Development and Validation of an Instrument to Measure the Usability of Educational Artifacts Created with Web 2.0 Applications

Tihomir Orehovački and Nikolina Žajdela Hrustek

University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{tihomir.orehovacki,nikolina.zajdela}@foi.hr

Abstract. The emergence of Web 2.0 applications has provided new opportunities for all participants in the educational process. Students are encouraged to create and share educational artifacts and thereby actively contribute to the development of knowledge repository. On the other hand, teachers are enabled to publish lecture resources, communicate with students, comment on shared and integrated artifacts, and evaluate completed educational e-activities. Considering that usability represents a necessary condition for an effective learning, it affects the adoption and use of created artifacts in e-learning settings. Although Web 2.0 applications are widely used for educational purposes, a consolidated methodology for the assessment of artifacts resulting from their use is still not available. The work presented in this paper is the first step towards a comprehensive framework for evaluating the usability of educational artifacts created with Web 2.0 applications. Following the standard procedure for instrument development, we conducted an empirical study during which specific pedagogical and technical attributes that capture certain usability facets of educational artifacts created with Web 2.0 applications were identified.

Keywords: Web 2.0, Usability Evaluation, Educational Artifacts, Study Results.

1 Introduction

Web 2.0 refers to a shift from the read-only paradigm of static web pages where users were passive recipients of published information towards user-centered web applications that support different kinds of interaction among users [16]. Orehovački et al. [18] distinguish eighteen types of Web 2.0 applications with educational potential that support communication and collaboration among students and enable them to create, organize, share, and integrate various artifacts. Results of two consecutive studies [17][20] pointed out that blogs, social networks, video podcasting services, and wikis are most commonly used types of Web 2.0 applications in the e-learning environment. In addition, findings of a study conducted by Bubaš et al. [4] suggest that integration of educational artifacts created with Web 2.0 applications into wiki, blog, or e-portfolio enables students to create personal learning environment and become active members of a learning community.

A. Marcus (Ed.): DUXU/HCII 2013, Part I, LNCS 8012, pp. 369–378, 2013.

Usability plays an important role in the assessment of Web 2.0 applications and artifacts created by their use. In educational environment, usability is composed of two different but rather complementary concepts. While technical usability deals with the evaluation of pragmatic aspects of Web 2.0 applications, pedagogical usability consists of the assessment of educational benefits gained while using created artifacts. The latest research related to Web 2.0 applications has mainly focused on modeling their adoption [8][25] as well as on their assessment from either the objective [1][5] or the subjective [3][14][16][22] perspective, or both [15][19]. Literature related to the usability evaluation in the educational settings offers diverse frameworks [2][26][27], guidelines [10][23], and instruments [12][28]. Nevertheless, research on evaluation of educational artifacts created with Web 2.0 applications has not attracted enough attention from the HCI community. With an aim to overcome the set forth gap, we initiated a research on the development of a framework that would enable the assessment of various usability facets in the context of educational artifacts created by means of Web 2.0 applications. As a first step towards this goal, we present results of an empirical study carried out to validate an instrument aimed for measuring technical and pedagogical usability dimensions of educational artifacts.

2 Related Work

Current research on usability evaluation in the educational context deals with the development and application of questionnaires, heuristics, and methodologies. The scope of this section is to provide concise overview of their most prominent representatives.

Questionnaires have been widely used in the usability evaluation of e-learning systems and educational artifacts. For instance, Nokelainen [12] developed self-rating Pedagogically Meaningful Learning Questionnaire (PMLQ) composed of 92 multiple-choice items meant for concurrent usability evaluation of a learning management system and digital learning materials. The PMLQ includes ten most widely applied technical usability dimensions (accessibility, learnability and memorability, user control, help, graphical layout, reliability, consistency, efficiency of use, memory load, and errors) as well as ten pedagogical usability criteria (learner control, learner activity, cooperative learning, goal orientation, applicability, added value, motivation, valuation of previous knowledge, flexibility, and feedback). By conducting two empirical studies, Zaharias and Poylymenakou [28] introduced questionnaire-based evaluation method for e-learning applications. The final version of questionnaire contained 49 items grouped into seven separate categories: self-assessment and learnability, interactivity, accessibility, navigation, visual design, learning and support, and content.

