22] and the others were created for the purpose of this experiment. Participants had to assign a score for each criteria, according to a Likert scale from 0 to 6, where: 0 = I strongly disagree; 1 = I disagree; 2 = I slightly disagree; 3 = I neither agree nor disagree (indifferent); 4 = I slightly agree; 5 = I agree; 6 = I strongly agree.

⁶ We used anonymized data from a high school level online learning environment with more than 6000 active Brazilian students.

⁷ The available visualizations depended on the pedagogical issue chosen in step 1.

Table 1. Scores for accomplishing the tasks (LW = Light Weight and HW = Heavy Weight).

PDMP	MIN LW/HW	MAX LW/HW	MED LW/HW	AVG LW/HW	SD LW/HW
Step 1	0/0.67	1/1	1/1	0.94/0.97	0.15/0.10
Step 2	0/0	1/1	0.93/1	0.84/0.93	0.3/0.19
Step 3	0/0	1/1	0/0.55	0.33/0.52	0.44/0.49
Step 4	1/0.1	1/1	1/1	1/0.7	0/0.84
All tasks	0/0	1/1	1/1	0.84/0.87	0.32/0.29

Table 2. Score comparison for completing the tasks (Wilcoxon-Mann-Whitney test).

PDMP	P-VALUE	HIGHER SCORE
Step 1	0.0001984*	<i>HW</i>
Step 2	0.08513	No difference
Step 3	0.01698*	<i>HW</i>
Step 4	0.0004147*	<i>LW</i>
All tasks	0.02922*	<i>HW</i>

* = Statistically significant; LW = Light Weight; HW = Heavy Weight.

4 Results and Discussion

Regarding the participants, we had 45 complete and valid participations, with $n = 20$ for the Light Weight version and $n = 25$ for the Heavy Weight version. They were all higher education professors from Brazil, with ages ranging from 32 to 63 years old. Their years of experience ranged from 6 years to more than 15 years as higher education professors. Their level of familiarity and professional use of educational technologies ranged from good to very good.

For part 1, the Score for each step was normalized⁸. We calculated the minimum score (MIN), maximum score (MAX), median score (MED) average score (AVG) and the standard deviation (SD) for each step. The results are shown in Table 1. In Table 2 we applied the Wilcoxon-Mann-Whitney to test for statistical relevance of the differences in scores for the two versions.

The results show that, regarding the tasks in Steps 1, 2 and the sum of the tasks in all steps, the scores in the Heavy Weight version (HW) were higher, suggesting the HW version allowed instructors to make better decisions (score higher in doing what was expected from them) than the LW version. The Heavy Weight version offers greater detailing and control to instructors (human intelligence), which is represented by more options to make more detailed decisions. In the Light Weight version this control and detailing is mostly done by the system. We believe that having the system handle some parts of the decision confused the participants, affecting their comprehension and proper completion of the task.

⁸ All scores are in the 0 to 1 interval.

