MEEGA+: A Method for the Evaluation of the Quality of Games for Computing Education

Giani Petri *(Autor)*
Colégio Politécnico da UFSM
Universidade Federal de Santa Maria (UFSM)
Santa Maria – RS, Brasil
gpetri@inf.ufsm.br

Christiane Gresse von Wangenheim *(Orientadora)*
Departamento de Informática e Estatística (INE)
Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC, Brasil
c.wangenheim@ufsc.br

***Abstract*—Games are assumed to be an effective and efficient instructional strategy for computing education. However, it is essential to systematically evaluate such games to obtain sound evidence of their quality. In this regard, the objective of this research is to develop and evaluate an evaluation method (MEEGA+) providing a comprehensive support for evaluations of the quality of games used as an instructional strategy for computing education. The MEEGA+ method has been systematically developed and evaluated following a rigorous research method. Results of a reliability and validity analysis based on data collected from 62 case studies involving a population of 1048 computing students, as well as based on the perspective of 19 experts in educational games, indicate that the MEEGA+ is a valid and reliable method and provides a systematic support for quality evaluations of games. It can be used by game creators in the industry and/or academy, instructors, and researchers to evaluate the quality of games as a basis for their improvement and effective and efficient adoption in practice for computing education.**

Keywords: educational game; evaluation; method; computing education;

I. INTRODUCTION

In the last years, games have also been used more and more in educational contexts [1, 2, 3]. Educational games are specifically designed to teach people about a certain subject, expand concepts, reinforce development, or assist learners in learning a skill or change an attitude [1]. Especially in computing education, games are believed to result in a wide range of benefits, like increasing learning effectiveness, increasing interest and motivation as well as a reduction of teaching time and instructor load. However, in practice, the expected benefits of this kind of games are still questionable due to a lack of studies providing evidence of these benefits [2, 4, 5, 6]. This, consequently, may impair the effective and efficient adoption of games as an instructional strategy as well as misguide their development and/or improvement. Therefore, it is essential to systematically evaluate such games in order to obtain sound evidence of their quality.

As result of systematic literature reviews analysing the state of the art and the practice (thesis chapter 3) [6, 7], we identified that a reason for the lack of studies providing evidence of games' benefits may be that most games used for computing education are evaluated without explicitly defining an evaluation objective, research design, measurement, data collection instruments, and/or data analysis methods. This lack of scientific rigor leaves the reliability and validity of their results and, thus, the quality and/or effectiveness of such games questionable.

Another reason may be that there are only few approaches that provide a support for game evaluations. Existing approaches to game evaluations, e.g., MEEGA (Model for the Evaluation of Educational Games) [8] and the EGameFlow scale [9], do not provide a comprehensive support to guide game evaluations. Currently, the MEEGA model is widely used for game evaluations in practice [4, 6]. Yet, although initially an acceptable reliability has been identified, a more comprehensive analysis of the MEEGA model based on a series of case studies with 723 students indicated some improvement opportunities regarding its validity [10]. These improvement opportunities are related to an overlap of theoretical concepts of its quality factors motivation and user experience, as well as a lack of understanding of the wording of some questionnaire items [10]. Thus, evaluations using the MEEGA model may lead to imprecise results on the game's quality, not correctly identifying evidence of their benefits regarding the overlapped concepts. This, consequently, may impair the effective and efficient adoption of games as an instructional strategy for computing education. In addition, the existing approaches, including MEEGA, do not provide a more comprehensive support, for example, defining a process step by step in order to guide researchers in the planning, execution and analysis of game evaluations. Therefore, there is a lack of a valid and reliable method that provides a systematic support for the definition, execution and analysis of quality evaluations of games used for computing education.

Thus, the question focused in this research is: how to systematically conduct an evaluation of the quality of educational games used as an instructional strategy for computing education? In order to answer this question, the main contribution of this research is the design and evaluation of a new method (MEEGA+) improving the initial version of the MEEGA model and providing a comprehensive support for the evaluation of the quality of games.

