Giani Petri

A METHOD FOR THE EVALUATION OF THE QUALITY OF GAMES FOR COMPUTING EDUCATION

Tese submetida ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Ciência da Computação.
Orientadora: Prof.ª Dr.ª rer. nat. Christiane Gresse von Wangenheim, PMP.

Florianópolis
2018

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Petri, Giani
A method for the evaluation of the quality of games for computing education / Giani Petri ; orientadora, Christiane Gresse von Wangenheim, 2018.
335 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós Graduação em Ciência da Computação, Florianópolis, 2018.

Inclui referências.

1. Ciência da Computação. 2. Jogo educacional. 3. Avaliação. 4. Método. 5. Ensino de Computação. I. Gresse von Wangenheim, Christiane . II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciência da Computação. III. Título.

Giani Petri

A METHOD FOR THE EVALUATION OF THE QUALITY OF GAMES FOR COMPUTING EDUCATION

Esta tese foi julgada adequada para obtenção do título de doutor e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Florianópolis, 29 de novembro de 2018.

Prof. José Luiz Almada Güntzel, Dr.
Coordenador do Programa

Banca Examinadora:

Prof.ª Christiane Gresse von Wangenheim, Dr.ª
Orientadora
Universidade Federal de Santa Catarina

Prof.ª Tayana Uchôa Conte, Dr.ª
Universidade Federal do Amazonas
(videoconferência)

Prof.ª Giliane Bernardi, Dr.ª
Universidade Federal de Santa Maria
(videoconferência)

Prof. Ricardo Azambuja Silveira, Dr.
Universidade Federal de Santa Catarina

To my lovely family.

ACKNOWLEDGEMENTS

At the end of this four-year journey, I could summarize it with just one word: gratitude.

Gratitude to God, for the life, for my family, for the professional opportunities and, especially, for the force in difficult times during the accomplishment of this work.

Gratitude to my lovely family, especially, to my parents and my sister. For the support, love, and mainly, for understanding when I was absent.

Gratitude to my fiancée Caroline, for the affection, love and understanding. Thank you for listening and motivating me every day in the pursuit of my dreams. You are also part of this achievement!

I am also profoundly grateful to Prof. Christiane, for the orientations, revisions, opportunities and all suggestions that contributed not only to the quality of this research, but also in my professional and personal development. You are my ‘gold standard’ of professor and researcher!

I am also grateful to Prof. Adriano, for making me understand (a little) the universe of statistics.

I would like to thank the colleagues of the GQS research group, especially to Paulo, Rafael, Prof. Jean and Nathalie, for all support in daily activities, for the learning, shared knowledge, and coffee breaks.

Thanks also to the researchers of the SPI&FM research group of the University of Cádiz/Spain, for receiving me so well and giving me all support during my stay in Spain, especially, to Prof. Mercedes, Alejandro, Tatiana and Edson. It was an amazing experience!

I thank to the CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) (grant n. 88881.131485/2016-01), for the financial support during my research stay in Spain.

I would like to thank to the “serious boys”, for the friendship and all support during this period.

I also thank the Federal University of Santa Maria, for the opportunity to devote myself fully to research during this period.

Finally, I would like to thank everyone who contributed to this research, especially to the students who participated in the game evaluations and the researchers who participated in the final evaluation.

I am profoundly grateful!

For every disciplined effort there are a multiple reward.

– Jim Rohn

RESUMO

Jogos educacionais são considerados uma estratégia instrucional eficiente e efetiva em diversas áreas do conhecimento, inclusive na computação. Especialmente para o ensino de engenharia de software, jogos têm sido utilizados com o objetivo de fornecer mais oportunidades de aprendizagem prática, contribuindo para a aprendizagem, motivação, engajamento, etc. dos estudantes de computação. No entanto, para confirmar esses benefícios esperados é importante avaliar sistematicamente esses jogos. Porém, atualmente, há poucas abordagens, que forneçam um suporte sistemático para a avaliação de jogos. Entre as poucas abordagens existentes, um modelo de avaliação proeminente parece ser o MEEGA (*Model for the Evaluation of Educational Games*), que avalia jogos em termos de motivação, experiência do usuário e aprendizagem. No entanto, resultados de uma avaliação em larga escala do modelo MEEGA, mostram limitações em termos de sua validade, indicando uma sobreposição conceitual entre os fatores de motivação e experiência do usuário. Além disso, o modelo MEEGA não define um processo de avaliação de jogos. Nesse contexto, há uma ausência de abordagens válidas e confiáveis, e que forneçam um suporte sistemático, para a definição, execução e análise de dados de avaliações de jogos educacionais. Diante disso, o objetivo desta pesquisa é desenvolver e avaliar um método de avaliação, MEEGA+, que forneça um suporte abrangente para a avaliação da qualidade de jogos usados como estratégia instrucional para o ensino de computação/engenharia de software. Ele é composto por um modelo de avaliação que define fatores de qualidade para serem avaliados por meio de um instrumento de medição padronizado; uma escala, que classifica o jogo avaliado em níveis de qualidade; e um processo de avaliação de jogos. 62 estudos de caso foram conduzidos de modo a avaliar a confiabilidade e validade do instrumento de medição do método MEEGA+, envolvendo uma amostra de 1048 estudantes. Os resultados da análise da confiabilidade indicam uma excelente consistência interna do instrumento de medição do método MEEGA+ (Alfa de Cronbach α=.927). Além disso, os resultados da análise da validade confirmam a estrutura original do método MEEGA+, indicando que a qualidade de jogos para o ensino de computação/engenharia de software é avaliada em termos de usabilidade e experiência do jogador. Adicionalmente, resultados de uma avaliação sob a perspectiva de especialistas, indicam que o método MEEGA+ também é correto, autêntico, consistente e não ambíguo. Portanto, os resultados desta pesquisa indicam que o MEEGA+ é um método válido e confiável e que fornece um suporte sistemático para a avaliação da qualidade de jogos. Assim, o método MEEGA+ pode ser usado por desenvolvedores de jogos, instrutores e pesquisadores de modo a avaliar a qualidade de jogos como base para identificar melhorias e/ou adotá-los na prática para o ensino de computação/engenharia de software.

Palavras-chave: Jogo educacional. Avaliação. Método. Ensino de Engenharia de Software. Ensino de Computação.

RESUMO EXPANDIDO

Introdução

Nos últimos anos, os jogos educacionais têm sido utilizados como uma estratégia instrucional inovadora para o ensino de computação. Esses jogos educacionais são projetados especificamente para ensinar as pessoas sobre um determinado assunto, expandir conceitos, ou ajudá-las a explorar ou aprender uma habilidade ou uma mudança de atitude. No ensino de computação, principalmente na área de engenharia de software, os jogos educacionais são usados tipicamente para fornecer mais oportunidades práticas de aprendizagem para os estudantes por meio de um ambiente seguro e controlado. Assim, acredita-se que os jogos educacionais possam ser uma estratégia instrucional eficaz e eficiente para o ensino de computação, contribuindo para a aprendizagem, motivação, engajamento, etc. dos estudantes. No entanto, para confirmar esses benefícios esperados é importante avaliar sistematicamente esses jogos. Porém, atualmente, há poucas abordagens, que forneçam um suporte sistemático para a avaliação de jogos. Entre as abordagens para a avaliação de jogos encontradas na literatura, um modelo proeminente e amplamente utilizado na prática para avaliação de jogos parece ser o MEEGA (*Model for the Evaluation of Educational Games*). MEEGA é um modelo desenvolvido para a avaliação de jogos educacionais, que mede a qualidade do jogo em termos de motivação, experiência do usuário e aprendizagem por meio de um questionário. No entanto, resultados de uma avaliação em larga escala do modelo MEEGA, mostram limitações em termos de sua validade, indicando uma sobreposição conceitual entre os fatores de motivação e experiência do usuário. Além disso, o modelo MEEGA não fornece um suporte mais abrangente, por exemplo, definindo um passo a passo de modo a orientar pesquisadores no planejamento, execução, e análise de resultados de avaliações de jogos. Nesse contexto, há uma ausência de abordagens válidas e confiáveis e que forneçam um suporte sistemático para a definição, execução e análise de dados de avaliações de jogos educacionais usados para o ensino de computação/engenharia de software. Diante disso, a pergunta que guia a realização desta pesquisa consiste em: como realizar sistematicamente a avaliação da qualidade de jogos educacionais utilizados como estratégia instrucional para o ensino de computação/engenharia de software?

Objetivos

O objetivo geral desta pesquisa é desenvolver e avaliar um método (MEEGA+) para a avaliação da qualidade de jogos educacionais utilizados como estratégia instrucional para o ensino de engenharia de software. De modo a alcançá-lo, são definidos os seguintes objetivos específicos: identificar o estado da arte de quais abordagens existem para avaliar jogos educacionais; identificar o estado da prática de como jogos para o ensino de computação/engenharia de software são avaliados; reavaliar a confiabilidade e validade de uma abordagem de avaliação proeminente; evoluir a abordagem de avaliação com base no levantamento do estado da arte e prática e na análise de sua versão inicial; desenvolver um processo que forneça um suporte sistemático para a avaliação da qualidade de jogos para o ensino de engenharia de software; aplicar e avaliar o método de avaliação em diferentes disciplinas, cursos e instituições de ensino.

Metodologia

Os procedimentos metodológicos desta pesquisa são definidos com base na proposta de Saunders, Lewis e Thornhill (2009), que estruturam o processo da ciência, classificando o método científico em camadas. Em termos de filosofia, esta pesquisa é predominantemente interpretativista, pois assume-se que o objeto de pesquisa (jogos educacionais) é interpretado e avaliado do ponto de vista social, pelos atores envolvidos na pesquisa (alunos de computação). No aspecto da abordagem da pesquisa, é utilizada uma abordagem indutiva, pois não parte de uma hipótese pré-estabelecida, mas sim, procura atingir a solução do problema a partir das conclusões inferidas de estudos de casos particulares do objeto de estudo (jogos educacionais). Como estratégia de pesquisa, é classificada como multimétodo, pois são utilizados diferentes métodos qualitativos e quantitativos, tais como: pesquisa bibliográfica, mapeamentos sistemáticos da literatura, estudos de caso, GQM (*Goal/Question/Metric*), guia de desenvolvimento de escalas, guia para o design de questionários, métodos estatísticos, dentre outros. O horizonte de tempo desta pesquisa é classificado como transversal, pois o objeto de estudo (jogos educacionais) é analisado pelos atores (alunos) em eventos dissociados no tempo. Quanto à natureza, esta pesquisa é classificada como aplicada, pois objetiva gerar conhecimentos para a aplicação prática, contribuindo para a solução de problemas específicos que ocorrem na realidade. E, quanto aos objetivos, é classificada como exploratória, pois se interessa em proporcionar uma maior compreensão do fenômeno que é investigado.

Resultados e Discussão

A principal contribuição desta pesquisa é o desenvolvimento de um método (MEEGA+) para a avaliação da qualidade de jogos usados para o ensino de computação/engenharia de software. O método MEEGA+ foi sistematicamente desenvolvido com base nos resultados da análise do estado da arte e da prática e nos resultados da avaliação em larga escala da versão inicial do modelo MEEGA, identificado como a abordagem de avaliação proeminente e amplamente utilizada na prática. O método MEEGA+ é composto por um modelo de avaliação (Modelo MEEGA+), que sistematicamente define fatores de qualidade para avaliar jogos por meio de um instrumento de medição padronizado e fornece uma escala, que classifica o jogo avaliado em níveis de qualidade. Além disso o método MEEGA+ também define um processo (Processo MEEGA+), que descreve em detalhes as fases, atividades e produtos de trabalho para orientar instrutores e pesquisadores na condução de avaliações de jogos para o ensino de computação. Uma série de estudos de caso foi conduzida de modo a avaliar a confiabilidade e validade do instrumento de medição do método MEEGA+. Os resultados da análise da confiabilidade indicam uma excelente consistência interna do instrumento de medição do método MEEGA+ (Alfa de Cronbach α=.927). Além disso, os resultados da análise da validade, obtidos por meio de uma análise fatorial exploratória, confirmam a estrutura original do método MEEGA+, indicando que a qualidade de jogos para o ensino de computação/engenharia de software é avaliada em termos de usabilidade e experiência do jogador. Adicionalmente, resultados de uma avaliação do método MEEGA+, sob a perspectiva de especialistas, fornecem uma primeira indicação de que o método MEEGA+ também é correto, completo, autêntico, consistente e não ambíguo.

Considerações Finais

Observando uma lacuna no estado da arte em termos de abordagens sistemáticas e que forneçam um suporte abrangente para a avaliação da qualidade de jogos para o ensino de computação/engenharia de software, essa pesquisa objetiva desenvolver e avaliar o método de avaliação de jogos MEEGA+. O método MEEGA+ é composto por um modelo de avaliação que define fatores de qualidade para serem avaliados por meio de um instrumento de medição padronizado, uma escala, que classifica o jogo avaliado de acordo com seu nível de qualidade, e um processo, que define fases, atividades e produtos de trabalho, guiando pesquisadores em como planejar, executar e analisar os resultados de avaliações de jogos. O método MEEGA+ foi aplicado em 62 estudos de caso, envolvendo uma

população de 1048 estudantes. Com base nos dados coletados nos estudos de caso, o método foi avaliado em termos de validade, confiabilidade por meio de análise estatística. Além disso, o MEEGA+ também foi avaliado em termos de corretude, completude, consistência, validade, não ambiguidade, flexibilidade, compreensibilidade, usabilidade e autenticidade, por meio de um painel de especialistas. Com base nos resultados positivos obtidos em ambas as avaliações, obtém-se uma indicação de que o MEEGA+ é um método válido e confiável para a avaliação da qualidade de jogos e pode ser usado por desenvolvedores de jogos, instrutores e pesquisadores de modo a avaliar a qualidade de jogos como base para identificar melhorias e/ou adotá-los na prática para o ensino de computação/engenharia de software.

Palavras-chave: Jogo educacional. Avaliação. Método. Ensino de Engenharia de Software. Ensino de Computação.