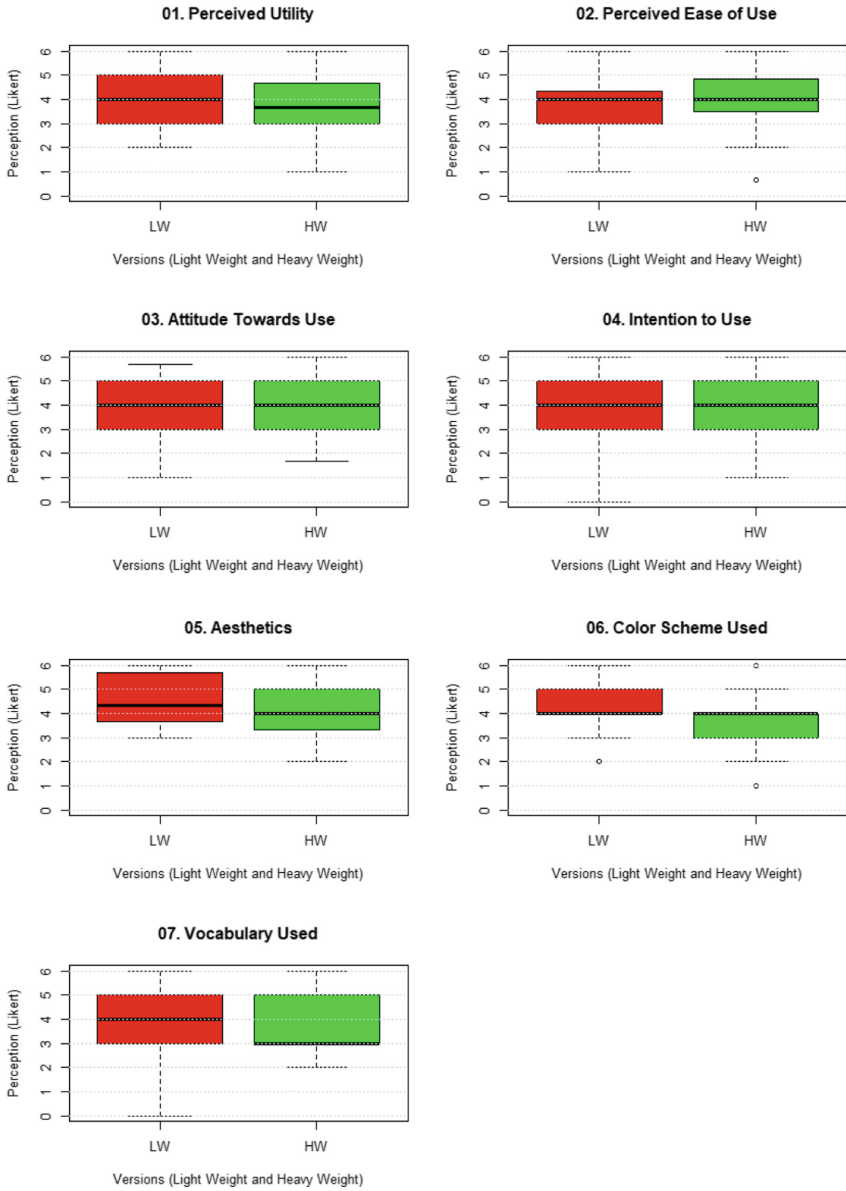


Fig. 3. Participants' perceptions about the T-Partner's versions.

We need to further investigate other variations of this human/computer balance for each step of the process.

The results show higher standard deviation, regarding the scores in Steps 3 and 4 in the Heavy Weight version, suggesting participants had difficulties

completing the respective tasks. This may be due to these steps were the ones with a higher amount of tasks to complete (in the HW version). It may be the case to improve clarity of the steps and/or divide these steps into sub-steps (further investigation is necessary).

The results of the participants' perceptions for all metrics of both versions (heavy weight and light weight) were positive and similar (Fig. 3), which is a good and desired result, showing that the participants had a favourable perception regarding the T-Partner and the process, independent of the version. The median value was 4, which corresponds to the answer "I slightly agree." This shows that participants' perceptions were positive (above neutral/indifferent).

5 Conclusion

We proposed the Pedagogical Decision-Making Process (PDMP) and an authoring tool (T-Partner) that implements it. The objective was to have artificial and human intelligence work, collaboratively, to help online instructors managing their courses/students, offering personalised assistance. However, we did not know how to balance these two "intelligences" in the final tool. Therefore, we created two versions of T-Partner: the light weight version, where most part of the decisions are made by the system, and the heavy weight version, where most part of the decisions are made by the instructors (professors, teachers and tutors). We evaluated both versions of T-Partner, regarding its capacity to support instructors' pedagogical decision-making as well as their perceptions on its utility and use.

Overall, the results showed that the participants were able to properly perform the demanded tasks, supporting online instructors' pedagogical decision-making, with some minor issues in the tasks for steps 3 and 4 from the Heavy weight version (with higher number of tasks to complete). Instructors show positive perceptions regarding all the metrics considered, independent of the version, stating that: (1) the tool would be useful to help them manage their courses/students; (2) The tool was easy to use; (3) They had a positive perception towards using the tool; (4) They would use the tool if it was available in their workplace; (5) The visualizations provided were informative; (6) The color scheme helped them understand how serious the students' situation was and; (7) The vocabulary used was appropriate.

We believe the proposed process and tool are a step towards augmenting human intelligence with artificial intelligence in the education area. However, we noticed that some situations require more research and experiments, for example: what is the ideal balance between human and artificial intelligence for making pedagogical decisions? Does this balance change in different contexts? Does the experience of the instructor affect his/her interest in more or less control over the pedagogical decisions they make and the technology support they receive? We will research these and other questions in future works.

Acknowledgements. We would like to thank EyeDuc® team for granting us access to the data used in this research. We acknowledge the financial support from FAPEAL and CAPES/CNPq.