II. RESEARCH METHOD

To develop the MEEGA+ method, a multi-method research (detailed in thesis section 1.4) is adopted, following these steps:

Step 1. Identify the state-of-the-art & practice. To identify the state-of-the-art and practice we performed two systematic literature reviews (thesis chapter 3). The state-of-the-art aims at identifying existing approaches (methods, models, frameworks, scales) to systematically evaluate educational games [6]. And, the state-of-the-practice, aims at identifying how games used for computing education are evaluated [7].

Step 2. Conduct a large-scale evaluation of the initial version of a prominent evaluation approach. To better understand shortcomings of existing approaches, we conduct a large-scale evaluation of the MEEGA model, currently the most used one in practice in this context (thesis chapter 4). Following the case study approach as proposed by Yin [11], the study objective has been defined and decomposed into quality aspects and analysis questions. In the execution phase, we collected the data from selected studies that evaluated educational games using the initial version of a prominent evaluation approach, then we pooled the data collected in a single sample for data analysis. In the analysis phase, the data collected were analysed in order to identify the reliability and construct validity of the prominent evaluation approach.

The design of the MEEGA+ method is organized into two steps (Steps 3 and 4) (thesis chapter 5):

Step 3. Design the MEEGA+ model. The MEEGA+ model has been developed, as an evolution of the initial version of the MEEGA model [8] identified as a prominent evaluation approach, widely used in practice for game evaluations [6, 7] and the large-scale analysis of the initial version of the MEEGA model [10]. The design of the MEEGA+ model follows the procedure of the scale development guide proposed by DeVellis [13], defining an evaluation model, a standardized measurement instrument, the data analysis process, and the game quality scale.

Step 4. Design the MEEGA+ process. The MEEGA+ process aims to provide a systematic support, guiding researchers and instructors, in the conduction of game evaluations adopting the MEEGA+ model. The process is modelled in a prescriptive way following the approach proposed by Acuña and Ferré [12], defining how the process should be performed, establishing phases, activities, and work products.

Step 5. Apply and evaluate the MEEGA+ method. The evaluation of the MEEGA+ method has been performed in two steps. The first step aims to evaluate the reliability and validity of the MEEGA+ model. In this step, a series of case studies [11] are conducted, data are collected, grouped and analysed in order to identify the reliability and validity of the MEEGA+ (thesis chapter 6.1).

The second step aims to evaluate the quality of the MEEGA+ method from the experts' perspective in terms of authenticity, validity, usability, correctness, completeness, consistency, understandability, unambiguousness, and flexibility, through an expert panel (thesis chapter 6.2). The evaluation is defined and decomposed into analysis questions and metrics, which are collected through a questionnaire answered by experts in educational games after analysing the MEEGA+ method.

III. THE MEEGA+ METHOD

The MEEGA+ method (Figure 1) aims to provide a systematic support for the evaluation of the quality of games for computing education. It is composed of an evaluation model (MEEGA+ Model) defining quality factors to be evaluated through a standardized measurement instrument, a scale, which classifies the evaluated game according to its quality level, and a process (MEEGA+ Process), defining phases, activities and work products, describing how to plan, execute and analyse the results of game evaluations.

