Evaluation of Serious Games: A Holistic Approach

Citation for published version (APA):

Steiner, C., Hollins, P., Kluijfhout, E., Dascalu, M., Albert, D., Westera, W., & Nussbaumer, A. (2015). Evaluation of Serious Games: A Holistic Approach. In L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.), *ICERI 2015 Proceedings: 8th International Conference of Education, Research and Innovation Seville, Spain. 18-20 November, 2015.* (pp. 4334-4342). IATED Academy. ICERI Proceedings https://library.iated.org/view/STEINER2015EVA

Document status and date:
Published: 04/10/2015

Document Version:
Peer reviewed version

Document license:
CC BY-NC-SA

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

https://www.ou.nl/taverne-agreement

Take down policy
If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from https://research.ou.nl/ on date: 29 Nov. 2021

Open Universiteit
www.ou.nl

Evaluation of Serious Games: A Holistic Approach

Christina M. Steiner[1], Paul Hollins[2], Eric Kluijfhout[3], Mihai Dascalu[4], Alexander Nussbaumer[1], Dietrich Albert[1], & Wim Westera[3]

[1]Graz University of Technology (AUSTRIA)
[2]University of Bolton (UNITED KINGDOM)
[3]Open University of the Netherlands (NETHERLANDS)
[4]University Politehnica of Bucharest (ROMANIA)

Abstract

Digital games constitute a major emerging technology that is expected to enter mainstream educational use within a few years. The highly engaging and motivating character of such games bears great potential to support immersive, meaningful, and situated learning experiences. To seize this potential, meaningful quality and impact measurements are indispensible. Although there is a growing body of evidence on the efficacy of games for learning, evaluation is often poorly designed, incomplete, biased, if not entirely absent. Well-designed evaluations demonstrating the educational effect as well as the return on investment of serious games may foster broader adoption by educational institutions and training providers, and support the development of the serious game industry. The European project RAGE introduces a comprehensive and multi-perspective framework for serious game evaluation, which is presented in this paper.

Keywords: Serious games, evaluation, empirical evidence, game development, learning effectiveness.

1 INTRODUCTION

Serious games or so-called applied games serve a primary purpose that goes beyond the aspect of pure entertainment. There are multiple genres of serious games based on their outcomes, out of which the most common in this field are learning (or educational) games. Digital learning games represent an e-learning technology that is increasingly recognized by educational practitioners [1]. With their highly engaging and motivating character games constitute effective educational tools for creating authentic learning tasks and meaningful, situated learning [2]. One main reason why games can be so effective for learning is their ability to induce a 'flow experience' – a positively perceived experience and state of full immersion in an activity that typically goes along with a loss of sense of time [3].

Digital learning games correspond to the current Zeitgeist of using information technology as an integral part of our everyday life, and they meet the trend of pedagogical paradigms calling for active, constructive, and playful learning. Serious games are considered a major emerging technology that is expected to enter mainstream use as educational tool in K-12 and higher education within the next two to three years ([1], [4]). The market potential of this kind of learning technology is not yet fully exploited. Reasons for that are, among others, the high effort required for the creation of successful learning games and the challenge, as well as the lack of thorough impact measurements. Although there is a growing body of evidence on the efficacy of games for learning, evaluation is often poorly designed, incomplete, biased, if not entirely absent (e.g. [5]). Evaluations rarely consist in randomised controlled trials. Furthermore, methodological flaws consist in the common use of only post game experience questionnaires; these are often applied for reasons of ease, simplicity or ignorance about alternative methods, but their scope is largely qualitative (e.g. player attitude) and their validity may be questioned, as item validation is often neglected. A critical aspect of research on the effectiveness of educational games is, in fact, how to approach and operationalize the measurement methodologies.

The European RAGE project (http://rageproject.eu/) aims at fostering the adoption of digital game-based learning in game industry and in education. In the context of RAGE, a holistic and multi-perspective framework for serious game evaluation is developed, which is described in the present paper. The framework also serves as a common reference point and guidance for investigating and demonstrating the quality and benefits of the achieved research and technology outputs.

The remainder of this paper is structured as follows: Section 2 gives an overview of the objectives of RAGE and outlines the challenges existing in the evaluation of the related project achievements. Subsequently, section 3 summarises relevant related work that has inspired and has been incorporated in the comprehensive RAGE evaluation framework, which is presented in more detail in section 4. Section 5 presents conclusions and an outlook to future work.