References

1. Abreu, D., Bittencourt, I., Paiva, R., Dermeval, D.: Pedagogical recommendation to improve the quality of writing: a case study in a public school. In: 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), pp. 75–76. IEEE (2018)
2. Baker, R.S.: Stupid tutoring systems, intelligent humans. *Int. J. Artif. Intell. Educ.* **26**(2), 600–614 (2016)
3. Burden, P., Fuller, A.: The teacher as a decision maker. In: *Methods for Effective Teaching Meeting the Needs of All Students* (2014)
4. Chou, C.Y., Huang, B.H., Lin, C.J.: Complementary machine intelligence and human intelligence in virtual teaching assistant for tutoring program tracing. *Comput. Educ.* **57**(4), 2303–2312 (2011)
5. Dağ, F., Durdu, L., Gerdan, S.: Evaluation of educational authoring tools for teachers stressing of perceived usability features. *Proc.-Soc. Behav. Sci.* **116**, 888–901 (2014)
6. Dermeval, D., Paiva, R., Bittencourt, I.I., Vassileva, J., Borges, D.: Authoring tools for designing intelligent tutoring systems: a systematic review of the literature. *Int. J. Artif. Intell. Educ.* **28**(3), 336–384 (2018)
7. Duhaney, D.C.: Teacher education: preparing teachers to integrate technology. *Int. J. Instruct. Med.* **28**(1), 23 (2001)
8. AC EAD: Br: Online learning in Brazil - analytic report 2017/2018. Ibpx, Curitiba (2018)
9. Jain, A.K.: Data clustering: 50 years beyond k-means. *Pattern Recogn. Lett.* **31**(8), 651–666 (2010)
10. Kim, C., Kim, M.K., Lee, C., Spector, J.M., DeMeester, K.: Teacher beliefs and technology integration. *Teach. Teach. Educ.* **29**, 76–85 (2013)
11. Kopcha, T.J.: Teachers’ perceptions of the barriers to technology integration and practices with technology under situated professional development. *Comput. Educ.* **59**(4), 1109–1121 (2012)
12. Murray, T.: An overview of intelligent tutoring system authoring tools: updated analysis of the state of the art. In: Murray, T., Blessing, S.B., Ainsworth, S. (eds.) *Authoring Tools for Advanced Technology Learning Environments*, pp. 491–544. Springer, Dordrecht (2003). https://doi.org/10.1007/978-94-017-0819-7_17
13. Paiva, R., Barbosa, A., Batista, E., Pimentel, D., Bittencourt, I.I.: Badges and XP: an observational study about learning. In: *Frontiers in Education Conference (FIE)*, pp. 1–8. IEEE (2015)
14. Paiva, R., Bittencourt, I.I.: The authoring of pedagogical decisions informed by, data, on the perspective of a MOOC. In: *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, vol. 6, p. 15 (2017)
15. Paiva, Ranilson, Bittencourt, Ig Ibert: Helping MOOC teachers do their job. In: Cristea, Alexandra Ioana, Bittencourt, Ig Ibert, Lima, Fernanda (eds.) *HEFA 2017. CCIS*, vol. 832, pp. 52–67. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-97934-2_4

16. Paiva, R., Bittencourt, I.I., Tenório, T., Jaques, P., Isotani, S.: What do students do on-line? Modeling students' interactions to improve their learning experience. *Comput. Hum. Behav.* **64**, 769–781 (2016). <https://doi.org/10.1016/j.chb.2016.07.048>, <http://www.sciencedirect.com/science/article/pii/S0747563216305386>
17. Paiva, R.O.A., Bittencourt, I.I., da Silva, A.P., Isotani, S., Jaques, P.: Improving pedagogical recommendations by classifying students according to their interactional behavior in a gamified learning environment. In: *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, pp. 233–238. ACM (2015)
18. de Santana, S.J., Ranilson Paiva, I.I.B.e.a.: Evaluating the impact of mars and Venus effect on the use of an adaptive learning technology for Portuguese and mathematics. In: *The 16th IEEE International Conference on Advanced Learning Technologies - ICALT2016* (2016)
19. Seaman, J., Allen, I., Seaman, J.: *Grade increase: tracking distance education in the United States*. Babson Survey Research Group, Oakland (2019)
20. Sheridan, S., Mouza, C., Pollock, L.: Professional development for computer science education: design and outcomes from a case study teacher. In: *Society for Information Technology & Teacher Education International Conference*, pp. 1115–1123. Association for the Advancement of Computing in Education (AACE) (2019)
21. Stephenson, J.: *Teaching & Learning Online: New Pedagogies for New Technologies*. Routledge, Abingdon (2018)
22. Teo, T.: Factors influencing teachers' intention to use technology: model development and test. *Comput. Educ.* **57**(4), 2432–2440 (2011)