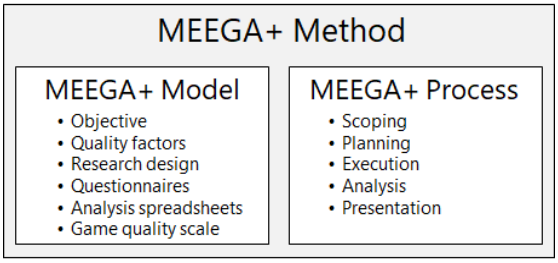

Figure 1. The MEEGA+ method

A. MEEGA+ Model

The MEEGA+ model aims to evaluate the quality of educational games in terms of usability and player experience from the students' perspective [14]. Based on the results of the literature reviews [6, 7] and results of a preliminary statistical analysis of the MEEGA+ model (exploratory factor analysis), analysing a sample of 718 students from 40 case studies [14], the MEEGA+ model is decomposed into two quality factors (usability and player experience) and their dimensions (Figure 2). We define usability as the degree to which a product (educational game) can be used by specified users (students) to achieve specified goals with effectiveness and efficiency in a specified context of use (computing education), being composed of the following dimensions: aesthetics, learnability, operability, and accessibility. Player experience is a quality factor that covers a deep involvement of the student in the gaming task, including his/her perception of learning, feelings, pleasures, and interactions with the game, environment and other players, being composed of the following dimensions: focused attention, fun, challenge, social interaction, confidence, relevance, satisfaction, and perceived learning.

Research Design. To conduct the game evaluation in a quickly and non-intrusive way, not interrupting the normal flow of the class and to not impair the participants involved in the study, a case study design is chosen for the evaluation that allows an in-depth research of an individual, group or event. Thus, the evaluation is conducted as a one-shot post-test only, in which the case study begins with the application of the treatment (educational game) and then a measurement instrument is answered by the students (self-assessment) in order to collect data on their perceptions about the game.

The MEEGA+ measurement instrument. Data collection is operationalized through a measurement instrument (questionnaire). The measurement instrument items are derived based on the defined quality factors/dimensions (Figure 2), customizing and unifying existing standardized questionnaires found in literature. The MEEGA+ measurement instrument is composed of 31 items that systematically measure the defined quality factors/dimensions. As response format, we adopt a 5-point Likert scale with response alternatives ranging from strongly disagree to strongly agree. A complete description of the questionnaire is available in thesis section 5.1.

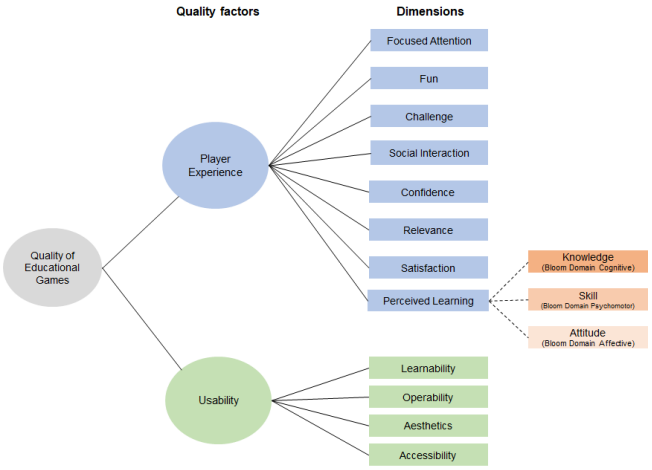

Figure 2. The MEEGA+ model

Data Analysis. Data collected in the evaluations are analysed in terms of frequency distribution (through frequency graphs) and central tendency (median) for each quality factor (usability and player experience) and their dimensions. The MEEGA+ model provides a spreadsheet for the analysis of the data collected, assisting in the organization of the information and automatic generation of graphs, visualizing the results of the evaluation.

Game quality scale. The scale aims to classify the evaluated game on a quality level. The scale has been developed by adopting the Item Response Theory (IRT) (thesis section 5.1.1), which allows to express through mathematical models the relationship between observable variables (questionnaire items) and latent traits (game quality). Based on the results of the analysis, three levels of quality are defined to classify the evaluated game: low, good, and excellent quality.

B. *MEEGA+ Process*

To guide the application of the model, the MEEGA+ method contains a systematic process (thesis section 5.2). The process specifies steps, activities and work products, guiding researchers and instructors in the conduction of game evaluations. The MEEGA+ process is organized into five phases: scoping, planning, execution, analysis, and presentation, as presented in Figure 3.

The complete material of the MEEGA+ method is available in English, Brazilian Portuguese and Spanish at: http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/ under the Creative Commons License.