Heuristic evaluation is usability inspection technique in which expert evaluators examine the system and determine to what extent it conforms to recognized set of usability principles known as heuristics. Ssemugabi and De Villiers [23] adapted initial set of Nielsen's heuristics [11] on general interface usability (visibility of system status; match between designer and user model; learner control and freedom; consistency and adherence to standards; prevention of peripheral usability-related errors; recognition rather than recall; flexibility and efficiency of use; aesthetics and minimalism in design; recognition, diagnosis, and recovery from errors; and help and documentation) to the e-learning context. They also suggested additional heuristics related to educational websites (simplicity of site navigation, organization and structure; and relevance of site content to the learner and the learning process) and learner-centred instructional design (clarity of goals, objectives, and outcomes; effectiveness of collaborative learning; level of learner control; support for personally significant approaches to learning; cognitive error recognition, diagnosis and recovery; feedback, guidance and assessment; context meaningful to domain and learner; and learner motivation, creativity and active learning). Drawing on prior research, Mehlenbacher et al. [10] proposed a set of heuristics for the design of e-learning environments. The outlined set of heuristics was based on following dimensions of instructional situations: learner background and knowledge (accessibility; customizability and maintainability; error support and feedback; navigability and user movement; and user control, error tolerance, and flexibility), social dynamics (mutual goals and outcomes; and communication protocols), instructional content (completeness; examples and case studies; readability and quality of writing; and relationship with real-world tasks), interaction display (aesthetic appeal; consistency and layout; typographic cues and structuring; and visibility of features and self-description), instructor activities (authority and authenticity; and intimacy and presence), and environment and tools (help and support documentation; metaphors and maps; organization and information relevance; and reliability and functionality).

Existing methodologies intended for the assessment of e-learning applications represent a synthesis of different usability evaluation methods. For example, Milano-Lugano Evaluation method (MiLE) [26] is an experience-based inspection framework that combines scenario-driven and heuristic-based techniques in the evaluation of e-learning web applications. Ardito et al. [2] employed Systematic Usability Evaluation (SUE) [9] methodology that suggests coupling of inspection and user-testing activities thus enabling evaluation of four usability dimensions (presentation, hypermediality, application proactivity, and user activity) of e-learning platforms. MiLE+ [27] presents an evolution of MiLE and SUE into structured usability evaluation framework that is more convenient for novice evaluators than its predecessors.

3 Methodology

3.1 Participants

A total of 102 Information Science students enrolled in the Data Structures course took part in the study. The sample was composed of 81.37% male, and 18.63%

female students. They ranged in age from 19 to 23 years (M = 20.23, SD = .843). At the time the study was carried out, the majority (85.29%) of participants were second-year undergraduate students.

3.2 Procedure

With an objective to validate an instrument for measuring usability of educational artifacts created with Web 2.0 applications, we adopted scale development process proposed by Straub et al. [24]. As a starting point, an initial pool of 47 items was drafted based on the questionnaires, heuristics, and methodologies outlined in previous section. We then conducted a pilot study [21] to explore the appropriateness and perceived meaning of generated pool of items. Drawing on collected data, ten distinguishable usability facets of educational artifacts (completeness, usefulness, availability, learning flow, content quality, added value, adaptability, presentation quality, memorability, and learnability) were identified. In the follow-up to the aforementioned steps, initial pool of items was supplemented with statements inspired by
authors' practical experience in use and evaluation of Web 2.0 applications. Subsequently, an online questionnaire comprising 120 items was designed. Answers to the items were given on a five-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree).