ABSTRACT

Educational games are supposed to be an effective and efficient instructional strategy for teaching and learning in diverse knowledge areas, including computing. Especially in software engineering education, games have been used in order to provide more practical learning opportunities to computing students, contributing to students' learning, motivation, engagement, etc. Thus, in order to confirm these expected benefits, it is important to systematically evaluate such games. However, currently, there are only few approaches, which provide a systematic and comprehensive support for game evaluations. Among the few existing approaches, a prominent evaluation model seems to be MEEGA (Model for the Evaluation of Educational Games), which evaluates games in terms of motivation, user experience and learning. However, results of a large-scale analysis of the MEEGA model, have identified limitations regarding its validity, indicating an overlap of the theoretical concepts of motivation and user experience. In addition, the MEEGA model does not provide a process for game evaluations. In this context, there is a lack of a valid and reliable approach and that provides a systematic support in the definition, execution and data analysis from evaluations of educational games. In this regard, the objective of this research is to develop and evaluate an evaluation method (MEEGA+) providing a comprehensive support for quality evaluations of games used as an instructional strategy for computing/software engineering education. The MEEGA+ method has been systematically developed based on the results of systematic mapping studies. It is composed of an evaluation 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, defining phases, activities and work products, guiding researchers on how to plan, execute and analyse the results of game evaluations. In order to evaluate the MEEGA+ method in terms of reliability and validity of its measurement instrument, 62 case studies were conducted involving a sample of 1048 students. Results of the reliability analysis indicate an excellent internal consistency of the MEEGA+ measurement instrument (Cronbach's alpha α=.927). In addition, results of the validity analysis confirm the original structure of the MEEGA+ model, indicating that the quality of games for software engineering education is evaluated in terms of usability and player experience. Moreover, results of a comprehensive evaluation of the MEEGA+ method, based on the experts' perspective, 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 quality evaluation of games. It can be used by game creators, instructors and researchers in order to evaluate the quality of games as a basis for their improvement and effective and efficient adoption in practice for computing/software engineering education.

Keywords: Educational game. Evaluation. Method. Software Engineering education. Computing education.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ACM – Association for Computing Machinery
ADDIE – Analyse, Design, Develop, Implement, and Evaluate
AQS – Analysis questions
ARCS – Attention, Relevance, Confidence, Satisfaction
BPMN – Business Process Modelling Notation
CTF – Computational Thinking Framework
GBL – Games-based Learning
GQM – Goal Question Metric
GQS – Software Quality Group
IEEE – Institute of Electrical and Electronic Engineers
INCoD – Brazilian Institute for Digital Convergence
INE – Department of Informatics and Statistics
IRT – Item Response Theory
IT – Information Technology
IU – Instructional units
KMO – Kaiser-Meyer-Olkin index
MEC – Ministry of Education and Culture
MEEGA – Model for the Evaluation of Educational Games
MUMMS – Measuring the Usability of Multi-Media Systems
PC – Personal Computer
PPGCC – Graduate Program in Computer Science
RPG – Role-playing game
SBC – Brazilian Computer Society
SE – Software Engineering
SUS – System Usability Scale
TAM – Technology Acceptance Model
TUP – Technology, Usability and Pedagogy)
UDESC - Santa Catarina State University
UFAM – Federal University of Amazonas
UFSC – Federal University of Santa Catarina
UFSM – Federal University of Santa Maria
Uninorte – Centro Universitário do Norte
UNISUL – Universidade do Sul de Santa Catarina
USES – Usability and Software Engineering Research Group

SUMMARY

1 INTRODUCTION

1.1 CONTEXTUALIZATION

Computing has an important role in our society nowadays, strongly influencing the development of education, science, engineering, business, and many other knowledge areas. In this context, computing professionals will continue to play an essential role, changing and shaping our future (ACM; IEEE-CS, 2013). Therefore, it is important that computing education attracts quality students in order to prepare them to be qualified and responsible computing professionals for them to contribute to the development of diverse knowledge areas (PARSONS, 2011; ACM; IEEE-CS, 2013).

Students of higher computing education need to develop competencies including knowledge, skills and attitudes in computing areas as algorithms, data structures, software engineering, security, operating systems, among others (ACM; IEEE-CS, 2013). These competencies need to be acquired on several learning levels, from simply remembering concepts, classifying and structuring the information, up to being able to apply the knowledge in concrete situations in different knowledge areas (ACM; IEEE-CS, 2013).

Especially in Software Engineering (SE), students are not only expected to successfully cope with technical challenges, but also to deal with non-technical issues, including management, communication, teamwork, etc. SE is a discipline concerned with the application of theory, knowledge, and practice to effectively and efficiently build reliable software systems that satisfy the requirements of customers and users (ACM; IEEE-CS, 2013; BORQUE; FAIRLEY, 2014). However, one of the main challenges when teaching SE is to give students sufficient hands-on experience in building software (GRESSE VON WANGENHEIM; SHULL, 2009).

Although computing systems/technologies and the profile of computing students have expressively changed in the past decades, most of undergraduate computing programs are still taught in traditional ways (PARSONS, 2011; FREEMAN et al., 2014). Yet, in today's world, traditional instructional strategies may not be adequate for effective and efficient learning (PARSONS, 2011; FREEMAN et al., 2014; DOLAN; COLLINS, 2015). Therefore, more active instructional strategies are required, focusing on the students, allowing them to learn by doing and, thus, enabling more effective learning (FREEMAN et al., 2014; DOLAN; COLLINS, 2015).

In this context, educational games are expected to contribute to a deep and more active learning, through which students learn from their own experiences (CONNOLLY et al., 2012; GIBSON; BELL, 2013; BOYLE et al., 2016; KOSA et al., 2016; KORDAKI; GOUSIOU, 2016; CALDERÓN; RUIZ; O'CONNOR, 2018). 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 (ABT, 2002; PRENSKY, 2007; RITTERFELD; CODY; VORDERER, 2010; DJAOUTI et al., 2011). With this objective, educational games have been used as an innovative instructional strategy for computing education (BACKLUND; HENDRIX, 2013; BATTISTELLA; GRESSE VON WANGENHEIM, 2016). Driven by the need to provide more hands-on opportunities for computing students, various educational games have been used to contribute to the learning process in different computing knowledge areas. For example, educational games used for teaching Software Engineering (e.g., SimSE (NAVARRO; VAN DER HOEK, 2007), Requirements Collection and Analysis Game (RCAG) (HAINEY et al., 2011)), Software Project Management (e.g., SCRUMIA (GRESSE VON WANGENHEIM; SAVI; BORGATTO, 2013), ProDec (CALDERÓN; RUIZ, 2013)), or Security (e.g., CyberCIEGE (RAMAN et al., 2014), CounterMeasures (JORDAN et al., 2011)), among other. Digital games represent about 64% of the games developed for computing education, mainly PC games (Personal Computer). However, there is also a considerable trend towards non-digital (or traditional) games (paper and pencil, board games, cards, etc.) (BATTISTELLA; GRESSE VON WANGENHEIM, 2016). Simulation games, which allow practice skills in a realistic environment while keeping the students involved, are prevalent. Additionally, some games also have as learning objectives at lower cognitive levels, often being used as a complementary instructional strategy to revise and reinforce previously taught knowledge (BATTISTELLA; GRESS VON WANGENHEIM, 2016).

Educational 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 (PRENSKY, 2007; GRESSE VON WANGENHEIM; SHULL, 2009; BACKLUND; HENDRIX, 2013). Games are expected to provide a fun and safe environment, where students can try alternatives and see the consequences, learning from their own mistakes and practical experiences (PFAHL; RUHE; KOVAL, 2001; BACKLUND; HENDRIX, 2013). They are supposed to be an effective and efficient instructional strategy for computing education. However, in practice, the expected benefits of this kind of games, including digital and non-digital ones, are still questionable due to a lack of studies providing evidence of these benefits (GIBSON; BELL, 2013; BOYLE et al., 2016; ALL; CASTELLAR; LOOY, 2016; KORDAKI; GOUSIOU, 2017; PETRI; GRESSE VON WANGENHEIM, 2017). Therefore, it is essential to systematically evaluate such games in order to obtain sound evidence of their quality.

1.2 PROBLEM

A reason for the lack of studies providing evidence of games' benefits may be that most of games used for teaching computing/software engineering are evaluated without explicitly defining an evaluation objective, research design, measurement program, data collection instruments, and data analysis methods (CALDERÓN; RUIZ, 2015; ALL; CASTELLAR; LOOY, 2016; BOYLE et al., 2016; KOSA et al., 2016, PETRI; GRESSE VON WANGENHEIM, 2017; KORDAKI; GOUSIOU, 2017). And, often, data are collected in the evaluations only in form of students' informal comments and/or through questionnaires developed in an ad-hoc manner. This lack of scientific rigor and the fact that most studies are conducted with a small sample size without replications leaves the reliability and validity of their results and, thus, the quality and/or effectiveness of such games questionable (CALDERÓN; RUIZ, 2015; ALL; CASTELLAR; LOOY, 2016; KOSA et al., 2016, PETRI; GRESSE VON WANGENHEIM, 2017; TAHIR; WANGMAR, 2017).

Another reason for the lack of scientific rigor of the evaluations of such games may be that, currently, there are only few approaches that provide a systematic support for game evaluations (CALDERÓN; RUIZ, 2015; ALL; CASTELLAR; LOOY, 2016; PETRI; GRESSE VON WANGENHEIM, 2016; KORDAKI; GOUSIOU, 2017; TAHIR; WANGMAR, 2017; SANTOS et al., 2018). Among the few existing approaches, MEEGA (Model for the Evaluation of Educational GAmes) (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011) and EGameFlow (FU; SU; YU, 2009) stand out. The EGameFlow scale (FU; SU; YU, 2009) aims to evaluate the level of enjoyment of games from the users' point of view through a standardized questionnaire. Although this scale has been developed in a systematic way and evaluated in terms of validity and reliability, its evaluation involves only data from four games applied by the authors themselves with a population of 166 students and, in addition, the scale seems to have been proposed by the authors and then discontinued, not improving its initial version based on the results of its evaluation. MEEGA (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011) is a model developed for the evaluation of educational games for computing/software engineering education in terms of motivation, user experience, and learning, by measuring the reaction of students after the game play through a standardized questionnaire. Currently, the MEEGA model is widely used for game evaluations in practice (CALDERÓN; RUIZ, 2015; PETRI; GRESSE VON WANGENHEIM, 2017; CALDERÓN; RUIZ; O'CONNOR, 2018). Yet, although initially an acceptable reliability has been demonstrated, a more comprehensive analysis of the MEEGA model based on a sample of 1000 responses indicated some improvement opportunities regarding its validity (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2016). These improvement opportunities are related to an overlap of theoretical concepts of the quality factors motivation and user experience, as well as a lack of understanding of the wording of some questionnaire items (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2016; PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017). 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 (motivation and user experience). This, consequently, may impair the effective and efficient adoption of games as an instructional strategy for computing education as well as misguide their development and/or improvement. In addition, the existing approaches to game evaluations, including the MEEGA, do not provide a more comprehensive support, for example, defining a step by step in order to guide researchers in the planning, execution and analysis of results of game evaluations.

In this context, there is a lack of a valid and reliable method that provides a systematic support in the definition, execution and analysis of data from quality evaluations of educational games used for computing/SE education. Thus, the question guiding the realization of this research consists of:

Research question: How to systematically conduct a quality evaluation of educational games used as an instructional strategy for computing/SE education?

1.3 OBJECTIVES

The main objective of this research is to develop and evaluate a method for quality evaluation of games used as an instructional strategy for computing/SE education. In order to achieve this main objective, the following specific objectives are defined:

O1. Identify the state-of-the-art of approaches used to evaluate educational games.

O2. Identify the state-of-the-practice on how games for computing/SE education are evaluated.

O3. Conduct a reanalysis in terms of reliability and validity of a prominent evaluation approach

O4. Evolve a prominent evaluation approach based on the results of the state-of-the-art and practice and the analysis of its initial version.

O5. Develop a process that provides a systematic support for the evaluation of games for computing/SE education.

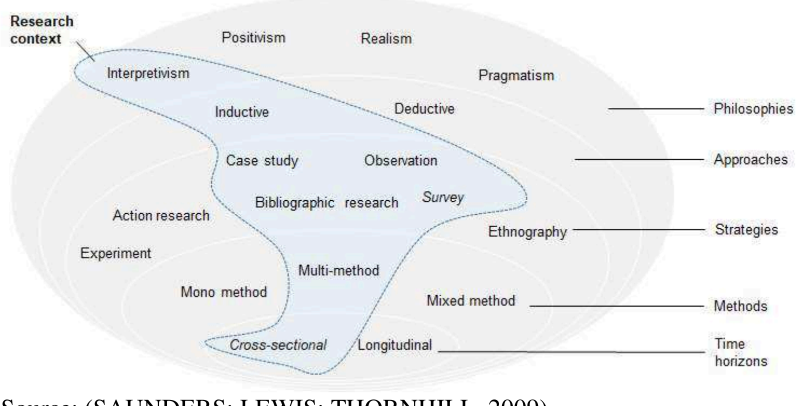
O6. Apply and evaluate the evaluation method in different computing courses and educational institutions.

1.4 RESEARCH METHOD

1.4.1 Research context

The methodological procedure of this research is defined based on Saunders, Lewis and Thornhill (2009), which structure the process of science, classifying the scientific method in layers (SAUNDERS; LEWIS; THORNHILL, 2009). Figure 1 presents the context of this research in the universe of the scientific method.

Figure 1 - Research context



Source: (SAUNDERS; LEWIS; THORNHILL, 2009).

Regarding the philosophy, this research is predominantly interpretivist, due to the fact that the object of study (educational games) is interpreted and evaluated from the social perspective by the actors involved in the research (computing students) (SAUNDERS; LEWIS; THORNHILL, 2009). In terms of approaches, this research adopts an inductive approach, which seeks to reach the solution of the problem from the inferences from particular case studies of the object of study (educational games). As research strategy, it is classified as a multi-method, adopting different qualitative and quantitative methods such as bibliographic research, systematic mapping studies (PETERSEN et al., 2008; KITCHENHAM, 2010), case studies (YIN, 2017), GQM (*Goal/Question/Metric*) (BASILI; CALDIERA; ROMBACH, 1994); scale development guide (DEVELLIS, 2016), guide for questionnaire design (KASUNIC, 2005), statistical methods (TROCHIM; DONNELLY, 2008), among others. The time horizon is classified as cross-sectional, because the object of study (educational games) is analysed by the actors (students) in events dissociated in time.

This research is also characterized, in terms of nature, as an applied research, because it aims to generate knowledge for practical application, contributing to the solution of specific problems that occur in reality (MARCONI; LAKATOS, 2010). In terms of objectives, it is classified as exploratory, because it is interested in providing an understanding of the phenomenon that is investigated.

As a result, this research proposes a new method (MEEGA+) for the evaluation of the quality of games for teaching computing/SE. In this study, we understand a method as a systematic approach to achieve a certain objective or result and, which describes the characteristics of an ordered process or a procedure used in the engineering of a product (IEEE, 2002; IEEE, 2010).