IV. Evaluation of the MEEGA+ Method

To evaluate the MEEGA+ method in terms of reliability and validity, 62 case studies were conducted at eight educational institutions in Brazil and Spain, involving a sample of 1048 computing students. Results of the reliability analysis indicate an excellent internal consistency of the MEEGA+ measurement instrument (Cronbach's alpha α=.927). Results of the validity analysis, through an exploratory factor analysis and items correlation, confirm the original structure of the MEEGA+ model, indicating that the quality of games for computing education is evaluated in terms of usability and player experience. Moreover, results of a comprehensive evaluation of the MEEGA+ method, based on the perspective of 19 experts in educational games, indicate that the MEEGA+ method is also correct, authentic, consistent, and unambiguous. Thus, the results of our research indicate that the MEEGA+ is a valid and reliable method and provides a systematic support for evaluations of the quality of games for computing education. A detailed description of the application and evaluation of the MEEGA+ method can be found in thesis chapter 6.

V. Scientific Publications

During the development of this research, several articles have been published presenting partial results as journal, conference papers, and book chapters. Some of them are presented in Table I. Two articles present the first results of our research through systematic literature reviews, presenting the state of the art (Id 4) and the practice (Id 1).

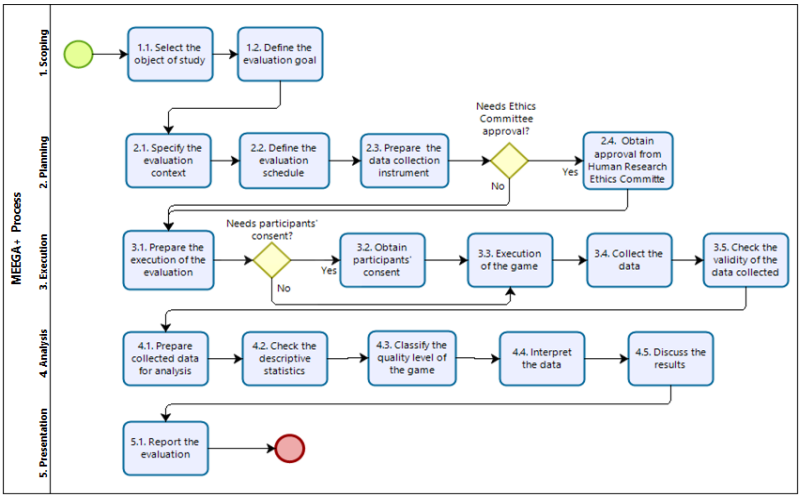

Figure 3. The MEEGA+ process

A conference article (Id 2) presents the results of a large-scale evaluation of the MEEGA model, identified through the literature review as a prominent evaluation approach. During the design of the MEEGA+ method, partial results were published in conferences and books (Id 7, Id 9, Id 12). In addition, several articles have been published presenting the results obtained from the case studies conducted to evaluate the MEEGA+ method (Id 3, Id 5, Id 6, Id 8, Id 10, Id 11, Id 13, Id 14).