2 THE RAGE PROJECT AND ITS EVALUATION CHALLENGES

The primary objective of the RAGE project is to make available and accessible advanced software tools, methodologies, and expertise for serious games development and application [6]. Two main groups of stakeholders are targeted: game developers, on the one hand, and training providers and their learners, on the other hand. Serious game industry build-up shall be supported with advanced technology and know-how, for easier, faster, and more cost-effective development of serious games. In this way, game development for education and training shall be boosted, and thus, the use of games to support skill development and knowledge acquisition.

RAGE will provide a collection of reusable, and interoperable software components (so called “gaming assets”) for game development, which are currently under development. These assets will provide functionalities to undertake various data analyses, like competence assessment, emotion detection, comprehension measurement, or motivation identification. Another group of assets will enable game intelligence and adaptation, e.g. in terms of competence-based personalisation, natural language processing, motivational adaptation, cognitive interventions, and social gamification. The gaming assets will be provided via an online “ecosystem”, which will also make available a broad range of literature and training material, as well as collaboration tools. The RAGE ecosystem will therefore serve as a central access point and affinity social space. The RAGE technologies will be applied and tested in six specific, asset-based learning games (mobile and desktop implementations). The games will address different types of employability skills.

The tools and methods that RAGE produces and will make available are of interest to the wider serious gaming research and industry communities as they strive to improve the quality of serious games. To demonstrate the effectiveness of these research and development outcomes, and to ensure that they meet the needs of industrial and educational stakeholders, a comprehensive and multi-perspective evaluation approach is required.

Aside from a systematic analysis of the games’ effectiveness for learning, which is traditionally done in serious game evaluations (e.g. [5]), the broader benefit for training providers or educational organisations will be taken into account in assessing empirical evidences. This is also necessary for being able to seize the great potential of serious games, in general. However, focusing purely on the effectiveness of games would not be sufficient for gaining a comprehensive understanding of the added value of RAGE technologies for the game industry. Therefore, the underlying processes of using these new tools and methods for actual game development will be additionally subjected to evaluation.

3 RELATED WORK

When aiming at addressing the perspective of both the gaming industry and the educational stakeholders in evaluation, relevant related work covers research on the empirical evidence about the impact and outcomes of serious games. In addition, literature about the evaluation of game authoring tools as well as of digital repositories is relevant, to serve as an inspiration for assessing RAGE gaming assets and ecosystem.

Since educational games are fundamentally different from traditional learning environments or other software products, evaluation approaches valid for those applications may fall short when used in serious games evaluation. Universal evaluation frameworks for e-learning (e.g. [7]), or for training programmes (e.g. [8], [9]) may only serve as a starting point for assessing serious games. Given the complexity of digital game environments and the embedding of non-leisure and learning purposes in the game, there is a need to select and adapt suitable evaluation methodologies. Several evaluation models or frameworks have been suggested in the literature to specifically frame the research and evaluation of serious games (e.g. [10], [11], [12], [13]).

Evaluation goals in the context of serious games are usually two-fold, aiming at the measurement of the software quality of the game, on the one hand, and at the assessment of its effectiveness in terms of reaching their goals of learning and engagement (in a wider sense), on the other hand. As a result, usability, learning effectiveness and game enjoyment are the evaluation criteria commonly addressed. Usability in the context of (serious) games is referred to as the degree to which a player is able to learn, control and understand a game [14]. Techniques applied for usability evaluation cover heuristics, think-aloud user testing (e.g. [15]) and observational methods (e.g. [16]). Learning, i.e. the educational effectiveness of games, is typically evaluated by applying a pre- and post-test design, i.e. the assessment of learning outcomes of a certain unit of study (e.g. [17]). Alternative approaches consist of the use of self-reports, where people are asked to indicate what they feel they have learned from undertaking an activity (e.g. [18]), or of built-in assessment procedures of the educational game simulation (e.g. [19]). User engagement, flow, satisfaction and motivation are aspects subsuming a range of attributes related to the subjective experience and enjoyment of games (e.g. [20], [21]). Common approaches to evaluate engagement, motivation and other aspects of user experience are questionnaires or interviews (e.g. [22]), attendance rates, measurement of (voluntary) time-on-task (e.g. [23]). More sophisticated techniques include observations [24] or non-intrusive assessment based on users' interaction with the system (e.g.[25]).