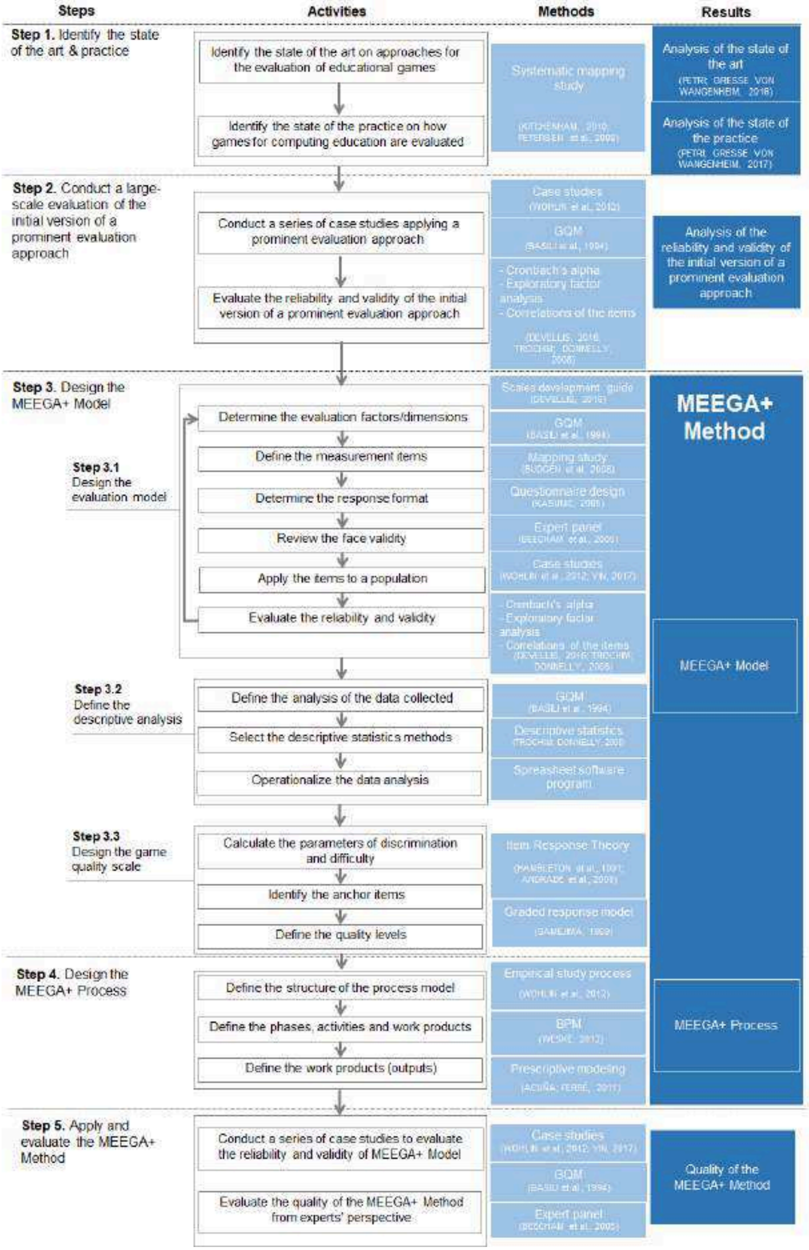
The MEEGA+ method is expected to contribute to the evaluation of the quality of games, evolving the initial version of a prominent evaluation approach in terms of reliability and validity and, in addition, providing a systematic support for researchers in the planning, execution and analysis of results of game evaluations.

In accordance with the classification of research, section 1.4.2 presents the research steps for the development and evaluation of the MEEGA+ method.

1.4.2 Research steps

In order to develop the MEEGA+ method, a multi-method research is adopted, following the steps shown in Figure 2.

Figure 2 - Research steps



Source: developed by the author.

Step 1. Identify the state-of-the-art & practice. In order to identify the state-of-the-art and practice on the evaluation of educational games for computing/SE education, we performed two Systematic Mapping Studies following the procedure defined by Petersen et al. (2008) and Kitchenham (2010). The analysis of the state-of-the-art aims at identifying existing approaches (methods, models, frameworks, scales) to systematically evaluate educational games. And, the analysis of the state-of-the-practice, aims at identifying how games used for computing/SE education are evaluated.

The systematic mapping process is divided into three phases: definition, execution and analysis. In the definition phase, research objectives are identified, and a systematic review protocol is defined. The protocol specifies the central research questions and the procedures that will be used to conduct the review, including the definition of inclusion/exclusion criteria, quality criteria, data sources, and search strings. The execution phase consists of the search and identification of relevant studies, and their selection in accordance with the inclusion/exclusion and quality criteria established in the protocol. Once identified, data related to the research question(s) are extracted from the relevant studies, analysed and synthesized during the analysis phase.

Step 2. Conduct a large-scale evaluation of the initial version of a prominent evaluation approach. In order to perform a large-scale evaluation of the initial version of a prominent evaluation approach, we conduct a case study (WOHLIN et al., 2012; YIN, 2017). The case study is divided into three phases: definition, execution and analysis. In the definition phase, the study objective is defined using the GQM goal template (BASILI; CALDIERA; ROMBACH, 1994). Following the GQM approach, the study objective is decomposed into quality aspects and analysis questions. The execution phase was organized in three steps. In step 1, we identified and selected potential studies that evaluated educational games (digital and/or non-digital) in different contexts using the initial version of a prominent evaluation approach. In step 2, we collected the data from the selected studies, contacting (via e-mail) the authors requesting the collected data. In step 3, we pooled the data collected in a single sample for data analysis. In the analysis phase, the data collected were analysed in order to answer our analysis questions. Data were analysed in terms of reliability and construct validity following the definition of Trochim and Donnelly (2008) and the scale development guide proposed by DeVellis (2016). In terms of reliability, we measured the internal consistency through the Cronbach's alpha coefficient (CRONBACH, 1951). Construct validity was analysed using an exploratory factor analysis and based on evidence of convergent and discriminant validity, obtained through the degree of correlations of the items (TROCHIM; DONNELLY, 2008; DEVELLIS, 2016). The results of the statistical analysis were interpreted by researchers in the context of computing/SE education in order to identify the reliability and validity of the measurement instrument of the evaluation approach.

Step 3. Design the MEEGA+ model.

Step 3.1 Design the evaluation model. The MEEGA+ model has been developed, as an evolution of the initial version of the MEEGA model (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011) identified as a prominent evaluation approach widely used in practice for game evaluations, based on the results of the literature reviews (PETRI; GRESSE VON WANGENHEIM, 2016; PETRI; GRESSE VON WANGENHEIM, 2017) and the large-scale analysis of the initial version of the MEEGA model (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017). The development of the MEEGA+ model follows the procedure of the scale development guide proposed by DeVellis (2016) and the guide for questionnaire design by Kasunic (2005):

Determine the evaluation factors/dimensions. Adopting the GQM (Goal/Question/Metric) approach (BASILI; CALDIERA; ROMBACH, 1994), the evaluation objective is defined and systematically decomposed into factors to be measured. The factors are defined in order to support the development of the measurement instrument (questionnaire), based on a mapping study of their concepts following the procedure proposed by Budgen et al. (2008).

Define the measurement items. The measurement of the factors is operationalized by decomposing the factors into measurement instrument items. For the definition of the items we also took into consideration other standardized questionnaires found in literature. We analysed the pool of items in terms of similarity and redundancy, customizing and unifying the selected items for the defined evaluation factors. In order to standardize the selected items, all items were refined and transformed into positive statements.

Determine the response format. The response format for the items of the measurement instrument is defined. This definition is based on response formats typically used in standardized questionnaires following the scale development guide proposed by DeVellis (2016).

Review the face validity. Face validity (TROCHIM; DONNELLY, 2008) is analysed by an expert panel (BEECHAM et al., 2005). The expert panel is composed of a multidisciplinary group of senior researchers with backgrounds in computing and/or statistics. The review aims at analysing clarity, relevance, consistency, and completeness of the measurement instrument items of the MEEGA+ model. The suggestions of the experts, including changes in the wording and text formatting, were considered in the development of the MEEGA+ measurement instrument.

Apply the items to a population. A series of case studies applying several games for computing/SE education is conducted. In each case study, after the game session (treatment), the MEEGA+ measurement instrument is used for data collection on the participants' perceptions about the game. We use a non-probability sampling technique in each case study applying the convenience sampling method (TROCHIM; DONNELLY, 2008), in which our sample is composed by students enrolled in computing courses.

Evaluate the reliability and validity. We pooled the data collected in each case study in a single sample for data analysis. Data were analysed in terms of reliability and construct validity following the definition of Trochim and Donnelly (2008) and the scale development guide proposed by DeVellis (2016). In terms of reliability, we measured the internal consistency through the Cronbach's alpha coefficient (CRONBACH, 1951). Construct validity was analysed using an exploratory factor analysis and based on evidence of convergent and discriminant validity, obtained through the degree of correlations of the items (TROCHIM; DONNELLY, 2008; DEVELLIS, 2016). The results of the statistical analysis were interpreted by researchers in the context of computing education in order to identify the reliability and validity of the MEEGA+ measurement instrument. Results of this step are used to identify the quality factors that represent the responses to the MEEGA+ measurement instrument.

Step 3.2 Define the descriptive analysis. In order to provide a support in the analysis of the data collected through the MEEGA+ measurement instrument, the analysis of the data collected is defined and analysed through descriptive statistics methods.

Define the analysis of the data collected. In accordance with the evaluation objective defined and the respective analysis questions, following the MEEGA+ model and the GQM approach (BASILI; CALDIERA; ROMBACH, 1994), the analysis of the data collected is defined.

Select the descriptive statistics methods. Descriptive statistics methods are used to describe, and graphically present interesting aspects of the data collected (WOHLIN et al., 2012). Thus, in order to analyse data collected through the MEEGA+ measurement instrument, descriptive statistics methods are selected (TROCHIM; DONNELLY, 2008; WOHLIN et al., 2012), such as measures of central tendency (median, average and frequency of responses) and graphical visualization (frequency charts).

Operationalize the data analysis. In this step, the descriptive statistics methods selected are used to calculate the measures of the central tendency and provide a graphical visualization of the data collected. Thus, following the MEEGA+ model and the GQM approach, the descriptive statistics are interpreted in order to answer the defined analysis questions and, thus, achieving the evaluation objective.

Step 3.3 Design the game quality scale. In order to use the MEEGA+ measurement instrument for the classification of the quality level of the game being evaluated, a MEEGA+ scale is developed. The MEEGA+ scale aims to classify the evaluated game regarding to its quality level, based on students' perception, allowing to identify which requirements correspond to the lowest or highest level of quality. The MEEGA+ scale is developed by adopting the statistical technique Item Response Theory (IRT) (PASQUALI; PRIMI, 2003), which allows to express through mathematical models the relationship between observable variables (questionnaire items) and latent traits (game's quality) based on the students' perceptions (PASQUALI; PRIMI, 2003).

Calculate the parameters of discrimination and difficulty. In order to calculate the parameters of discrimination (*a*) and difficulty (*b*) we use the probabilistic model proposed by Samejima (1969). The model of Samejima (1969) is used due to the nature of the analysed data, distributed in categories on a gradual scale (Likert scale). The parameter *a* is associated with how much the questionnaire item discriminates (differentiates) the students in relation to the latent trait (game's quality), where the higher its value, the more associated with the latent trait is the questionnaire item. The parameter *b* is associated with the degree of difficulty of the item, where the higher its value, the more difficult the students to agree with the questionnaire item in relation to the game's quality.

Identify the anchor items. In order to identify the anchor items, which determine the categories of the latent trait (game's quality), we consider the probability parameter $P_{i,k}(\theta)>=0,50$ (ANDRADE; TAVARES; VALLE, 2000).

Define the quality levels. The quality levels determined by the anchor items are defined considering the latent trait (game's quality), thus, defining and describing the quality levels of a measurement scale.

The quality level of the scale is defined using an average 50 and standard deviation 15, scale (50.15), applying the formula $\theta_{50.15}=50+15*\theta_{0.1}$ (ANDRADE; TAVARES; VALLE, 2000).

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. The process is modelled in a prescriptive way (ACUÑA; FERRÉ, 2001). The goal of prescriptive modelling is to define the required or recommended means of executing the process, thus, defining how the process should be performed, establishing rules, guidelines and standards (ACUÑA; FERRÉ, 2001).

Define the structure of the process model. The process is organized in phases, activities and work products (ACUÑA et al., 2000; BENALI; DERNIAME, 1992; FINKELSTEIN; KRAMER; NUSEIBEH, 1994). A phase is a step of the process, presenting a set of activities in a structured sequence. Activities are the stages of a process that implements procedures to transform a product. Work products are the inputs and outputs of an activity from a process, they may be produced and consumed throughout the process. In order to provide a graphical and standardized notation of the MEEGA+ process, we adopt the Business Process Modelling Notation (BPMN) (WESKE, 2012).

Define the phases, activities and work products. The phases, activities, and work products of the MEEGA+ process are defined, based on the empirical study process as proposed by Wohlin et al. (2012) and practical experiences in the conduction of game evaluations.

Define the work products (outputs). A template for each work product (output) generated during the process is defined. The templates are aimed to organize and standardize the information generated during the process in order to assist researchers in the definition, planning, execution, analysis, and presentation of the evaluation.

Step 5. Apply and evaluate the MEEGA+ method. The MEEGA+ method is applied for the evaluation of games for computing/SE education, as well as the evaluation of the MEEGA+ method by experts is conducted. The application of the MEEGA+ method is carried out through a series of case studies (WOHLIN et al., 2012; YIN, 2017) defined based on the GQM approach (BASILI; CALDIERA; ROMBACH, 1994) and adopting the MEEGA+ method. The case studies are conducted in different courses of higher computing education, from educational institutions in Brazil and abroad, using different educational games (digital and non-digital).

The evaluation of the MEEGA+ method is performed in two steps. The first step aims to evaluate the reliability and validity of the measurement instrument of the MEEGA+ model. In this step, the data collected in the conducted case studies are grouped and analysed in terms of reliability and validity, based on the definition of Trochim and Donnelly (2008) and on the scale development guide proposed by DeVellis (2016).

The second step aims to evaluate the quality of the MEEGA+ method from the experts' perspective. The MEEGA+ method is evaluated in terms of authenticity, validity, usability, correctness, completeness, consistency, understandability, unambiguousness, and flexibility (DAVIS, 1989; RITTGEN, 2010; MATOOK; INDULSKA, 2009; IEEE, 2010), through an expert panel (BEECHAM et al., 2005). By adopting the GQM approach (BASILI; CALDIERA; ROMBACH, 1994), the evaluation is defined and decomposed into analysis questions and metrics, which are collected through a questionnaire answered by the experts after analysing the MEEGA+ method.