TABLE I. PUBLICATIONS

Id	Reference	Qualis Computer Science
1	PETRI, G., & GRESSE VON WANGENHEIM, C. (2017). How games for computing education are evaluated? A systematic literature review, **Computers & Education**, 107, pp. 68-90. DOI: https://doi.org/10.1016/j.compedu.2017.01.004	A1
2	PETRI, G., GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. (2017). A Large-scale Evaluation of a Model for the Evaluation of Games for Teaching Software Engineering. In Proc. of the **39th International Conference on Software Engineering**: Software Engineering Education and Training Track (pp. 180-189). Buenos Aires/Argentina. DOI: https://doi.org/10.1109/ICSE-SEET.2017.11	A1
3	PETRI, G.; GRESSE VON WANGENHEIM, C.; BORGATTO, A. F. Quality of Games for Teaching Software Engineering: An Analysis of Empirical Evidence of Digital and Non-digital Games. In Proc. of the **39th International Conference on Software Engineering**: Software Engineering Education and Training Track (pp. 150-159). Buenos Aires/Argentina. DOI: https://doi.org/10.1109/ICSE-SEET.2017.18	A1
4	PETRI, G., & GRESSE VON WANGENHEIM, C. (2016). How to evaluate educational games: a systematic literature review. **Journal of Universal Computer Science**, 22(7), pp. 992-1021. DOI: http://dx.doi.org/10.3217/jucs-022-07-0992	B1
5	PETRI, G., CALDERÓN, A, GRESSE VON WANGENHEIM, C., BORGATTO, A. F., & RUIZ, M. (2018). Games for Teaching Software Project Management: An Analysis of the Benefits of Digital and Non-Digital Games. **Journal of Universal Computer Science**, 24(10), 1424,1451. Available at <http://www.jucs.org/jucs_24_10/games_for_teaching_software>	B1
6	PETRI, G., BATTISTELLA, P., CASSETTARI, F., GRESSE VON WANGENHEIM, C., & HAUCK, J. (2016). Um Quiz Game para a revisão de conhecimento em Gerenciamento de Projetos. In Proc. of the **27° Simpósio Brasileiro de Informática na Educação** (SBIE) (pp. 320-329). Uberlândia/MG. DOI: http://dx.doi.org/10.5753/cbie.sbie.2016.320	B1
7	SOARES, R., PETRI, G., GRESSE VON WANGENHEIM, C., CONTE, T., & MARQUES, A. B. (2018). AssistantMEEGA+: Uma ferramenta de apoio para avaliação de jogos educacionais usando modelo MEEGA+. In Proc. of the **29° Simpósio Brasileiro de Informática na Educação** (SBIE), Fortaleza/CE. DOI: http://dx.doi.org/10.5753/cbie.sbie.2018.615	B1
8	BATTISTELLA, P. E., PETRI, G., GRESSE VON WANGENHEIM, C., VON WANGENHEIM, A., & MARTINA, J. E. (2016). SORTIA 2.0: Um jogo de ordenação para o ensino de Estrutura de Dados. In Proc. of the **12º Simpósio Brasileiro de Sistemas de Informação** (SBSI) (558-565). Florianópolis/SC. Available at http://www.lbd.dcc.ufmg.br/colecoes/sbsi/2016/073.pdf	B2
9	PETRI, G., GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. (2017). Evolução de um Modelo de Avaliação de Jogos para o Ensino de Computação. In Proc. of the **25° Workshop sobre Educação em Computação** (CSBC/WEI) (pp. 2327-2336). São Paulo/SP. Available at http://csbc2017.mackenzie.br/public/files/25-wei/25.pdf	B3
10	PETRI, G., CALDERÓN, A, GRESSE VON WANGENHEIM, C., BORGATTO, A. F., & RUIZ, M. (2018). Benefícios dos Jogos Não-Digitais no Ensino de Computação. In Proc. of the **26° Workshop sobre Educação em Computação** (CSBC/WEI). Natal/RN. Available at http://portaldeconteudo.sbc.org.br/index.php/wei/article/view/3481	B3
11	PETRI, G., GRESSE VON WANGENHEIM, C., BONIATI, B., & WEBER, A. (2018). Avaliação de uma Dinâmica Vivencial para o Ensino de Gerenciamento de Projetos em Cursos de Computação. In Proc. of the **26° Workshop sobre Educação em Computação** (CSBC/WEI). Natal/RN. Available at http://portaldeconteudo.sbc.org.br/index.php/wei/article/view/3480	B3 (honors mention)
12	PETRI, G., GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. (2018). MEEGA+, Systematic Model to Evaluate Educational Games. In: Lee N. (Eds.) **Encyclopedia of Computer Graphics and Games** (pp. 1-7). Springer. DOI: https://doi.org/10.1007/978-3-319-08234-9_214-1	Book chapter
13	PETRI, G., GRESSE VON WANGENHEIM, C., BORGATTO, A. F., CALDERÓN, A., & RUIZ, M. (2018). Digital Games for Computing Education: What are the Benefits? In: Krassmann et al. (Eds.) **Handbook of Research on Immersive Digital Games in Educational Environments**. IGI Global. Chap 2. DOI: https://doi.org/10.4018/978-1-5225-5790-6.ch002	Book chapter
14	PETRI, G., GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. (2018). Qualidade de jogos digitais e não digitais utilizados para o ensino de engenharia de software no Brasil. **Revista de Gestão e Avaliação Educacional**, 7 (14), pp. 9-29. DOI: http://dx.doi.org/10.5902/2318133829046	-