Prior to data collection, each student had to solve four different programming problems by completing following educational e-activities: (a) use mind mapping service to illustrate the decomposition of the programming problem; (b) employ diagramming application to create a flowchart that depicts the algorithmic solution; (c) develop a solution to the programming problem; (d) use screencasting service to narratively explain the main parts of the code and demonstrate how it works; (e) post the snippets of the code on the social network aimed for collaborative programming thus enabling teacher, experts in the field, and other students to provide constructive feedback; (f) integrate all the aforementioned artifacts or links to their locations on the wiki page together with a concise explanation on completed e-activities; and (g) use created artifacts when learning for the midterms. Students' reflections on usability of created and disseminated educational artifacts were obtained by administering an online questionnaire at the end of the semester.

The psychometric properties of the online questionnaire were evaluated in terms of construct validity and reliability. Construct validity determines the extent to which items capture the essence of the underlying construct. In this study, construct validity was established through convergent and discriminant validity. The former is the degree to which items of the underlying factor are in agreement while the latter refers to the extent to which items of different factors are distinct. Convergent and discriminant validity were assessed by means of a principal component analysis (PCA) with equamax rotation and Keiser normalization. They were both verified by observing correlations between items of the extracted factors. The Kaiser-Meyer-Olkin test of sampling adequacy and Bartlett's test of sphericity were examined to check that the requirements for factor extraction were met. An eigenvalue greater than 1 was used as a criterion to determine the number of factors. Only items with factor loadings above .40 and cross-loadings below .40 were retained [7]. Reliability refers to the extent to which a strong mutual association exists among items assigned to the same factor. This study employed Cronbach's Alpha reliability coefficient for measuring internal consistency of extracted factors. Considering a minimum value of .40 [6] the item-total correlation was used to improve the levels of Cronbach's Alpha coefficient.

4 Results

The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO = .693) and Bartlett's test of sphericity ($\chi 2 = 2275.114$, $p = .000$) both confirmed that the data satisfied the requirements for carrying out the principal component analysis (PCA). After the first PCA iteration, items that did not meet the loading cut-off or loaded on more than one factor were dropped. The remaining items were then subjected to another round of PCA. The process was continued until a meaningful factor structure had been reached. A total of 81 items were eliminated after 27 PCA iterations. In addition, 7 items were removed because their item-total correlation was below the .40 threshold. The final PCA iteration uncovered 11 distinct usability dimensions of educational artifacts created with Web 2.0 applications. They accounted for 82.032% of the sample variance. All 32 retained items had a factor loading greater than the .50 cut-off [7]. The communalities ranged from 0.666 to 0.900 with an average value of 0.820, indicating that yielded factors explained an acceptable amount of the variance in each of the items. Cronbach's Alpha values for all extracted factors exceeded the .70 threshold [13]. Since all reported results were above the cut-off values for exploratory research, they provided support for the validity and reliability of identified factor structure.

As a follow-up to afore discussed findings, factors that determine the usability of educational artifacts created with Web 2.0 applications were defined. The first one, *learnability* appeared to be the most important because it explained the largest portion (9.174%) of the total variance after the equamax rotation. It consisted of 4 items meant for measuring ease and pace of learning from the educational artifacts. The second factor, *productivity*, explained 8.432% of the variance. It had 3 items that addressed students' productivity in solving programming problems. The third factor, *content quality*, accounted for 8.146% of the variance. It employed 3 items for evaluating the credibility, truthfulness, and accuracy of the content included in the educational artifacts. The fourth factor, *learning performance*, represented 7.958% of the variance and consisted of 3 items measuring students' learning effectiveness and efficiency. The fifth factor, *consistency*, explained 7.914% of the total variance. It was assessed with 3 items aimed for measuring the extent to which terminology, structure, and layout of the educational artifacts are consistent. The sixth factor, *satisfaction*, explained 7.672% of the variance and used 3 items for evaluating the degree to which the educational artifacts meet students' expectations and needs. The seventh factor, *availability*, accounted for 7.349% of the variance. It applied 3 items to measure the extent to which the educational artifacts are continuously available. The eighth factor, *added value*, represented 7.226% of the total variance. It encompassed 4 items for evaluating pedagogical benefits of the educational artifacts created with certain types of Web 2.0 applications. The ninth factor, *usefulness*, accounted for 6.517% of the variance. It employed 2 items for assessing the degree to which the educational artifacts affect existing students' knowledge and are valuable for learning theoretical course concepts. The tenth factor, *adaptability*, represented 6.105% of the variance and consisted of 2 items measuring the extent to which the educational artifacts are customized to students' learning style and pace. The last factor, *learning flow*, explained 5.540% of the total variance and consisted of 2 items that addressed students' concentration when learning from the educational artifacts.