The RAGE gaming assets aim at supporting serious game creation and development; they will provide authoring tools for entering relevant domain data and for including and configuring features of game analytics and intelligence. Although these authoring tools and the game development process in RAGE will be quite different from common content authoring, evaluation approaches applied for conventional e-learning and game authoring software may inspire the evaluation of gaming assets. Authoring tools for course or game development addresses professional instructional or game designers and developers, but may also aim at supporting pedagogical practitioners and content providers [26]. These different user groups have different levels of expertise in programming and game authoring, and therefore have also different needs and expectations towards authoring tools, which need to be taken into account in the evaluation of the quality and the benefits of this kind of software. The most commonly addressed evaluation topic is usability of these authoring tools, covering different aspects of a tool's suitability, effectiveness and efficiency for a given task, ease of use and learnability, as well as user satisfaction (e.g. [29]). Standardised usability scales or heuristic checklists provide systematic instruments for evaluating usability features (e.g. [27], [28]). The research design oftentimes consists in task-based evaluations, presenting an authoring task to evaluation participants, who carry out the task themselves then or, alternatively, by giving instructions to other persons who operate the system (e.g. [26], [30]). Data collected through standard questionnaires is usually complemented by more in-depth feedback gathered through think-alouds or focus groups (e.g. [26], [30]). These also allow establishing a better understanding of the authoring process and how users experience and use the software. In this way, more detailed information about the specific benefits for authoring can be captured. At earlier stages of development evaluation is sometimes also performed through cognitive walkthroughs [31].

The RAGE ecosystem constitutes a combination of a digital library, media archive, and software repository. As a result, evaluation methods used for this kind of repositories and environment provide a useful starting point for framing the evaluation of the ecosystem. Software and media repositories store reusable assets and make them available. Digital libraries and virtual research environments (VRE) are digital repositories equipped with a variety of additional tools supporting users in the exploration, search and interaction with repository contents, like e.g. cultural artefacts. In case of such software or media repositories, software technological aspects are oftentimes of key interest, while aspects of user interaction and experience sometimes remain off-stage. Most research therefore focuses on technical details and methods for storing and retrieving repository content, while the evaluation of the effectiveness of a repository is oftentimes rather informal or vague [32]. Evaluation approaches for digital repositories and VREs fall into three main categories: a) user-oriented evaluations addressing users' requirements, preferences, interaction and satisfaction with a VRE; b) system-oriented evaluations focusing on technological aspects of digital information representation and retrieval (e.g. precision, recall); and c) systematic evaluations covering user-oriented as well as system-oriented evaluation goals [33]. Methods for evaluating software repositories include gathering direct feedback from repository users or managers, for example via questionnares or structured interviews (e.g. [34], [35]), or expert evaluations against a pre-defined set of evaluation criteria, like scalability, extensibility, interoperability, ease of deployment etc. (e.g. [36]).

4 A COMPREHENSIVE MULTI-PERSPECTIVE EVALUATION FRAMEWORK FOR APPLIED GAMING TECHNOLOGY

4.1 The Framework in General

The comprehensive RAGE evaluation framework integrates the perspectives of the different stakeholder groups present in the project. The central goal is to collect evidence for the effectiveness of serious game technologies in a scientific and methodologically sound way. Fig. 1 presents an overview of the evaluation framework. Evaluations will address both levels of stakeholders and benefits for these groups: game developers, on the one hand, and end users, represented by the actual learners/gamers as well as by the educational providers/institution, on the other hand. The evaluation framework thus includes two main dimensions of evaluation: game development and learning (see Fig. 1). This multi-perspective approach described in more detail, per level, in subsequent sections, will yield a holistic understanding of the quality and impact of the serious game technologies, from the asset-based game development process, to the actual interactions with and impact of the resulting serious games.

All evaluation data will be collected in the context of the project's application cases. The evaluation framework provides the common reference point, with shared methodologies across the RAGE pilots, where possible and appropriate, while nevertheless providing flexibility to accommodate to the specific conditions of the individual application scenarios. With regards to evaluation instruments, the framework calls for a mixed-method approach in evaluations, enabling the integration and triangulation of qualitative and quantitative data from multiple sources and perspectives.

The framework covers a cyclic approach aligned with both asset and game development phases (e.g. [37]). Individual game assets will be thoroughly validated before being integrated in specific serious games. In addition to formative and summative evaluations during and, respectively, at the end of the game development process, preprototype and baseline evaluations (e.g. [39]) will be conducted to incorporate participatory design ideas [38] and to gather benchmark data for later comparative analyses on the impact of project outputs.