Table 1 presents the mapping between the research steps and the specific objectives of this research.

Table 1 - Mapping of research steps and research objectives

Research step	Specific objectives
Step 1 – Identify the state-of-the-art & practice	O1, O2
Step 2 – Conduct a large-scale evaluation of the initial version of a prominent evaluation approach	O3
Step 3 – Design the MEEGA+ model	O4
Step 4 – Design the MEEGA+ process	O5
Step 5 – Apply and evaluate the MEEGA+ method	O6

Source: developed by the author.

1.5 ORIGINALITY

Based on the analysis of the state-of-the-practice, analysing two decades, we identified that the use of games for computing/SE education has been increased (BATTISTELLA; GRESSE VON WANGENHEIM, 2016; PETRI; GRESSE VON WANGENHEIM, 2017). However, we also observed that most of the evaluations of such games are performed in an ad-hoc manner in terms of research design, measurement, data collection and analysis. Typically, the game is applied and then a subjective feedback is collected through non-standard questionnaires with open questions, thus, indicating the lack of scientific rigor adopted in the evaluations (PETRI; GRESSE VON WANGENHEIM, 2017).

Consistent with the lack of scientific rigor applied in the evaluations, results of the analysis of the state-of-the-art also indicate that there is a lack of systematic approaches to evaluate games. Most of the few existing approaches are frameworks rather than comprehensive evaluation methods, indicating a lack of support on how to conduct and operationalize such evaluations. In addition, most of the existing approaches also seem to be developed in a rather ad-hoc manner, not providing an explicit definition of an objective, measures or data collection instruments (PETRI; GRESSE VON WANGENHEIM, 2016).

Therefore, based on the results of the state-of-the-art and practice, we observe that there is no a systematic method for the definition, execution, and data analysis of quality evaluations of games for computing/SE education. However, there are some related evaluation models and scales, such as MEEGA (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011) and EGameFlow (FU; SU; YU, 2009). MEEGA and EGameFlow are developed in a systematic way, defining a set of factors for the evaluation and also provide a questionnaire for data collection. However, analysing the validity and reliability of the initial version of the MEEGA model (Chapter 4), we observed limitations in its validity, in terms of overlapping of concepts between the evaluation factors, as well as in the validity of the items of the data collection instrument (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017). And, the EGameFlow scale, although evaluated in terms of validity and reliability, seems to have been proposed by the authors and then discontinued, not improving its initial version based on the results of its evaluation. In addition, these related approaches only propose an evaluation model/scale, not providing a systematic process, detailing phases, activities and work products for the conduction of evaluations.

In general, there are other related studies proposing generic approaches to game evaluations (FREITAS; OLIVER, 2006; CONNOLLY; STANSFIELD; HAINEY, 2009; CARVALHO, 2012; AK, 2012; ABDELLATIF; MCCOLLUM; MCMULLAN, 2018). However, besides these approaches have been developed in an ad-hoc manner, they do not define evaluation factors supported by theories nor provide data collection and analysis instruments. In addition, they do not provide a systematic process in order to guide researchers and instructors in the conduction of evaluations.

Thus, the originality of this research is the design and evaluation of a new method (MEEGA+) providing a comprehensive support for game evaluations. The MEEGA+ method is composed by an evaluation model (MEEGA+ Model), defining quality factors to be evaluated through a standardized measurement instrument, provides instruments for data analysis of the data collected in the evaluations and, provides a scale, which classifies the evaluated game according to its quality level. In addition, the MEEGA+ method contains a process (MEEGA+ Process), which describes in detail the phases, activities and work products to guide researchers and/or instructors in the conduction of evaluation of games for computing/SE education.

1.6 CONTRIBUTIONS

This research has produced scientific, technological, and social contributions.

Scientific contributions. This research has two major scientific contributions. The first one is the synthesis of the state-of-the-art and practice on evaluations of games for computing/SE education. This synthesis indicated a gap in the current field of evaluation of educational games for computing/SE education, which this research aims to cover. Another scientific contribution is the MEEGA+ method, providing a comprehensive support for a systematic evaluation of games used for computing/SE education. The MEEGA+ method 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, guiding researchers on how to plan, execute and analyse the results of game evaluations.

Additional scientific contributions of this research are the results of two systematic mappings of the literature and the large-scale evaluation of the initial version of the MEEGA model, identified as a prominent evaluation approach. Through the systematic mappings, the state-of-the-art and practice were obtained, identifying the existing approaches to game evaluation (PETRI; GRESSE VON WANGENHEIM, 2016), and analysing how games for computing education are evaluated (PETRI; GRESSE VON WANGENHEIM, 2017). In addition, by conducting a large-scale evaluation of the initial version of the MEEGA model, we obtained the reliability and validity analysis of the model, as well as the identification of its limitations and improvement opportunities (PETRI; GRESSE VON WANGENHIEM; BORGATTO, 2017).

Technological contributions. The main technological contribution of this research is the MEEGA+ method itself. In addition, secondary technological contributions are the instruments and the game quality scale provided by the MEEGA+ method. The method provides a standardized measurement instrument for data collection and a data analysis spreadsheet, both available in English, Spanish and Brazilian Portuguese. Additionally, the game quality scale adopts mathematical models to classify the evaluated game regarding to its quality level.

Social contributions. The social contribution of this research is the quality evaluation of the games, which are used as an instructional strategy in computing programs in educational institutions. Through the MEEGA+ method, it is possible to systematically evaluate the quality of the games currently used, providing a systematic support to instructors in deciding which games to apply in their courses. Thus, based on the results of the use of this systematic method, it is possible to choose which games contribute to computing/SE education, thus, contributing positively to the teaching of computing/SE professionals, enabling them to work on the development solutions for society in general.

1.7 DELIMITATION AND SCOPE OF WORK

The MEEGA+ method developed in this research is exclusively to evaluate games used as an instructional strategy in the computing/SE education context. Thus, the scope of this research is limited to the evaluation of games used to develop knowledge, skills and/or attitudes in any computing knowledge area. The results of reliability and validity of the MEEGA+ method presented in this research are also specific to the computing/SE context. The development of a new educational game and the evaluation of a game of other knowledge area are out of the scope of this research.

1.8 ALIGNMENT WITH THE PPGCC RESEARCH AREA

This research has been developed in the research area of Software Engineering of the Graduate Program in Computer Science (PPGCC). The objective of the Software Engineering research line is "to train individuals capable of conducting the software development process and to investigate new methodologies, techniques and tools for system design" (PPGCC, 2017).

Software Engineering is defined as a discipline that involves the application of theory, knowledge, and practice to effectively and efficiently build reliable software systems that satisfy the requirements of customers and users (ACM; IEEE-CS, 2013). Software engineering uses engineering methods, processes, techniques, and measurements. It benefits from the use of tools for managing software development; analysing and modelling software artefacts; **assessing and controlling quality**; and for ensuring a disciplined, controlled approach to software evolution and reuse (ACM; IEEE-CS, 2013).

In this context, the objective of this research, which aims at developing a new method for quality evaluation of games, is related to the evaluation of **software quality** (more specifically in the evaluation of software quality of educational games). Software quality is a major topic of Software Engineering, as reported in the Curriculum Guidelines for Computer Science (ACM; IEEE-CS, 2013), the Guide to the Software Engineering body of knowledge (SWEBOK) (BORQUE; FAIRLEY, 2014), and the Reference Curriculum of the Brazilian Computer Society (SBC, 2005). Thus, the development of a new method for **the evaluation of the quality of games**, a kind of specialized system (ACM; IEEE-CS, 2013), is considered adherent to the objectives of the Software Engineering research line of the PPGCC. Furthermore, the present work also contributes directly to Software Engineering Education, an important area for the formation of future SE professionals, being a research area/track of important international conferences such as the *International Conference of Software Engineering* (ICSE)[1] and the *Conference on Software Engineering Education and Training* (CSEE&T)[2].

1.9 WORK STRUCTURE

This work is divided into 8 chapters. In the following chapter (chapter 2), the theoretical background in order to facilitate the understanding of the main concepts used in this research is described. In chapter 3, a synthesis of the state-of-the-art and practice on evaluations of games for computing/SE education is presented. In chapter 4, the large-scale evaluation study of the initial version of the MEEGA model is presented. Chapter 5 presents the design of the MEEGA+, a method for the evaluation of the quality of games for computing/SE education. Chapter 6 presents the application and the evaluation of the MEEGA+ method, through a series of case studies conducted and an expert panel. In the chapter 7, the results of this research are presented. Lastly, the conclusions and future works are presented in chapter 8.

---

[1] https://www.icse2018.org/track/icse-2018-Software-Engineering-Education-and-Training

[2] http://conferences.computer.org/cseet/

2 THEORETICAL BACKGROUND

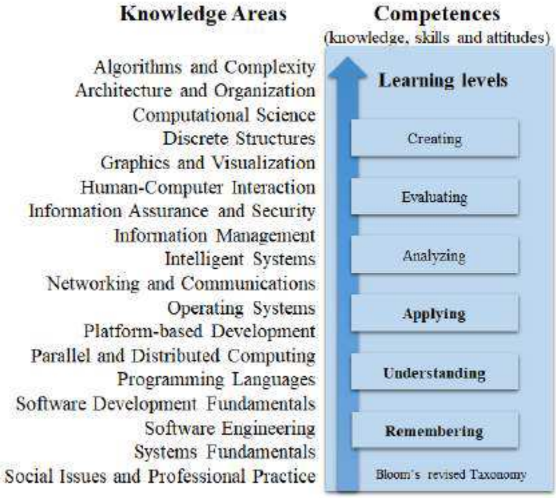
In this chapter, the theoretical foundations used in this work are presented in order to provide a better understanding of this research. The chapter is organized in five sections. Section 2.1 presents an overview of computing education, including the competencies that are expected to develop by computing students. Section 2.2 and 2.3 present the definitions of learning and teaching and the activities involved in these processes. In the section 2.4 the concepts of educational games and their use in computing/SE education are addressed. And, Section 2.5 presents the instructional design process, which is adopted to develop learning activities.

2.1 COMPUTING EDUCATION

In accordance with the Association for Computing Machinery (ACM), in a general way, we can define computing to mean any goal-oriented activity requiring, benefiting from, or creating computers. It involves the designing and building of hardware and software systems for different purposes (ACM; AIS; IEEE, 2005; ACM; IEEE-CS, 2013).

Education of students at the undergraduate level in computing generally occurs through five programs: Computer Engineering, Computer Science, Information Systems, Information Technology and Software Engineering (ACM, AIS, IEEE, 2005). In Brazil, the Ministry of Education and Culture (MEC), through the Brazilian Computing Society (SBC), also includes another program specifically for the training of computing teachers. Each computing program has a basic structure with an emphasis on a specific computing area (ACM; AIS; IEEE, 2005). However, in general, computing students are expected to develop computing competencies on different learning levels, as presented in Figure 3.

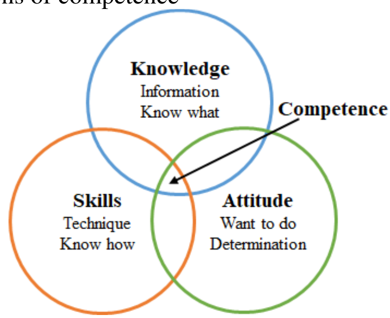
Figure 3 - Competencies to be developed by computing students



Source: (ACM; IEEE-CS, 2013; ANDERSON; KRATHWOHL; BLOOM, 2001).

The development of competencies (Figure 4) includes three dimensions: knowledge, skills and attitudes, integrating technical knowledge, work-related attitudes and behavioural and cognitive issues (DURAND, 1999; BRANDÃO; GUIMARÃES, 2001; COELHO; FUERTH, 2009). Knowledge refers to the set of data, information, concepts and perceptions acquired through education and experience. Skills refer to the capability to develop (know how) physical and intellectual tasks; and attitude refers to the predisposition of a person in relation to work, objects or situations (BRUNO-FARIA; BRANDÃO, 2003).

Figure 4 - Dimensions of competence



Source: (DURAND, 1999; BRANDÃO; GUIMARÃES, 2001)

In computing programs, students need to develop competencies on several learning levels, from simply remembering concepts, classifying and structuring the information, up to being able to apply the knowledge in concrete situations in different knowledge areas (ACM; IEEE-CS, 2013). In this respect, knowledge should typically developed on the learning levels of remembering, understanding and applying in accordance with the revised version of the Bloom's taxonomy (Table 2) (ANDERSON; KRATHWOHL; BLOOM, 2001; ACM; IEEE-CS, 2013).

Table 2 - Revised version of Bloom's taxonomy

Learning level	**Description**
1. Remembering	The ability to retrieving, recalling, or recognizing knowledge from memory.
2. Understanding	The ability to construct meaning by interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining.
3. Applying	The ability to use learned material through executing or implementing.
4. Analyzing	The ability to decompose material or concepts into parts, determining how the parts relate or interrelate to one another or to an overall structure or purpose.
5. Evaluating	The ability to judge, check, and critique the value of material for a given purpose.
6. Creating	The ability to put parts together to form a coherent or unique new whole.

Source: (ANDERSON; KRATHWOHL; BLOOM, 2001).

Besides technical knowledge, computing students also need to develop skills such as problem solving, interpersonal communication, teamwork, leadership, and management (ACM; IEEE-CS, 2013). The development of such skills is classified in accordance with the psychomotor levels of learning (Table 3) (SIMPSON, 1972).

Table 3 - Learning levels of the psychomotor domain

Learning level	**Description**
1. Perception	The ability to use sensory cues to guide activity, such as, e.g. detecting non-verbal communication cues.
2. Set	Readiness to act, such as, knowing and acting upon a sequence of steps in a communication process.

3. Guided response	The early stages in learning a complex skill that includes imitation and trial and error. Adequacy of performance is achieved by practicing, e.g., performs a communication process as demonstrated.
4. Mechanism	This is the intermediate stage in learning a complex skill. Learned responses have become habitual and the movements cab be performed with some confidence and proficiency.
5. Complex overt response	The skillful performance of acts that involve complex patterns without hesitation, and automatic performance.
6. Adaptation	Skills are well developed, and the individual can modify movement patterns to fit special requirements, e.g. respond effectively to unexpected experiences.
7. Origination	Creating new movement patterns to fit a particular situation or specific problem. Learning outcomes emphasize creativity based upon highly developed skills, such as the construction of a new communication theory.

Source: (SIMPSON, 1972).