It should be noted that learnability, content quality, consistency, satisfaction, availability, and adaptability are technical usability aspects while productivity, learning performance, added value, usefulness, and learning flow are pedagogical usability facets of the educational artifacts created with Web 2.0 applications. The summary of the study findings is presented in Table 1 (see Appendix).

5 Conclusion

The work reported in this paper was motivated by the need to address the specificities of measuring the usability of educational artifacts created with Web 2.0 applications. Therefore, we initiated a study on the development and evaluation of a multi-dimensional scale. The results of the empirical validation revealed six technical and five pedagogical usability dimensions and provided evidence for the validity and reliability of the 32-item instrument.

As with most empirical studies, our work has some limitations. The employment of a homogenous group of students as study participants represents limitation to the generalizability of the results. Heterogeneous sample of students may have decidedly different attitudinal structure with respect to usability dimensions of created educational artifacts. Another limitation is the applicability of the results to educational artifacts created by means of Web 2.0 applications in general. It is likely that the type of Web 2.0 application is moderating the relationships between generated artifacts and their perceived usability.

In spite of study limitations, this paper provides valuable guidelines for assessing educational artifacts created with Web 2.0 applications. Teachers of the programming-related courses can use the developed instrument to examine and improve the usability of created educational artifacts. In addition, the validated instrument adds to extant body of knowledge by establishing a basis for further advances on evaluating educational artifacts created with Web 2.0 applications. Given that this paper presents results of an on-going research, our future work will be focused on modeling the interplay among identified usability dimensions of educational artifacts created by means of Web 2.0 applications.

References

1. Almeida, J.M., Gonçalves, M.A., Figueiredo, F., Pinto, H., Belém, F.: On the Quality of Information for Web 2.0 Services. IEEE Internet Computing 14(6), 47–55 (2010)
2. Ardito, C., Costabile, M.F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T., Rossano, V.: An approach to usability evaluation of e-learning applications. Universal Access in the Information Society 4(3), 270–283 (2006)
3. Brown, A., Jay, C., Chen, A.Q., Harper, S.: The uptake of Web 2.0 technologies, and its impact on visually disabled users. Universal Access in the Information Society 11(2), 185–199 (2012)
4. Bubaš, G., Ćorić, A., Orehovački, T.: The integration of students' artifacts created with Web 2.0 tools into Moodle, blog, wiki, e-portfolio and Ning. In: Proceedings of the 34th International Convention MIPRO, pp. 1084–1089. IEEE Press, Opatija (2011)
5. Cappiello, C., Daniel, F., Koschmider, A., Matera, M., Picozzi, M.: A Quality Model for Mashups. In: Auer, S., Díaz, O., Papadopoulos, G.A. (eds.) ICWE 2011. LNCS, vol. 6757, pp. 137–151. Springer, Heidelberg (2011)
6. Churchill, G.A.: A paradigm for developing better measures of marketing constructs. Journal of Marketing Research 16(1), 64–73 (1979)
7. Hair Jr., J.F., Black, W.C., Babin, B.J., Anderson, R.E.: Multivariate Data Analysis, 7th edn. Prentice Hall, Englewood Cliffs (2009)
8. Hartshorne, R., Ajjan, H.: Examining student decisions to adopt Web 2.0 technologies: theory and empirical tests. Journal of Computing in Higher Education 21(3), 183–198 (2009)
9. Matera, M., Costabile, M.F., Garzotto, F., Paolini, P.: SUE Inspection: An Effective Method for Systematic Usability Evaluation of Hypermedia. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 32(1), 93–103 (2002)
10. Mehlenbacher, B., Bennett, L., Bird, T., Ivey, M., Lucas, J., Morton, J., Whitman, L.: Usable E-Learning: A Conceptual Model for Evaluation and Design. In: Proceedings of the 11th International Conference on Human-Computer Interaction, pp. 1–10. Lawrence Erlbaum Associates, Las Vegas (2005)
11. Nielsen, J.: Heuristic evaluation. In: Nielsen, J., Mack, R.L. (eds.) Usability Inspection Methods, pp. 25–62. Wiley & Sons, New York (1994)
12. Nokelainen, P.: Conceptual Definition of the Technical and Pedagogical Usability Criteria for Digital Learning Material. In: Cantoni, L., McLoughlin, C. (eds.) Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004, pp. 4249–4254. AACE, Chesapeake (2004)
13. Nunnally, J.C.: Psychometric Theory. McGraw-Hill, New York (1978)
14. Orehovački, T.: Proposal for a Set of Quality Attributes Relevant for Web 2.0 Application Success. In: Proceedings of the 32nd International Conference on Information Technology Interfaces, pp. 319–326. IEEE Press, Cavtat (2010)
15. Orehovački, T.: Development of a Methodology for Evaluating the Quality in Use of Web 2.0 Applications. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 382–385. Springer, Heidelberg (2011)
16. Orehovački, T.: Perceived Quality of Cloud Based Applications for Collaborative Writing. In: Pokorny, J., et al. (eds.) Information Systems Development – Business Systems and Services: Modeling and Development, pp. 575–586. Springer, Heidelberg (2011)