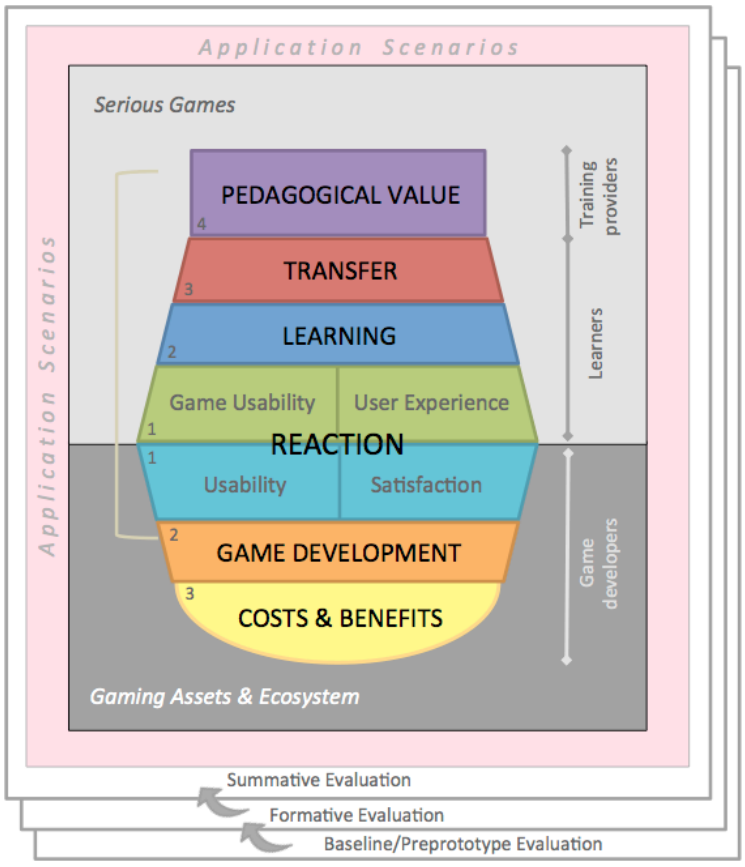

Fig. 1: The RAGE evaluation framework.

4.2 The Learning Dimension

Evaluations addressing the learning dimension, i.e. the educational effectiveness of serious games, adapt Kirkpatrick's [8] idea of an integral evaluation in terms of a four level process (cf. Fig. 1). The Kirkpatrick model forms the main theoretical foundation that has guided the elaboration of the evaluation framework and thus, it constitutes the basis for gathering comprehensive proof of the significance and added value of the addressed gaming technologies. On the first level, evaluation shall target the degree to which learners react favourably to a serious game. This reaction level entails two facets, perceived software quality operationalized by usability, as well as user experience and game enjoyment including variables like satisfaction, engagement or flow. Level two relates to the intended learning objectives and outcomes of the scenario in question. On this level, evaluation will investigate whether and to what degree learners acquire the targeted knowledge and skills by interacting with the serious game. Level three addresses the question whether learners are able to apply the knowledge and competences acquired during gaming in real world settings (transfer). Evaluation at this level is very challenging and ways for actually capturing at least partial evidence on this level still need to be further explored and implemented with the evolvement of the concrete application scenarios. At level four, evaluation addresses the organisational or institutional perspective, in terms of the pedagogical value and benefit of the serious games for training providers and/or educational institutions. This is slightly different to the 'results' level of the original Kirkpatrick model and includes subjective perception and reaction on the serious games' pedagogical effectiveness. This evaluation covers the perspective of the stakeholder group, their experience of the co-design process (i.e. their involvement in the game design), as well as an analysis of costs and benefits for introducing and using this type of learning technology.

A two-tier approach of evaluating the learning dimension is proposed in order to realise a systematic investigation of specific evaluation questions and effects on learning, on the one hand, as well as the analysis and demonstration of the significance in educational practice, on the other hand. In terms of the concrete research designs this means that a mix of real-life pilot studies and laboratory studies will be deployed. In order to obtain sound evidence of the effect of serious games on learning and transfer (level two and three) a comparative approach using baseline (pre-post) measurements or control groups designs shall be implemented.