Furthermore, computing students need to develop professional attitudes, such as a mature behaviour of the student, considering professional and legal issues as well as an ethical attitude in the profession (ACM; IEEE-CS, 2013). The development of such competencies typically targets the awareness and growth in attitudes, emotion, and feelings and can be classified in accordance to the taxonomy of the affective domain (Table 4) (KRATHWOHL; BLOOM; MASIA, 1973).

Table 4 - Learning levels of the affective domain

Learning level	**Description**
1. Receiving	The student passively pays attention.
2. Responding	The student actively participates in the learning process and reacts in some way expressing his/her compliance/willingness to respond (motivation).
3. Valuing	The student attaches a value to an object, phenomenon, or piece of information.
4. Organizing	The student can put together different values, information, and ideas and accommodate them within his/her own schema by contrasting different values, resolving conflicts between them, and creating a n unique value system.
5. Characterizing	The student holds a particular value or belief that now exerts influence on his/her behavior so that it becomes a characteristic.

Source: (KRATHWOHL; BLOOM; MASIA,1973).

Especially for the area of Software Engineering, which is a discipline concerned with the application of theory, knowledge, and practice to effectively and efficiently build reliable software systems that satisfy the requirements of customers and users (ACM; IEEE-CS, 2013; BORQUE; FAIRLEY, 2014), students are expected to master technical and non-technical competencies. Such SE competencies include the theoretical foundations of the discipline, the design methods, and the technology and tools of the discipline (ACM; IEEE-CS, 2013; BORQUE; FAIRLEY, 2014; GHEZZI; MANDRIOLI, 2005). In addition, students must be able to keep their knowledge current with respect to the new approaches and technologies, interact and communicate with other people (including the other culture), recognize a recurring problem, and reuse or adapt known solutions; manage a process and coordinate the work of different people (ACM; IEEE-CS, 2013; BORQUE; FAIRLEY, 2014; GHEZZI; MANDRIOLI, 2005).

The development of such competencies is, typically, a result of a teaching and learning process (FREITAS; BRANDÃO, 2005).

2.2 LEARNING

The learning process is a subject of study of the cognitive psychology. Different researchers such as Piaget, Vygotsky, Wallon, and Ausubel present learning theories that address various aspects and help understand how the learning occurs (RATIER, 2010). In general, the conceptions of these authors focus on the meanings that occur with a person, considering the indirect evidence to explain how individuals perceive, interpret and use the acquired knowledge (SALLA, 2012; GSI TEACHING & RESOURCE CENTER, 2015). Similarly, neuroscience has also been trying to understand the learning process, focusing on behavioural experiments and the observation of changes in the students' brain (GSI TEACHING & RESOURCE CENTER, 2015; SIGMAN et al., 2014). In both research areas there is evidence that factors such as emotion, motivation, attention, social environment and the association of new information with previous knowledge, directly influence the learning process (SALLA, 2012). The more emotion and affection in an event, it will be more meaningful and more easily it will be memorized in the brain (SALLA, 2012). Motivation stimulates the production of dopamine, mobilizing the attention of the student and potentializing the behaviour in relation to the event (HERCULANO-HOUZEL, 2007). The affectivity, motivation and the social environment interfere in the nervous system, stimulating the cognitive processes and causing changes in the brain and, consequently, improving concepts and knowledge already understood by the individual (SALLA, 2012).

Although it is difficult to define learning (HILGARD, 1975), some authors report their conceptions of this process. In accordance with Bruner (1977), learning is a process of understanding and it is related to the acquisition, transformation and evaluation of information arising from our experiences and experiences with the world (BRUNER, 1977). Learning implies the change of knowledge, skills and attitudes that will be applied in new situations in which the human will experience (POZO, 2002). According to Brandão and Borges-Andrade (2007), learning is related to changes and can be perceived in the comparison of the performance of a person before and after participating in a learning process. The performance acquired in the application of new competencies reveals that the individual learned something new (BRANDÃO; BORGES-ANDRADE, 2007). Driscoll (1994) explains that learning is the change in human performance resulting from interaction with the environment in which it lives. It is the construction of knowledge and skills resulting from behaviour stimulated by experiences (SCAGNOLATO, 2009).

In summary, defining learning is a complex task, it represents an intersection between different internal and external factors that contribute to this cognitive process. Thus, in this work, we consider that learning is a result of a teaching process that aims at develop/improve knowledge, skills and/or attitudes (COSENZA; WAR, 2011).

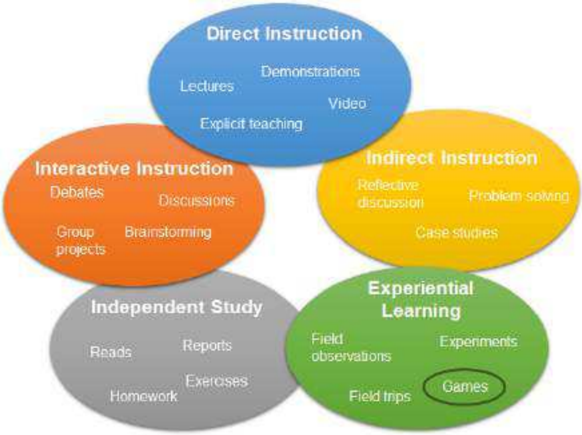
2.3 TEACHING

The development of competencies, resulting in learning, is facilitated by a teaching process (FERNÁNDEZ, 1998). Thus, teaching can be defined as any interaction that occurs the transmission of information from one individual to another, in order to improve its knowledge, skill or attitude (SPRINTHAL; SPRINTHAL, 1993).

In this context, the term instruction is also used and refers to the organization of activities or experiences guided by the instructor, with the purpose of to develop knowledge, skills and/or attitudes (DRISCOLL, 1994). Instructional units (IUs) are developed in order to structure the teaching process and facilitate the learning (BRANSFORD; BROWN; COCKING, 2000). They can have a different scope, for example, a course, a class, or a group of classes (BRANSFORD; BROWN; COCKING, 2000). Typically, an IU adopts one or more instructional strategies, which define how to address and expose the contents specified for that instructional unit (BRANSFORD; BROWN; COCKING, 2000; SASKATCHEWAN EDUCATION, 1991).

Instructional units that aim at developing competencies at higher learning levels, including the capability to be able to apply the knowledge in practice, and acquiring project experience, require the usage of active/experiential instructional strategies that enable deeper learning (WAGNER, 1970; GIBSON; BELL, 2013). An overview on different existing instructional strategies is presented in Figure 5.

Figure 5 - Classification of instructional strategies

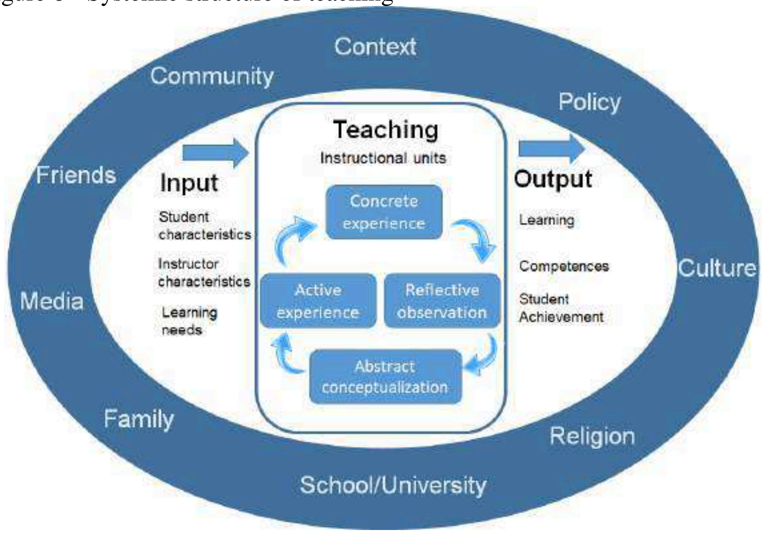


Source: (SASKATCHEWAN EDUCATION, 1991).

According to the classification of the instructional strategies (Figure 5) (SASKATCHEWAN EDUCATION, 1991), the direct instruction strategy is highly teacher-directed and is among the most commonly used. It is usually deductive, that is, the contents/concepts are presented and then illustrated with examples. Indirect instruction is student-centred, the student participates actively in the observation, investigation and formulation of hypotheses, the teacher acts as a facilitator. Interactive instruction is characterized by the sharing of information and ideas through debates and discussions mediated by the instructor. An individual study is student-centred, the student is responsible for the execution of activities that aim to enhance self-confidence and self-improvement. Experiential learning is inductive, learner centred, and activity oriented. It focuses on the application of the knowledge through active experiences. In this context, the object of study of this research is the educational games, a kind of experiential/active instructional strategy. Concepts and characteristics of educational games are presented in Section 2.4.

In summary, the teaching process represents a systemic structure composed of inputs, outputs and interrelated elements such as context, culture, family, friends, etc., which directly influence the learning results (FERNÁNDEZ, 1998). Figure 6 presents the characteristics and factors involved in the teaching process.

Figure 6 - Systemic structure of teaching



Source: (KOLB; FRY, 1975; OLIVEIRA, 1993).

The instructional units used in the teaching process need to be developed in a systematic way in order to create teaching experiences that make the construction of knowledge, skills, and/or attitudes more efficient, effective and attractive (MERRILL et al., 1996). Typically, instructional units are developed by adopting the instructional design process, presented in section 2.5.

2.4 EDUCATIONAL GAMES

Educational games are an active instructional strategy that typically involves competition and is organized by rules and restrictions to achieve a certain educational goal (ABT, 2002; PRENSKY, 2007; RITTERFELD; CODY; VORDERER, 2010; DJAOUTI et al., 2011). Besides promoting entertainment, they are specifically designed to teach people about a certain subject, expand and revise concepts, reinforce development, or assist them in drilling or learning a skill or seeking a change of attitude as they play (ABT, 2002; PRENSKY, 2007).

Educational games are characterized by a set of elements that intrinsically contribute to the use of games as an experiential instructional strategy. Such elements are presented in Table 5.

Table 5 - Elements of games

Element	Description
Goals	Every game has a goal that guides the player in the phases of the game. For example, arrive first or in the shortest time among the players.
Rules	The rules define what can and cannot be performed in the game, directing the players' decisions and ensuring that all players follow similar conditions.
Restrictions	Restrictions are introduced as incentives or resources such as money or lives. They can also determine dependencies by defining that a player can only perform certain actions if the conditions are met.
Narrative	The narrative of a game corresponds to the fictional history elaborated to guide the players and motivate their actions. This element is very present in Role Playing Games (RPG), however, there are games that do not have a narrative, such as quiz games.
Interaction	Interaction refers to how the game is played. The game can be played by a single player, competing against itself. It can have multi-players, where players compete with each other. Or, the game can be played in groups (multi-groups), competing with other groups.
Challenge, competition and conflict	Challenges play a role in keeping the player playing and prevents the game from becoming monotonous. Competition is one of the main factors that promote intrinsic motivation and that makes players stay focused and motivated to learn more and win the game. Conflicts are obstacles encountered throughout the game and make goals not so easily achieved.
Reward and feedback	The reward is a contribution to the player when completing tasks/objectives with success. The reward can be by enabling new features, or allowing, players to do things that were previously not possible. At the end of the game, the typical reward is to be the winner or to have a higher score in a ranking between players. Feedback is also an important element, especially in educational games, to show the results obtained by the player, clearly indicating where the student is right and wrong.

Source: (ABT, 2002; PRENSKY, 2007; GRESSE VON WANGENHEIM; VON WANGENHEIM, 2012).

In terms of platform, games are typically classified into digital and non-digital ones. Digital games are electronic games that involve human interaction with a user interface to generate visual feedback on an electronic device such as smartphones, computers, tablets, etc.

(MITAMURA; SUZUKI; OOHORI, 2012), whereas non-digital games are played with non-digital resources such as game boards, cards, pencils and papers, etc. (CONNOLLY; STANSFIELD; HAINEY, 2007). Table 6 presents a summary of the classification of educational games in terms of platforms (CONNOLLY et al., 2012; CAULFIELD; VEAL; MAJ, 2011; BATTISTELLA; GRESSE VON WANGENHEIM, 2016).

Table 6 - Games' classification in terms of platform

Classification			**Definition**
Digital game			Electronic game that involves human interaction with a user interface to generate visual feedback on an electronic device.
	PC (Personal computer)	*Stand-alone*	Game played on a personal computer.
		Online	Game played on a computer network (internet), using a personal computer.
	Console		Game played on a specialized electronic device that connects to a common television set or composite video monitor.
	Mobile		Game played on a mobile device, such as smartphone, tablet, media player, etc.
Non-digital game			Game that is not played on an electronic device, it is also called a traditional/handheld game.
	Board		Game that involves markers or pieces moved on a pre-marked board, following a set of rules.
	Cards		Game that uses cards as the primary resource.
	Paper & pencil		Game that can be played with only paper and pencil.
	Prop game		Game that is played using props (portable objects).

Source: (CAULFIELD; VEAL; MAJ, 2011; CONNOLLY et al., 2012; BATTISTELLA; GRESSE VON WANGENHEIM, 2016).

Besides of classifying in terms of platforms, games also cover a broad spectrum of genres, which determine the general style used by the game. Table 7 presents a classification of educational games in terms of genre (HERZ, 1997; WOLF, 2001; ADAMS; ROLLINGS, 2006; SCHIFFLER, 2006; KARNER; HÄRTEL, 2011; BATTISTELLA; GRESSE VON WANGENHEIM, 2014).

Table 7 - Games' classification in terms of genre

Genre	**Definition**
Action	A game that requires players to use fast reflections, accurately to overcome obstacles, solve challenges or answer questions.
Divination	A game in which the goal is to identify some kind of information, such as a word, from drawings or images.
Adventure	A game in which the player takes on the role of protagonist of the story and interacts through the exploration and the solution of challenges/missions (decoding messages, finding and using items).
Strategy	A game that requires tactical thoughts in order to achieve victory.

Puzzle	A game that involves characters that control a set of commands, to navigate through mazes or to organize objects correctly.
Quiz	A game in which the player needs to answer questions for a particular area of knowledge.
Role-playing game (RPG)	A game in which the player controls a protagonist's actions and with this character lives immersed in a fictional world.
Roll-and-move	Board games in which the markers are moved based on the results shown in one or more dice.
Simulation	A game in which the player is in control of a certain environment or activity, which seeks to be as realistic as possible.

Source: (HERZ, 1997; WOLF, 2001; ADAMS; ROLLINGS, 2006; SCHIFFLER, 2006; KARNER; HÄRTEL, 2011; BATTISTELLA; GRESSE VON WANGENHEIM, 2014).