17. Orehovački, T., Bubaš, G., Konecki, M.: Web 2.0 in Education and Potential Factors of Web 2.0 Use by Students of Information Systems. In: Proceedings of the 31st International Conference on Information Technology Interfaces, pp. 443–448. IEEE Press, Cavtat (2009)
18. Orehovački, T., Bubaš, G., Kovačić, A.: Taxonomy of Web 2.0 Applications with Educational Potential. In: Cheal, C., Coughlin, J., Moore, S. (eds.) Transformation in Teaching: Social Media Strategies in Higher Education, pp. 43–72. Informing Science Press, Santa Rosa (2012)
19. Orehovački, T., Granić, A., Kermek, D.: Exploring the Quality in Use of Web 2.0 Applications: The Case of Mind Mapping Services. In: Harth, A., Koch, N. (eds.) ICWE 2011. LNCS, vol. 7059, pp. 266–277. Springer, Heidelberg (2012)
20. Orehovački, T., Konecki, M., Radošević, D.: Web 2.0 technologies in university education. In: Proceedings of the 31st MIPRO International Convention on Computers in Education, pp. 269–273. MIPRO, Opatija (2008)
21. Orehovački, T., Žajdela Hrustek, N.: Towards a Framework for Usability Evaluation of Educational Artifacts created with Web 2.0 Applications: A Pilot Study. In: Proceedings of the 36th International Convention MIPRO, pp. 691–696. IEEE Press, Opatija (2013)
22. Pang, M., Suh, W., Hong, J., Kim, J., Lee, H.: A New Web Site Quality Assessment Model for the Web 2.0 Era. In: Murugesan, S. (ed.) Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications, pp. 387–410. IGI Global, Hershey (2010)
23. Ssemugabi, S., De Villiers, R.: Usability and Learning: A Framework for Evaluation of Web-Based e-Learning Applications. In: Montgomerie, C., Seale, J. (eds.) Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, pp. 906–913. AACE, Chesapeake (2007)
24. Straub, D., Boudreau, M., Gefen, D.: Validation Guidelines for IS Positivist Research. Communications of the Association for Information Systems 13(1), 380–427 (2004)
25. Suki, N.M., Ramayah, T., Ly, K.K.: Empirical investigation on factors influencing the behavioral intention to use Facebook. Universal Access in the Information Society 11(2), 223–231 (2012)
26. Triacca, L., Bolchini, D., Botturi, L., Inversini, A.: MiLE: Systematic Usability Evaluation for E-learning Web Applications. In: Cantoni, L., McLoughlin, C. (eds.) Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004, pp. 4398–4405. AACE, Chesapeake (2004)
27. Triacca, L., Inversini, A., Bolchini, D.: Evaluating Web usability with MiLE+. In: Proceedings of the 7th IEEE International Symposium on Web Site Evolution, pp. 22–29. IEEE, Lugano (2005)
28. Zaharias, P., Poylymenakou, A.: Developing a Usability Evaluation Method for e-Learning Applications: Beyond Functional Usability. International Journal of Human-Computer Interaction 25(1), 75–98 (2009)