Emerging technologies for unobtrusive data tracking and sensoring enable real-time, in-game data collection (e.g. [40]) and are made an integral part of evaluation in RAGE. Gaming assets providing different kinds of learning analytics and built into serious games – apart from their use for reporting learning success and dynamic adaptation within the game – will also serve assessing the effectiveness of the educational game itself. A dedicated software asset for in-game evaluation is under development and will be made available. This evaluation asset represents an instrument for continuous evaluation of the quality of learning games by providing insights to users' perception of games and their progress towards game goals. This is done by translating log and sensor data into meaningful information about game quality, user experience, and learning based on pre-defined, configurable evaluation metrics. The asset will facilitate the use of analytics for game evaluation purposes and will advance evaluation methods for serious games by complementing traditional instruments (such as questionnaires). This will enable a meaningful triangulation and cross-validation of different data sources and types, to derive more conclusive evidence on the quality and effect of the evaluated serious games.

4.3 The Game Development Dimension

The perspective of evaluating game development refers to questions on whether the provided game assets facilitate the creation of games for education and training and, in the end, render the serious game market more attractive. Similar to the learning dimension, evaluation of the impact on game development will cover several levels, as illustrated in Fig. 1. In this case, the evaluation objects are given by the gaming assets, on one hand, and the ecosystem making them available, on the other. Evaluation data on these technologies will be collected separately for individual (or groups of assets) and the ecosystem. While a meaningful evaluation of the ecosystem is only possible with an appropriate collection of software and media assets, feedback on the ecosystem shall be gathered in a corpus-agnostic manner. In contrast, assets may be evaluated both in the context of the ecosystem as central access point, or independent of it.

The first level of evaluation relates to the reaction of game developers in terms of usability perception and satisfaction (including aspects of usefulness and user acceptance) with respect to the gaming

assets and the ecosystem. This level may further be operationalized by adopting an evaluation model from the field of digital libraries and VRE [34], which identifies evaluation variables related to reaction on the interaction axes between users, system and user repository content. While data on this level may be gathered to a large extend by standard and customised survey instruments, evaluation at higher levels requires more extensive data collection and intensive dialogue with game developers and other stakeholders. The second level focuses on the analysis of the actual impact of technology use on game development. This refers to the perceived value of the pedagogical functionality provided and the added value for the game development process (benefits, which kind of design/development problems can be addressed). The collection of evaluation data will largely be framed by the game development for the application cases. In addition, specific task-based evaluations may be conducted to examine in more detail specific evaluation questions on individual assets. In the context of level two also the aspect of co-design may be considered, i.e. the extent to which training providers are involved in and co-creating the design of a serious game, and the ways game developers' experience the co-design processes. This, in turn, links back to level four of the learning dimension (see Fig. 1). The third level of game development evaluation refers to the aspect of the 'costs' for integrating and applying the game technologies and methodologies (e.g. [41]), and whether and how these can be balanced by their added value. This assessment is achieved by conducting systematic cost-benefit analysis for the use-cases (e.g. [42]), in order to examine the cost-effectiveness and the market readiness of game technologies. The synthesis of different approaches and perspectives in this analysis provides a thorough understanding of the benefits, opportunities, and challenges of the RAGE approach, in order to provide evidence of the congruency with market demands and developments, as well as a basis for potential future exploitation.

5 CONCLUSION AND FUTURE WORK

Due to the increasing interest in improving the quality of serious games, the tools and methods currently under development in the RAGE project will be of interest to wider applied gaming communities, both research and industry centered. Therefore, seizing the potential of serious games in terms of meaningful quality and impact assessment is indispensible. This paper has presented the background and elaboration of a comprehensive evaluation framework that has been defined to accommodate sound quality and impact measurements and that may serve as best practice for future serious game evaluations. It integrates multiple perspectives into a holistic approach for evaluating serious game technologies. Two main dimensions of evaluation are covered – the evaluation of technologies aiming at supporting serious game development and the evaluation of effectiveness of the resulting serious games. Each of these dimensions may be evaluated on different levels with the relevant stakeholders, i.e. game industry and, respectively, learners or their education providers.

The presented framework creates the basis for analysing the effectiveness of game technologies developed in the RAGE project and beyond, and for ensuring that they reflect the needs of industrial and educational stakeholders. By using this framework as a reference point, research designs will be established and implemented for creating scientifically sound, iterative, and mixed-method evaluations of project outcomes. This is accompanied by the definition of evaluation guidelines and a data management plan providing a standard procedure for organising and carrying out evaluation studies (e.g. administration of informed consents) and how data shall be processed, archived, and preserved. The procedures developed will also be fully compliant with national and European regulations on ethics. The evaluation data generated, collected, and processed in RAGE will be made openly accessible as part of the EU open research data pilot, thus making research reproducible and providing the possibility for further use and analysis by other researchers and future projects.