2.4.1 Educational Games for Computing/SE Education

Driven by the need to provide more hands-on opportunities for computing students, various educational games on different platforms and genres have been used to contribute to the learning process in different computing knowledge areas (BACKLUND; HENDRIX, 2013; BATTISTELLA; GRESSE VON WANGENHEIM, 2016; KOSA et al., 2016).

Currently, there exists a vast variety of educational games (more than 100) to teach computing competencies in higher education (BATTISTELLA; GRESSE VON WANGENHEIM, 2016) (Figure 7). The majority is digital games, principally PC games, with a considerable trend also to non-digital games (paper & pencil, board games, etc.). Predominant are simulation games, which allow students to practice competencies through the simulation of real-life situations in a realistic environment, while keeping them engaged in the game (BATTISTELLA; GRESSE VON WANGENHEIM, 2016). On the other hand, there also exist several games designed to teach computing that have learning objectives at lower cognitive levels (BATTISTELLA; GRESSE VON WANGENHEIM, 2016). Typically, these games are used to review and reinforce knowledge taught beforehand using different instructional strategies.

Figure 7 - Examples of games for computing education

Source: (BATTISTELLA; GRESSE VON WANGENHEIM, 2016).

The majority of games for computing education focuses on teaching Software Engineering (BATTISTELLA; GRESSE VON WANGENHEIM, 2016; PETRI; GRESSE VON WANGENHEIM, 2017) (e.g., SimSE (NAVARRO; VAN DER HOEK, 2007), Requirements Collection and Analysis Game (RCAG) (HAINEY et al., 2011), X-MED (GRESSE VON WANGENHEIM; THIRY; KOCHANSKI, 2009), Tower-Defense (RUSU et al., 2011), ProDec (CALDERÓN; RUIZ, 2013), SimSoft (BAVOTA et al., 2012), SCRUMIA (GRESSE VON WANGENHEIM; SAVI; BORGATTO, 2013)). This emphasis on Software Engineering can be explained by the possibility they offer to provide practical experience to the students in a safe and controlled environment, helping to achieve learning on the application level, which otherwise due to practice restrictions may not be possible in Software Engineering (NAVARRO; VAN DER HOEK, 2005; BATTISTELLA; GRESSE VON WANGENHEIM, 2016).

Many of these games for SE education are simulation games, typically covering topics of Software Engineering Management (e.g., DELIVER! (GRESSE VON WANGENHEIM; SAVI; BORGATTO, 2012), SimSoft (BAVOTA et al., 2012), SCRUMIA (GRESSE VON WANGENHEIM; SAVI; BORGATTO, 2013)), in which the learner assumes the role of a project manager and performs the planning, monitoring and control of a software project (BATTISTELLA; GRESSE VON WANGENHEIM, 2016); the Software Engineering Process (e.g., SimSE (NAVARRO; VAN DER HOEK, 2007), Problems and Programmers (BAKER; NAVARRO; VAN DER HOEK, 2003)), simulating the execution of a specific kind of software development process or requirements engineering; Software Testing (e.g., Secret Ninja Testing (BELL; SHETH; KAISER, 2011), U-Test (THIRY; ZOUCAS; SILVA, 2011)), in which the player must solve challenges by preparing unit test cases. In addition, there can also be observed a trend to quiz games in order to review SE knowledge (e.g., PM Master (GRESSE VON WANGENHEIM, 2012), PMQuiz (PETRI et al., 2016)).

Other computing knowledge areas for which a considerable number of games have been encountered are Software Development Fundamentals (e.g., Wu's Castle (EAGLE; BARNES, 2009), Light-Bot (GOUWS; BRADSHAW; WENTWORTH, 2013), C-Jump (SINGH; DORAIRAJ; WOODS, 2007)), Algorithms & Complexity (e.g., SortingCasino (HAKULINEN, 2011), Algorithms Recursive Game (ROSSIOU; PAPADAKIS, 2007)), Architecture & Organization (e.g., Age of Computers (SINDRE; NATVIG; JAHRE, 2009), Computer Architecture Mini-Game (MELERO; HERNÁNDEZ-LEO; BLAT, 2012)), and Information Assurance & Security (e.g., CyberCIEGE (RAMAN et al., 2014), CounterMeasures (JORDAN et al., 2011); Security Protocol Game (HAMEY, 2003), Anti-Phishing Phil (SHENG et al., 2007)). Further computing knowledge areas seem to adopt fewer games for computing education (BATTISTELLA; GRESSE VON WANGENHEIM, 2016; PETRI; GRESSE VON WANGENHEIM, 2017).

2.5 INSTRUCTIONAL DESIGN

Instructional design is defined as a process for identifying a learning need, and designing, developing, and evaluating a way to address this need (BRANCH, 2009; DICK; CAREY, 2006). In accordance with Filatro (2008) instructional design is a systematic planning and development action, where different methods, techniques, strategies, activities and educational products are applied to enhance the teaching and learning process.

Currently, there are different instructional design approaches, such as the Systems Approach Model (DICK; CAREY, 2006), Kemp (MORRISON; ROSS; KEMP; KALMAN, 2010), and the ADDIE (Analyse, Design, Develop, Implement, and Evaluate) (BRANCH, 2009). ADDIE is the most widely used and consolidated instructional design approach, due to the fact that it is generic and flexible for any kind of instructional unit and it represents a formalism, defining a process including phases and activities for the instructional design (BRANCH, 2009). The ADDIE approach is organized in 5 phases (Figure 8), describing a basic process to systematically develop an instructional unit.

Figure 8 - The ADDIE approach

Source: (BRANCH, 2009).

The phases of the ADDIE approach are described in the next sections, with an emphasis in the evaluation phase, which is the focus of this research.

2.5.1 Analyse

The purpose of the analyse phase is to identify the instructional needs of the intended audience, and characterize the learning environment (BRANCH, 2009). The main procedures often associated with the analyse phase are understand the instructional needs, characterize the intended audience, context and resources required, and determine the learning goals.

The main objective of the instructional design is to generate instruction in order to cover instructional needs. Instructional needs are defined by a performance gap, which is characterized by the discrepancy between a desired performance and the actual performance. Typically, the cause for a performance gap is the lack of knowledge or skills in a specific content. Thus, an individual has the resources to perform a task, but the intellectual or psychomotor skill to perform as desired is not evident. Therefore, the instruction is a response in order to cover this performance gap.

Besides identifying the instructional needs, it is necessary to characterize the learning context, resources and the target audience. The characterization of the learning environment directly influences the design of IU activities. Characteristics of the learning context include the educational institution, the program and/or course that the activities will be performed, the class time, the time available for the IU, and the frequency of meetings. In addition, the target audience should be characterized in terms of the number and location of the students, average age, gender distribution, educational level, and experience levels, which determines the previous knowledge in the subject of the IU. Furthermore, all types of materials and resources that will be required to complete the IU should be defined. It includes technological resources (e.g., computers, internet, smartphones, or any other electronic devices), non-technological resources (e.g., flip charts, writing utensils, dry marker boards, adhesive notepads, etc.), and instructional facilities (e.g., number of rooms required, the number of students that can be accommodated per room, number of computers available per students, etc.).

After identifying the instructional needs and characterize the context, the learning (or performance) goals are defined. A goal is defined as the end toward which all effort is directed, thus, a learning goal is a specific description of a knowledge, skill, and/or attitude that the students should develop when complete the IU. Typically, the learning goals are described, including a condition component, a performance component, and a criterion component:

- Performance: An observable behaviour that demonstrates that a student learned the desired competence. It is represented by an action verb that is measurable and observable.
- Condition: Important circumstances under which the performance is expected to occur.
- Criterion: The accuracy of performance that is considered acceptable. If the criterion is omitted, it is assumed that the accuracy is 100%.

Typically, Bloom's taxonomy is adopted in order to support the definition of such learning goals. According to the performance criterion of the defined learning goals, a different Bloom's taxonomy domain is used such as cognitive (knowledge) (Table 2) (ANDERSON; KRATHWOHL; BLOOM, 2001), psychomotor (skills) (Table 3) (SIMPSON, 1972), and affective (attitudes) (Table 4) (KRATHWOHL; BLOOM; MASIA, 1973).

2.5.2 Design

The purpose of the design phase is to prepare a set of functional specifications for closing the performance gap identified in the previous phase (BRANCH, 2009). The main procedures often associated with this phase are the generation of contents, definition of instructional strategies to achieve the defined learning goals and the definition of testing strategies to verify the desired performance.

Content is the focal point for engaging the student during the process of knowledge construction. However, content should be strategically introduced during the teaching and learning process. The generation of the contents used in the IU aims to meet the learning needs defined in the analysis phase. Typically, the learning contents are addressed adopting instructional strategies.

An instructional strategy defines the means used by instructors to achieve with effectiveness the learning goals (DICK; CAREY, 2006). According to the knowledge, skill or attitude addressed in the learning goals, a particular instructional strategy may be more effective in promoting student learning (DICK; CAREY, 2006, BRANCH, 2009). Instructional strategies are classified in direct instruction, indirect instruction, interactive instruction, individual study, and experiential learning, as previously presented (Figure 5). Besides the learning goals, the choice of the instructional strategies should also consider the interests and needs of students in order to facilitate and improve their learning (DICK; CAREY, 2006, BRANCH, 2009).

In order to verify if the learning goals were effectively achieved by the students testing strategies are applied. Testing strategies provide feedback to the instructor about whether learning is occurring and to the student about the progress s/he is making towards accomplishing the performance tasks. Examples of testing strategies are tests, research works, implementation, presentation, etc.

2.5.3 Develop

The purpose of this phase is to select and/or develop all learning resources that will be needed to undertake the planned for the IU. The main procedures are the selection and development of supporting media needed to implement the planned instruction, and the conduction of revisions and pilot tests (BRANCH, 2009).

In order to support the learning activities of the IU, media should be selected and/or developed. Instructional media are considered tools to extend the capability of the instructor and extend the capability of the student. Its selection or development should consider the context, learning goals, culture and students' learning levels. Examples of media are slides, videos, charts, storyboard for students to draw, educational software, games, etc.

In this phase, a formative evaluation is also conducted through revisions and pilot tests. A formative evaluation is the process of collecting data that can be used to revise the learning activities before implementation. A feedback is collected from students and instructors in order to improve the instructional material produced, thus making the instruction more effective. Pilot tests are used to identify restrictions in the IU, e.g., in terms of technological resources, time, physical space, etc.

2.5.4 Implement

The purpose of the implement phase is to execute the planned learning activities, resources, and/or media in the target audience in order to close the performance gap by achieving the learning goals (BRANCH, 2009). Common procedures associated with this phase are to prepare the instructor and to prepare the student to interact with the learning activities.

Before the application of the learning activities in the target audience, it is necessary to prepare the instructor(s). The instructor is responsible for facilitating the learning activities, setting the pace, providing guidance and assistance, furnishing subject matter expertise, and assisting in the evaluation.

The execution of the learning activities starts with an explanation about the rules and activities that the students will be perform, thus, preparing them to actively participate in the instruction and effectively interact with the newly developed learning resources in order to close the performance gap. During the interaction of the students with the learning activities, it is important observing and record the time spent on the activities, the interactions of the students with others, verify if the contents and activities are being performed as planned. These observations can be used to improve the IU.

After completing the execution of the learning activities, data should be collected from the participants in order to evaluate the IU. Details about the data collection process are addressed in the evaluation phase.

2.5.5 Evaluate

The purpose of the evaluate phase is to determine the level of success of the IU through a summative evaluation. A summative evaluation aims to determine the potential effectiveness of learning activities conducted, evaluating whether the target audience has achieved the defined learning goals, as well as identify improvement opportunities for subsequent projects that are similar in scope (BRANCH, 2009).

An evaluation represents a part of a cyclic process (Figure 9) that, through its results, provides feedback for improvement of the objectives of the IU and, consequently, contributes to the effectiveness of the teaching and learning process.

Figure 9 - Evaluation

Source: (GAGNE; BRIGGS; WANER, 1992).

Typically, the evaluation of an IU is conducted through an empirical study, which involves collecting data directly from the participants after completing the IU. The conduction of an empirical study follows a systematic process, which involves several phases as presented in Figure 10 (WOHLIN et al., 2012).

Figure 10 - Phases of an empirical study

Source: (WOHLIN et al., 2012).

In the **scoping phase**, the evaluation objectives are defined. This includes the explicit specification of the object of study, the factors to be evaluated, and the context. In order to support the definition of the evaluation the Goal/Question/Metric (GQM) (BASILI; CALDIERA; ROMBACH, 1994) approach may be adopted. GQM is a goal-oriented approach that defines a measurement program on three levels: conceptual, operational, and quantitative. In the conceptual level (goal) a goal is defined for an object of study, for a variety of reasons, with respect to various models of quality, from various points of view and relative to a particular environment. In the operational level (question) a set of analysis questions is used to define models of the object of study and then focuses on that object to characterize the evaluation or achievement of a specific goal. In the quantitative level (metric) a set of metrics is associated with each analysis question in order to answer it in a measurable way. In order to achieve the goals, the data collected are interpreted in a bottom-up way (BASILI; CALDIERA; ROMBACH, 1994).

Depending on the objective of the evaluation and the object of study, existing well-defined models, methods, process, scales or frameworks can also be adopted to conduct the research (HEVNER, 2010). A model defines a set of characteristics and sub characteristics and their relationships, providing a basis for specifying the constructs of a specific area (HEVNER, 2010; IEEE, 2010). A method is a systematic approach to achieve a certain objective or result and, which describes the characteristics of an ordered process or a procedure used in the engineering of a product or service (IEEE, 2002; IEEE, 2010). A process is a sequence of steps and interrelated activities that are performed for a particular purpose (IEEE, 2002; IEEE, 2010). Scales are systematic instruments to measure variables in a specific research area (HEVNER, 2010). A framework is, or contains, a (not completely detailed) structure or system for the realization of a defined result/goal (HEVNER, 2010).

In the **planning phase**, the evaluation level is defined, based, for example, on the Kirkpatrick's four-level model for evaluation (KIRKPATRICK; KIRKPATRICK, 2006), as shown in Table 8.

Table 8 - Kirkpatrick's four-level evaluation model

Level		**Evaluation description and characteristics**	**Examples of evaluation methods and instruments**
1	Reaction	Evaluates how the participants felt about the training or learning experience.	Feedback forms; verbal reactions; post-training questionnaires
2	Learning	Evaluates the increase in knowledge or skills.	Reviews and tests before and after training; interview and observations
3	Behaviour	Evaluates the degree to which new learning acquired actually transfers to the job performance.	Observations and interviews over time to assess changes

4	Results	Evaluation of the effect on the business environment by the learner.	Observation and measurement over time; interviews with participants, their managers and customer groups

Source: (KIRKPATRICK; KIRKPATRICK, 2006).