VI. CONCLUSIONS

As result of this research we develop and evaluate the MEEGA+ method, a well-defined, valid and reliable method, to provide a systematic and comprehensive support for evaluations of the quality of games used for computing education. Thus, answering our research question, indicating that one way to systematically evaluate games for computing education is adopting the MEEGA+ method. The list of publications and the wide adoption of the MEEGA+ method by the community (as mentioned in thesis chapter 8), show that this thesis has a strongly contribution to the state of the art, providing a systematic and comprehensive support for quality evaluations of games used for computing education and, thus, contributes to an effective and efficient teaching of computing in a fun way keeping students engaged and motivated through the use of games.

REFERENCES

[1] C. C. Abt. “Serious Games”, 2002, Lanhan, MD: University Press of America.

[2] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle. “A systematic literature review of empirical evidence on computer games and serious games”, Computers & Education, vol. 59, no. 2, 2012, pp. 661-686.

[3] P. Battistella and C. Gresse von Wangenheim. “Games for teaching computing in higher education – a systematic review”, IEEE Technology and Engineering Education Journal, vol. 9, no. 1, 2016, pp. 8-30.

[4] A. Calderón and M. Ruiz. “A systematic literature review on serious games evaluation: an application to software project management”, Computers & Education, vol. 87, Semptember 2015, pp. 396-422.

[5] M. Kosa, et al. “Software engineering education and games: a systematic literature review”, Journal of Universal Computer Science, vol. 22 n. 12, 2016, pp. 1558-1574.

[6] G. Petri and C. Gresse von Wangenheim. “How games for computing education are evaluated? A systematic literature review”, Computers & Education, vol. 107, 2017, pp. 68-90.

[7] G. Petri and C. Gresse von Wangenheim. “How to evaluate educational games: a systematic literature review”, Journal of Universal Computers Science, vol. 22, no. 7, 2016, pp. 992-1021.

[8] R. Savi, C. Gresse von Wangenheim, and A. F. Borgatto. “A model for the evaluation of educational games for teaching software engineering”, Proc. of the 25th Brazilian Symposium on Software Engineering, 2011, pp. 194-203. São Paulo/SP, Brazil.

[9] F. Fu, R. Su, and S. Yu. “EGameFlow: A scale to measure learners' enjoyment of e-learning games”, Computers & Education, vol. 52, no. 1, 2009, pp. 101-112.

[10] G. Petri, C. Gresse von Wangenheim, and A. F. Borgatto. “A Large-scale Evaluation of a Model for the Evaluation of Games for Teaching Software Engineering”, Proc. of the 39th International Conference on Software Engineering, 2017, pp. 180-189, Buenos Aires/Argentina.

[11] R. K. Yin. “Case study research: design and methods”, Thousand Oaks: Sage Publications, Inc, 2017.

[12] S. T. Acuña and X. Ferré, “Software Process Modelling”, Proc. of the World Multiconference On Systemics, Cybernetics and Informatics, 2001, pp. 1-6. Orlando, USA.

[13] R. F. DeVellis. “Scale development: theory and applications”, SAGE Publications, 2016.

[14] G. Petri, C. Gresse von Wangenheim, and A. F. Borgatto, “MEEGA+, Systematic Model to Evaluate Educational Games”, In Newton Lee (Eds) Encyclopedia of Computer Graphics and Games, pp. 1-7. Cham: Springer.