ACKNOWLEDGEMENT

This work has been partially funded by the EC H2020 project RAGE (Realising and Applied Gaming Eco-System); http://www.rageproject.eu/; Grant agreement No 644187. This document reflects only the views of the authors and the European Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2014). NMC Horizon Report: 2014 Higher Education Edition. Austin: The New Media Consortium.

[2] De Freitas, S. 2013. Learning in immersive worlds. A review of game-based learning. JISC E-learning programme. Retrieved March 1, 2013 from http://www.jisc.ac.uk/media/documents/programmes/elearninginnovation/gamingreport_v3.pdf

[3] Csikszentmihalyi, M., 1990. Flow: The Psychology of Optimal Experience. Harper Perennial, New York.

[4] Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2014). NMC Horizon Report: 2014 K-12 Edition. Austin: The New Media Consortium

[5] Connolly, T.M., Boyle, E.A, MacArthur, E., Hainey, T, & Boyle, J.M. (2013). A systematic literature review of empirical evidence on computer games and serious games. Computers & Education, 59, 661-686.

[6] Hollins, P., Westera, W., & Manero Iglesias, B. (2015). Amplifying applied game development and uptake. Paper accepted at ECGBL 2015.

[7] Cho, Y., Park, S., Jo, S.J., Jeung, C.-W., Lim, D.H. (2009). Developing an integrated evaluation framework for e-learning. In V.C.X. Wang (Ed.), Handbook of research on e-learning applications for career and technical education: Technologies for vocational training (pp. 707-722). Hershey: IGI Global.

[8] Kirkpatrick, D.L. & Kirkpatrick, J.D. (2009). Evaluating training programs. San Francisco: Berrett-Koehler Publishers.

[9] Pawson, R. & Tilley, N. (1997). Realistic evaluation. London: SAGE Publications.

[10] Connolly, T., Stansfield, M., & Hainey, T. (2009). Towards the development of a games-based learning evaluation framework. In T.M. Connolly, M.H. Stansfield, & E. Boyle (Eds.), Games-based learning advancement for multisensory human computer interfaces: Techniques and effective practices. Hershey: Idea-Group Publishing.

[11] De Freitas, S. & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? Computers & Education, 46, 249-264.

[12] Kriz, W.C. & Hense, J.U. (2006). Theory-oriented evaluation for the design of and research in gaming and simulation. Simulation & Gaming, 37, 268-283.

[13] Mayer, I., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., van Ruijven, T. et al. (2013). The research and evaluation of serious games: Toward a comprehensive methodology. British Journal of Educational Technology, 45, 502-527.

[14] Pinelle, D., Wong, N., & Stach, T. (2008). Heuristic evaluation of games: Usability principles for video game design. In Proceedings of CHI 2008 (pp. 1453-1462), Florence, Italy.

[15] Desurvire, H., Caplan, M., & Toth, J.A. (2004). Using heuristics to evaluate the playability of games. In: Extended Abstracts of CHI 2004 (pp. 1509–1512). ACM Press, New York.

[16] Moreno-Ger, P., Torrente, J., Hsieh, Y.G., & Lester, W.T. (2012). Usability testing for serious games: Making informed design decisions with user data. Advances in Human-Computer Interaction, 2012, Article ID 369637.

[17] Ebner, M. & Holzinger, A. (2007). Successful implementation of user-centred game-based learning in higher education: An example from civil engineering. Computers & Education, 49, 873-890.

[18] Whitton, N.J. (2007). An investigation into the potential of collaborative computer game-based learning in higher education. PhD Thesis: Napier University, UK

[19] Wesiak, G., Steiner, C., Moore, A., Dagger, D., Power, G., Berthold, M., Albert, D., & Conlan, O. (2014). Iterative augmentation of a medical training simulator: Effects of affective metacognitive scaffolding. Computers & Education, 76, 13-29.

[20] Boyle, E.A., Connolly, T.M., Hainey, T., & Boyle, J.M. (2012). Engagement in digital entertainment games: A systematic review. Computers in Human Behavior, 28, 771-780.