In accordance with the evaluation level, an appropriate research design is defined. In order to be considered a (true) experiment, the study needs to involve multiple groups with random assignment. If the study uses either multiple groups or multiple moments of measurement, yet no random assignment, the study is classified as quasi-experimental. If the study does not use multiple groups, but is conducted in a systematic way, such as case studies (YIN, 2017), the study is classified as non-experimental. Studies executed in an unsystematic manner, without an explicit definition of the study and the measurement beforehand, are classified as ad-hoc studies. Common study types and research designs used in evaluations are summarized in Table 9.

Table 9 - Common types of research design

Evaluation level **(KIRKPATRICK; KIRKPATRICK, 2006)**	**Study type**	**Design**	**Representation** **X=Treatment** **O=Measurement** **R=Random assignment**
1 – Reaction	Non-experimental	One-shot post-test only	X O
2 – Learning	Non-experimental	One-shot pre-test/post-test	O X O
	Quasi - experimental	Static group comparison group	X O O
		Static group pre-test – post-test	O X O O O
		Times Series	O O X O O
	Experimental	Randomizes post-test only	R X O R O
		Randomized pre-test/post-test	R O X O R O O
		Randomized pre-test/post-test control group	R O X1 O R O X2 O

Source: (SHADISH; COOK; CAMPBELL, 2002; GRESSE VON WANGENHEIM; SHULL, 2009).

In order to achieve the evaluation goal(s), measurement has to take place (FENTON; PFLEEGER, 1998; WOHLIN et al., 2012). Therefore, measures and data collection instruments have to be defined in a way that allows the evaluation goal to map to the data that is collected and also provide a framework for analysing and interpreting the data with respect to the goals. According to the type of data to be collected, different data collection instruments may be used, for example, questionnaires, interviews, observations, focus group discussion, knowledge tests, etc. (Table 10) (SAUNDERS; LEWIS; THORNHILL, 2009; JOHNSON; CHRISTENSEN, 2016).

Table 10 - Data collection instruments

Data collection instrument	Definition
Questionnaire	A data collection instrument that includes a series of questions and other prompts for the purpose of gathering information from respondents.
Interview	Interviews consist of collecting data by asking questions. Data can be collected by listening to individuals, recording, filming their responses, or a combination of methods.
Observation	Observation is an instrument that is based on the performance of trained observers to obtain certain types of information in the natural environment of individuals.
Focus group	Focus group is a structured discussion with the purpose of stimulating conversation around a specific topic.
Knowledge test	Knowledge test is an instrument where students respond to written tests (including objective and/or discursive questions) in order to identify their knowledge in a specific topic.

Source: (SAUNDERS; LEWIS; THORNHILL, 2009; JOHNSON; CHRISTENSEN, 2016).

Furthermore, according to the data collection instrument adopted in the evaluation, a different type of scale can be used in order to categorize the data collected, such as nominal, ordinal, interval, or ratio scale (Table 11) (SAUNDERS; LEWIS; THORNHILL, 2009; JOHNSON; CHRISTENSEN, 2016).

Table 11 - Scales

Scale	Definition
Nominal	Nominal scale is defined as a scale used for labelling variables into distinct classifications and does not involve a quantitative value or order.
Ordinal	The ordinal scale of measurement includes the variables that have the property of rank or order. The variables coming under this scale must have set of rankings. Every value evaluated on an ordinal scale has a unique meaning.
Interval	Interval scale of measurement includes categories in which the distances or intervals between the categories are to be compared.
Ratio	Ratio scale is defined as a variable measurement scale that not only produces the order of variables, but also makes the

	difference between variables known along with information on the value of true zero.

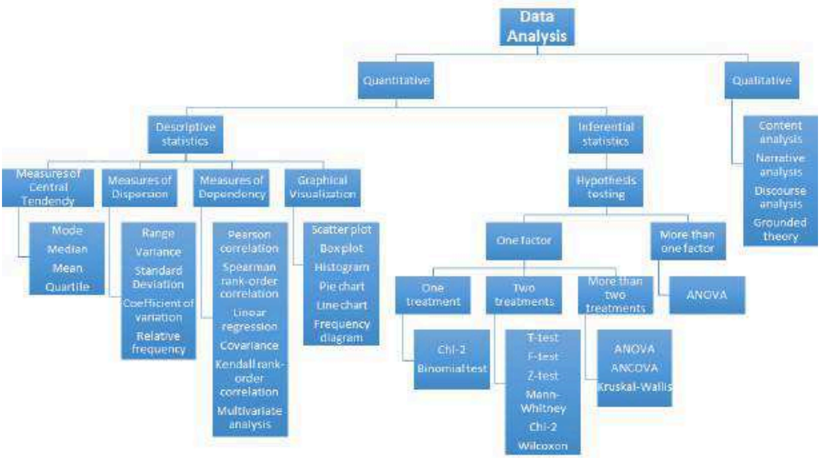
Source: (SAUNDERS; LEWIS; THORNHILL, 2009; JOHNSON; CHRISTENSEN, 2016).

In empirical studies, the Likert-scale (TROCHIM, DONEELY, 2008), a kind of rating scale, is the most widely used approach to scaling responses in researches that adopt questionnaires. It is a scale that allows the respondent to indicate how strongly s/he agrees or disagrees with a statement. Typically, it contains a set of statements presented with affirmations on a scale of 5 to 7 points ranging from strongly disagree to strongly agree (SAUNDERS; LEWIS; THORNHILL, 2009; JOHNSON; CHRISTENSEN, 2016).

The **operation phase** includes the preparation and execution of the study by applying the treatment (the educational game and optionally other instructional strategies for comparison). After the application of treatments with the target audience, the data collection instruments are applied in order to collect data as defined.

During the **analysis & interpretation phase**, data collected are analysed with respect to the evaluation goal(s). Depending on the nature of the collected data, this may be done by using qualitative and/or quantitative analysis methods ranging from descriptive statistics to inferential statistics as summarized in Figure 11 (FREEDMAN; PISANI; PURVES, 2007; WOHLIN et al., 2012).

Figure 11 - Data analysis methods



Source: (FREEDMAN; PISANI; PURVES, 2007; WOHLIN et al., 2012).

The results of the data analysis are interpreted, answering the analysis questions and, consequently, obtaining the evaluation goal.

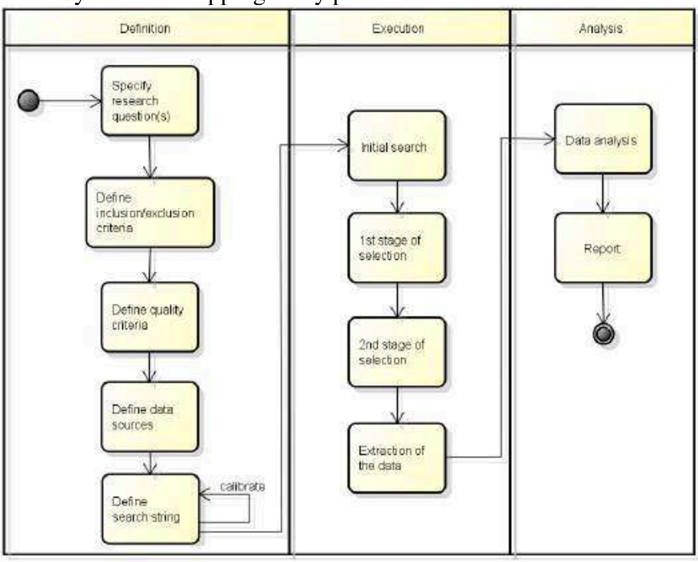
The last phase of the empirical study process involves **presenting and packing** of the results (WOHLIN et al., 2012). This includes documentation of the results, which can be made either through a research paper for publication, a technical report or as part of an experience base. Besides formally presenting all definitions of the study in a comprehensive document, it is important that the lessons learned are also documented in an appropriate way.

3 STATE OF THE ART AND PRACTICE

This chapter presents the state of the art and the practice obtained through two systematic mapping studies. Section 3.1 presents the state-of-the-art of existing approaches (methods, models, frameworks, scales) to systematically evaluate educational games. Section 3.2 presents the state of the practice on how games used for computing education are evaluated. Threats to the validity of the conducted mapping studies and mitigation strategies are presented in section 3.3. The main results obtained analysing the state of the art and the practice are discussed in Section 3.4.

Systematic mapping is a type of systematic literature review, which aims to identify, evaluate and interpret as many relevant studies as possible and available the research question, research topic or phenomenon of interest (KITCHENHAM, 2010). As mapping studies use the same basic methodology as systematic literature reviews, the two systematic mapping studies follow an adaptation of the procedure proposed by Petersen et al. (2008) and Kitchenham (2010), as presented in Figure 12.

Figure 12 - Systematic mapping study process



Source: (PETERSEN et al., 2008; KITCHENHAM, 2010).

A detailed description of the literature reviews can be found in Petri and Gresse von Wangenheim (2016) presenting the state-of-the-art, and Petri and Gresse von Wangenheim (2017), presenting the state of the practice.

3.1 STATE OF THE ART

In order to elicit the state of the art on how to systematically evaluate educational games, we conducted a systematic mapping study, following the steps defined in Figure 12.

3.1.1 Definition

This research aims at the elicitation of the state of the art on how to systematically evaluate educational games (PETRI; GRESSE VON WANGENHEIM, 2016). In accordance with this objective, we performed a mapping study, focusing on the following analysis questions (AQS):

Analysis Questions

AQS1: Which models, methods, scales, or frameworks (approaches) exist to systematically evaluate educational games?

AQS2: Which quality and/or sub-quality factors are evaluated?

AQS3: How data collection and analysis is operationalized?

AQS4: How these approaches have been developed?

AQS5: How these approaches have been evaluated?

Inclusion/Exclusion Criteria. In accordance with our research objective/questions, criteria for selecting only relevant studies were defined. We included only articles that presented a well-defined approach to systematically evaluate educational games for teaching any knowledge area. We focused only on articles written in English (or in Portuguese with an abstract in English), available via digital libraries published between January 1995 and August 2018.

On the other hand, we excluded:

- Any study not related to a well-defined approach;
- Articles that present the evaluation of an educational game, but do not use a well-defined approach.
- Any article published by the author in the scope of this work.

Quality criteria. In addition to our inclusion/exclusion criteria, we also superficially assessed the quality of the reported studies, considering only articles that provide substantial information on the evaluation approach.

Data Sources and Search String. Data sources have been chosen based on their relevance in the computing domain, including: ACM Digital Library, IEEE Xplore, Springer Link, Science Direct and Wiley Online Library. In addition, we also searched via Google Scholar, in order to also consider articles published outside the computing domain, but which may provide a relevant contribution.

In accordance with our research objective, we defined the search string by identifying core concepts such as model, educational games, and evaluation, including also synonyms as indicated in Table 12.

Table 12 - Keywords

Core Concepts	Synonyms
model	method, framework, scale
educational games	serious games, game-based learning
evaluation	assessment

Source: developed by the author.

Using these keywords, the search string has been calibrated and customized in conformance with the specific syntax for each of the data sources as presented in Table 13.

Table 13 - Search strings

Data source	Search String
ACM Digital Library	(model OR method OR framework OR scale) AND ("educational games" OR "serious games" OR "game-based learning") AND (evaluation OR assessment) for: ((model OR method OR framework OR scale) AND ("educational games" OR "serious games" OR "game-based learning") AND (evaluation OR assessment)) Published since January 1995
IEEE Xplore	((model OR method OR framework OR scale) AND ("educational games" OR "serious games" OR "game-based learning") AND (evaluation OR assessment) IN metadata) AND (pyr >= 1995 AND pyr <= 2018)
Springer Link	'(model OR method OR framework OR scale) AND ("educational games" OR "serious games" OR "game-based learning") AND (evaluation OR assessment)' published between 1995 - 2018
Science Direct	pub-date > 1994 and ((model OR method OR framework OR scale) AND ("educational games" OR "serious games" OR "game-based learning") AND (evaluation OR assessment))
Wiley Online Library	(model OR method OR framework OR scale) AND ("educational games" OR "serious games" OR "game-based learning") AND

	(evaluation OR assessment) in All Fields between years 1995 and 2018
Google Scholar	(model OR method OR framework OR scale) AND ("educational games" OR "serious games" OR "game-based learning") AND (evaluation OR assessment) Custom range: 1995-2018

Source: developed by the author.

3.1.2 Execution

The mapping study was conducted in October and November 2015 and updated in September 2018, by the author and was reviewed by a senior researcher (author's advisor). Table 14 summarizes the returned results per data source. From Google Scholar we selected only the 1,000 most relevant results (100 first pages), from ACM Digital Library and Science Direct the 1,000 most relevant results, observing a lack of relevancy after these quantities. From IEEExplore, SpringerLink and Wiley online library all returned articles were analysed. As a result, a total of 4,978 articles were analysed during the first stage.

Table 14 - Search results

	Google Scholar	ACM	IEEExplore	Springer Link	Science Direct	Wiley	Total
Total analysed during 1st stage	1,000	1,000	283	886	1,000	809	4,978
Selected after 1st stage	73	13	25	10	19	9	149
Selected after 2nd stage	5	0	5	0	3	1	14

Source: developed by the author.

During the first stage, the search results were quickly analysed based on their title and short summary. The abstract was read only in case the title did not provide evidence of any exclusion criteria. Irrelevant and duplicate papers were removed. This stage left us with 149 potentially relevant articles. Then, we performed a second stage of selection. As a result, 14 articles (describing a total of 10 approaches) were identified as primary studies. The selected articles are presented in Table 15.

Table 15 - Selected articles

Publication year	Title	Reference
2006	How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?	(FREITAS; OLIVER, 2006)
2008	Development a general framework for evaluating games-based learning	(CONNOLLY; STANSFIELD; HAINEY, 2008)
2009	Towards the development of a games-based learning evaluation framework	(CONNOLLY; STANSFIELD; HAINEY, 2009)
2009	EGameFlow: a scale to measure learners' enjoyment of e-learning games	(FU; SU; YU, 2009)
2010	A refined evaluation framework for games-based learning	(HAINEY; CONNOLLY; BOYLE, 2010)
2011	A model for the evaluation of educational games for teaching software engineering	(SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011)
2012	Is game-based learning suitable for engineering education?	(CARVALHO, 2012)
2012	A game scale to evaluate educational computer games	(AK, 2012)
2012	Towards a comprehensive methodology for the research and evaluation of serious games	(MAYER, 2012)
2013	A brief methodology for researching and evaluating serious games and game-based learning	(MAYER et al., 2013)
2014	The research and evaluation of serious games: Toward a comprehensive methodology	(MAYER et al., 2014)
2015	Towards a construction and validation of a serious game product quality model	(GARCIA-MUNDO; GENERO; PIATTINI, 2015)
2017	An efficient framework for game-based learning activity	(CHEW, 2017)
2018	Serious games: Quality characteristics evaluation framework and case study	(ABDELLATIF; MCCOLLUM; MCMULLAN, 2018)

Source: developed by the author.