Appendix

Table 1. Summary of the study findings

Items	Loading	Mean	SD
Learnability (Cronbach's α = .886)			
Students learn easier from the educational artifacts than they do from the course-related books.	.831	2.10	1.067
Students learn faster from the educational artifacts than they do from the course-related books.	.806	2.10	1.058
Students learn faster from the educational artifacts than they do from the resources published on the LMS*.	.787	2.85	.938
Students learn easier from the educational artifacts than they do from the resources published on the LMS*.	.765	2.97	.959
Productivity (Cronbach's α = .926)			
The educational artifacts increase students' effectiveness in solving the programming problems.	.898	2.29	.918
The educational artifacts increase students' productivity in solving the programming problems.	.880	2.35	.971
The educational artifacts increase students' efficiency in solving the programming problems.	.843	2.24	.869
Content Quality (Cronbach's α = .905)			
The content of the educational artifacts is trustworthy.	.917	1.82	.604
The content of the educational artifacts is true.	.909	1.76	.632
The content of the educational artifacts is accurate.	.839	1.83	.646
Learning Performance (Cronbach's α = .877)			
The educational artifacts enhance students' learning effectiveness.	.884	2.09	.785
The educational artifacts enhance students' learning efficiency.	.862	2.09	.834
The educational artifacts enhance students' learning productivity.	.832	2.17	.759
Consistency (Cronbach's α = .849)			
The terminology used in the educational artifacts is consistent.	.856	2.17	.661
The educational artifacts have consistent structure.	.854	2.11	.659
The educational artifacts have consistent layout.	.829	2.13	.792

* Learning Management System

Table 1. (*Continued.*)

Items	Loading	Mean	SD
Satisfaction (Cronbach's α = .884)			
The quality of the educational artifacts is satisfactory.	.836	2.11	.782
The content of the educational artifacts meets students' expectations.	.833	2.24	.811
The manner in which the educational artifacts present the course topics meets students' needs.	.731	2.23	.819
Availability (Cronbach's α = .846)			
Students can access the educational artifacts whenever they want to do so.	.883	1.52	.521
The educational artifacts are available when students need them.	.859	1.62	.527
The educational artifacts are continuously available.	.842	1.61	.566
Added Value (Cronbach's α = .846)			
Shared code snippets help students to develop their own solution to the programming problem.	.775	2.02	1.062
Mind map helps students to understand the relationship between the programming problem and its solution.	.709	2.15	.999
Screencast helps students to understand the relationship between theoretical and practical aspects of the course.	.687	2.34	1.039
Flowchart helps students to design the logical structure of the solution to the programming problem.	.681	2.04	.964
Usefulness (Cronbach's α = .868)			
The educational artifacts are advantageous for learning the data structures concepts.	.876	1.97	.777
The educational artifacts alter existing students' knowledge.	.824	1.91	.810
Adaptability (Cronbach's α = .815)			
The educational artifacts are adapted to the students' learning pace.	.821	2.61	.881
The educational artifacts are adapted to the students' learning style.	.817	2.86	.890
Learning Flow (Cronbach's α = .750)			
When learning from the educational artifacts, students are not aware of any noise.	.881	3.43	.939
When learning from the educational artifacts, students do not realize the time elapsed.	.869	3.36	1.124