[21] Law, E. L.-C., Hvannberg, E.T., & Hassenzahl, M. (2006) (Eds.). User experience. Towards a unified view. The 2nd COST294-MAUSE International Open Workshop. Oslo, Norway.

[22] Song, S.H., & Keller, J.M. (2001). Effectiveness of motivationally adaptive computer-assisted instruction on the dynamic aspects of motivation. Educational Technology Research and Development, 49, 5-22.

[23] Chapman, E. (2003). Alternative approaches to assessing student engagement rates. Practical Assessment, Research & Evaluation, 8(13).

[24] Law, E. L.-C. & Sun, X. (2012). Evaluating user experience of adaptive digital educational games with Activity Theory. International Journal of Human-Computer Studies, 70, 478-497.

[25] Cocea, M. & Weibelzahl, S. (2009). Log file analysis for disengagement detection in e-Learning environments. User Modeling and User-Adapted Interaction, 19, 341-385.

[26] Mehm, M., Göbel, S., Radke, S., & Steinmetz, R. (2009). Authoring environment for story-based digital educational games. In: M.D. Kickmeier-Rust (Ed.), Proceedings of the 1st International Open Workshop on Intelligent Personalization and Adaptation in Digital Educational Games (pp. 113-124). Graz, Austria.

[27] Prümper, J. (1993). Software-Evaluation based upon ISO 9241 Part 10. In: T. Grechenig & M. Tscheligi (Eds.), Human Computer Interaction (pp. 255-265). Berlin: Springer.

[28] Ibrahim, L.F.Md. & Yatim, M.H.M. (2013). Heuristic evaluation of children's authoring tool for game making. Journal of Education and Vocational Research, 4, 259-264.

[29] Dag, F., Durdu, L., & Gerdan, S. (2014). Evaluation of educational authoring tools for teachers stressing of perceived usability features. Procedia – Social and Behavioral Sciences, 116, 888-901.

[30] Gaffney, C., D. Dagger, & Wade, V. (2008). Evaluation of ACTSim: A composition tool for authoring adaptive soft skill simulations. In W. Nejdl, J. Kay, P. Pearl, & E. Herder (Eds.), Adaptive hypermedia and adaptive web-based systems. LNCS vol. 5149 (pp. 113-122). Berlin: Springer.

[31] Gaffney, C., Dagger, D., & Wade, V. (2010). Authoring and delivering personalised simulations – an innovative approach to adaptive eLearning for soft skills. Journal of Universal Computer Science, 16, 2780-2800.

[32] Lloyd, W.J. (n.d.) An evaluation of software repository effectivenss: A framework based approach. Retrieved September 18, 2015 from http://www.javaforge.com/displayDocument/ResearchExam.pdf?doc_id=56406

[33] Saracevic, T. (2000). Digital library evaluation: toward an evolution of concepts. Library Trends, 49, 350-369.

[34] Steiner, C.M., Agosti, M., Sweetnam, M.S., Hillemann, E.-C., Orio, N., Ponchia, C. et al. (2014). Evaluating a digital humanities research environment: the CULTURA approach. International Journal on Digital Libraries, 15, 53-70.

[35] Zuccala, A., Oppenheim, C., & Dhiensa, R. (2008). Managing and evaluating digital repositories. Information Research, 13, paper 333.

[36] Marill, J.L. & Luczak, E.C. (2009). Evaluation of digital repository software at the national library of medicine. D-Lib Magazine, 15.

[37] Van Velsen, L., Van der Geest, T., Klaasen, R., & Steehouder, M. (2008). User-centered evaluation of adaptive and adaptable systems: a literature review. The Knowledge Engineering Review, 23, 261-281.

[38] Danielsson, K., & Wiberg, C. (2006). Participatory design of learning media: Designing educational computer games with and for teenagers. Interactive Technology & Smart Education, 4, 275-292.

[39] Davis, F.D. & Venkatesh, V. (2004). Toward preprototype user acceptance testing of new information systems: Implications for software project management. IEEE Transactions on Engineering Management, 51, 31-46.

[40] Serrano-Laguna, A., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B. (2014). Application of learning analytics in educational videogames. Entertainment Computing, 5, 313-322.

[41] Sewell, M., & Marczuk, M. (2004). Using Cost Analysis in Evaluation. Tucson, AZ: University of Arizona, College of Agriculture and Life Sciences, Cooperative Extension.

[42] Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2001). Cost-benefit analysis: Concepts and practice (2 ed.). Upper Saddle River: Prentice Hall.