3.1.3 Data extraction

In accordance with the defined research questions, we systematically extracted information in a spreadsheet from each article selected for analysis. Table 16 shows the data items that were extracted.

Table 16 - Data items extracted

Research question	Data Item	Description
AQS1: Which models, methods, scales, or frameworks (approaches) exist to systematically evaluate educational games?	Reference	Reference of the study.
	Name	Acronym or name of the approach.
	Instructional strategy	The instructional strategy focused by approach.
AQS2: Which quality and/or sub-quality factors are evaluated?	Quality (sub-) factor(s)	Quality (sub-) factor(s) that are evaluated.
	Theoretical basis	The theoretical construct(s) used to define the quality factors that are evaluated.
AQS3: How data collection and analysis is operationalized?	Study type	Study type classified based on Table 9 following common research designs used in education contexts.
	Data collection instrument(s)	Instrument(s) used for data collection, such as questionnaires, interviews, or observations.
	Response format	Type of measurement scales used for data collection.
	Data analysis method(s)	Method(s) used for data analysis based on the classification presented in Figure 11.
AQS4: How these approaches have been developed?	Development methodology	Methodology used to develop the approach.
AQS5: How these approaches have been evaluated?	Evaluated factors	Factors used to evaluate the approach.
	Number of applications	Number of studies applying the approach.
	Data points	Number of data points collected during the applications used to evaluate the approach.
	Data analysis method(s)	Method(s) used for data analysis to evaluate the approach.

Source: developed by the author.

The articles were read thoroughly, and data were extracted by the author and reviewed by the author's advisor. Data extraction was hindered in several cases by the way in which the studies were reported. Most papers lack sufficient detail about the definition, development and validation of the evaluation approach. In some cases, the same approach was reported by more than one article. In these cases, we analysed each article in order to complement the information of the approach. A complete overview of the data extracted is available in Appendix A.

3.1.4 Analysis

In total, we identified 14 articles describing 10 approaches to evaluate educational games. Although we considered the last 23 years (1995-2018) in our review, we only encountered relevant publications after 2006. This shows that the interest in approaches to systematically evaluate educational games has been growing in the last years.

In order to present our findings, we analyse each of the analysis questions separately.

AQS1: Which models, methods, scales, or frameworks (approaches) exist to systematically evaluate educational games?

Analysing the selected studies, we identified 10 different approaches to systematically evaluate educational games. Five approaches present a framework (FREITAS; OLIVER, 2006; CONNOLLY; STANSFIELD; HAINEY, 2009; CARVALHO, 2012; CHEW, 2017; ABDELLATIF; MCCOLLUM; MCMULLAN, 2018), two approaches present a scale (FU; SU; YU, 2009; AK, 2012), one approach presents a generic methodology (MAYER, 2012) and two other approaches present a model (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011; GARCIA-MUNDO; GENERO; PIATTINI, 2015). We present a brief description of each one approach.

The Evaluation Framework for Effective Games-based Learning (GBL) (CONNOLLY; STANSFIELD; HAINEY, 2008; CONNOLLY; STANSFIELD; HAINEY, 2009; HAINEY; CONNOLLY; BOYLE, 2010) is a framework for GBL based on key measurements identified in the literature. The purpose of the framework is to identify what can potentially be evaluated in a GBL application. The approach proposes the evaluation of GBL with respect to learner performance, learner/academic motivation, learner/academic perceptions, learner/academic preferences, the GBL environment itself and the collaboration between players. The framework can be customized to particular requirements depending on particular analytical measurement is needed.

Another approach is the four-dimensional framework (FREITAS; OLIVER, 2006). This framework helps tutors to evaluate the potential of using games and simulation-based learning in their practice. The framework allows practitioners to be more critical about how they embed games and simulations into their lesson plans. It allows researchers and evaluators to develop metrics for supporting effective analysis of existing educational games and simulations and allows educational designers to consider a more user-based and specialized set of educationally specific factors. The four dimensions evaluated by the framework are: context, learner or learner group, internal representation world, and process of learning.

Carvalho (2012) presents an evaluation framework that assesses the efficiency of GBL focusing on engineering education. Covering the two first levels of Kirkpatrick's evaluation model (reaction and learning) (KIRKPATRICK; KIRKPATRICK, 2006), the framework is divided into three stages: alpha-testing, beta-testing and gamma-testing each with clear objectives, predefined protocols and data collection tools. The framework assesses the games' efficiency in terms of game play, game story, mechanisms, usability, knowledge, motivation, and satisfaction.

Chew (2017) presents a framework defining factors to consider in the design and analysis of game-based learning activities. The game factors considered in the framework are the cognitive engagement, behaviour engagement, emotional engagement, immersion, and challenge. In the design stage of the game-based learning activity, the authors adapted the game factors and applied the Design Thinking process to allow deeper considerations for the identified game factors. A questionnaire is developed based on the game factors to analyse the effectiveness of the learning activities.

Another framework was proposed by Abdellatif, Mccollum, and Mcmullan (2018), which aims to evaluate several dimensions of serious games by combining quality characteristics. It was designed to measure quality characteristics that do not require conducting an experiment and can be applied in a brief period. The framework includes characteristics that when absent will prevent serious games from delivering its educational content to a designated audience effectively, including usability, understandability, motivation, engagement, and user experience.

Fu, Su, and Yu (2009) present the EGameFlow, a scale that assesses user enjoyment of e-learning games to help developers to understand strengths and weaknesses from the students' perception in accordance to evaluation level 1 (reaction) (KIRKPATRICK; KIRKPATRICK, 2006). It evaluates the game's quality with respect to eight factors: immersion, social interaction, challenge, goal clarity, feedback, concentration, control, and knowledge improvement.

Another scale was proposed by Ak (2012). This scale aims at the selection of good educational computer games. The scale is intended to measure the quality of games before applying it in class. Game quality is measured in terms of enjoyment and learning.

A comprehensive methodology for the research and evaluation of serious games was proposed by (MAYER, 2012). This generic evaluation methodology for serious gaming, consists of a framework, conceptual models, research designs, evaluation constructs and scales, and data collection techniques. The methodology assesses serious games in three different moments (pre-game, in-game, and post-game) in terms of previous experiences/skills, game performance, game play, game experience, player satisfaction, and learning.

A model was proposed by Savi, Gresse von Wangenheim, and Borgatto (2011). The MEEGA model (Model for the Evaluation of Educational Games) is specifically developed for the evaluation of educational games for software engineering education. The model focuses on evaluation level 1 (reaction) (KIRKPATRICK; KIRKPATRICK, 2006), capturing the reaction of students after they played the game by applying a standardized questionnaire. MEEGA measures three quality factors of educational games: motivation, user experience, and learning.

Another model was proposed by Garcia-Mundo, Genero, and Piattini (2015). The QSGame-Model is a product quality model specific to serious games. The model is a customization of the ISO/IEC 25010 standard, changing and adapting its definitions and sub-characteristics in the context of serious games. The customization is mainly in terms of the sub-characteristics of usability and functional suitability.

AQS2: Which quality and/or sub-quality factors are evaluated?

In order to answer this question, we analysed the quality and/or sub-quality factors evaluated by the identified approaches. In summary, we identified 64 different quality and/or sub-quality factors that have been used by the approaches to evaluate educational games. All approaches use more than one quality factor to evaluate games. A complete list of the quality and/or sub-quality factors identified in the selected studies is available in Appendix A.

The most frequently used factors are learning (6 approaches), usability (4 approaches), social interaction (4 approaches), challenge (4 approaches), and immersion (4 approaches). Typically, evaluation of learning refers to the improvement of competence. Connolly et al. (2009) defines learning as an improvement in the performance of the learner as a result of the intervention. Other approaches evaluate learning/knowledge improvement based on students’ perceptions (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). Usability is defined in terms of awareness of progress, consistence of interface (colours, fonts), controls and visual feedback (CARVALHO, 2012). Only one approach defines usability based on a standard ISO/IEC 25010 (GARCIA-MUNDO; GENERO; PIATTINI, 2015). Social interaction refers to the creation of a feeling of shared environment and being connected with others in activities of cooperation or competition (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). Challenge means that a game needs to be sufficiently challenging with respect to the player's competency level. The increase of difficulty should occur at an appropriate pace accompanying the learning curve. New obstacles and situations should be presented throughout the game to minimize fatigue and to keep the students interested (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011; CHEW, 2017). Immersion allows the player to have an experience of deep involvement within the game, creating a challenge with real-world focus, so that s/he forgets about the outside world during gameplay (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011; CHEW, 2017).

AQS3: How data collection and analysis is operationalized?

In order to answer this question, we analysed how the approaches operationalize the evaluation, including research designs, data collection instruments, and data analysis methods.

Analysing the research designs, we classified the approaches in accordance to common study types as presented in Table 9. Four approaches (FU; SU; YU, 2009; CARVALHO, 2012; CHEW, 2017; ABDELLATIF; MCCOLLUM; MCMULLAN, 2018) provide an evaluation approach to be conducted in an ad-hoc manner, not clearly indicating the research design adopted. Only one approach (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011) proposes the conduction of the evaluation in form of a case study (non-experimental). The process defined by MEEGA explicitly defines the evaluation objective and provides a standardized questionnaire based on the defined model to be applied after the treatment (educational game) to collect data on the learners' perception. The approach proposed by (MAYER, 2012) defines a quasi-experimental design, similar to the experimental design, but without a random allocation of learners to the experimental or control group.

No information on the operationalization of the evaluation was given by (AK, 2012; CONNOLLY; STANSFIELD; HAINEY, 2009;

FREITAS; OLIVER, 2006; GARCIA-MUNDO; GENERO; PIATTINI, 2015).

Analysing the kind of data collection instruments, we identified that most of the approaches collect data via questionnaires (7 approaches), but only two approaches have systematically developed and statistically evaluated (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). Carvalho (2012) also used a questionnaire as a data collection instrument, but not provide information about its validity. Analysing the response format of these scales we identified that the Likert scale is the most used one (4 approaches), typically, representing the lowest and the highest degree to which respondents agree with the items. In addition, an ordinal scale also is used (3 approaches) to measure specific characteristics. Other data collection methods used include semi-structured interviews (CARVALHO, 2012) and tests in order assess the knowledge of the students (CARVALHO, 2012).

Analysing the data analysis methods of the selected studies, only two approaches (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011) provide information about which methods are used to analyse the data collected. Savi, Gresse von Wangenheim, and Borgatto (2011) use descriptive statistical methods (median/mode) and graphical visualization techniques such as histogram and frequency diagrams. Fu, Su, and Yu, (2009) also use descriptive statistics methods such as mean, standard deviation, and Pearson correlation coefficient to examine the dependency between variables. In addition, this approach also includes hypothesis testing in order to reject (or accept) a hypothesis with respect to a quality factor of the game. The t-test is used to compare two sample means, in a one factor-two treatments design and the ANOVA is used to evaluate the discrepancy in the level of psychological enjoyment between subjects (FU; SU; YU, 2009).

AQS4: How these approaches have been developed?

Analysing the selected studies, we identified that most of the approaches (7) do not report a systematic methodology to develop the approach. In general, the approaches seem to be developed in an ad-hoc manner (CARVALHO, 2012) or only based on theoretical constructs (AK, 2012; MAYER, 2012; CONNOLLY; STANSFIELD; HAINEY, 200), but not providing an explicit definition of the objective, measures or data collection instruments. One approach (QSGame-Model) (GARCIA-MUNDO; GENERO; PIATTINI, 2015) was developed by adopting a top-down approach methodology (FRACH; CARVALLO, 2003), used to adapt quality models to a specific domain, customizing general characteristics to a specific context.

On the other hand, two approaches report a systematic methodology for their development (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). MEEGA and EGameFlow follow the Scale Development Guide (DEVELLIS, 2003) to systematically develop a measurement instrument. In addition, MEEGA has been developed by using the GQM (Goal/Question/Metric) approach (BASILI; CALDIERA; ROMBACH, 1994) to explicitly define a measurement program for evaluating three quality factors of educational games: motivation, user experience and learning based on theoretical constructs.

AQS5: How these approaches have been evaluated?

In order to answer this question, we analysed the factors used to evaluate the approaches. We identified that most of the approaches (8) do not explicitly define criteria. Typically, the approaches (6) are proposed and partially evaluated through some case or pilot studies, applying the approach to evaluate an educational game in class (CARVALHO, 2012; MAYER, 2012; CONNOLLY; STANSFIELD; HAINEY, 2009; FREITAS; OLIVER, 2006). No information with respect to its evaluation was encountered for the approach proposed by Ak (2012) and Garcia-Mundo, Genero, and Piattini (2015).

On the other hand, two approaches present a systematic evaluation (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). MEEGA has been evaluated in terms of its applicability, utility, validity and reliability through three case studies in two different courses using three educational games (GRESSE VON WANGENHEIM; BORGATTO, 2011). A total of 79 data points was collected and analysed with respect to (DEVELLIS, 2003): correlation of items, item-total correlation, variance, mean, and Cronbach’s alpha coefficient.

EGameFlow has been evaluated in terms of its item analysis, reliability and validity through 4 game sessions in the same course, using different e-learning games (FU; SU; YU, 2009). A total of 166 data points was collected and analysed using the following tests: mean, standard deviation, extreme group comparison, test for homogeneity, t-test, ANOVA, Pearson’s correlation, and Cronbach’s alpha correlation.

3.1.5 Discussion

Analysing the selected approaches (AQS1), we identified that most of the studies propose a framework (5 approaches) to systematically evaluate educational games (FREITAS; OLIVER, 2006; CONNOLLY; STANSFIELD; HAINEY, 2009; CARVALHO, 2012; CHEW, 2017; ABDELLATIF; MCCOLLUM; MCMULLAN, 2018). Typically, the frameworks define a set of criteria ranging from the pedagogical perspective to gaming perspective, including context, environment, learner specifications, preferences, game play, user experience, etc. (FREITAS; OLIVER, 2006; CONNOLLY; STANSFIELD; HAINEY, 2009; CARVALHO, 2012). These criteria are used to guide and to help support instructors/tutors to evaluate educational games in a particular learning context and knowledge area (FREITAS; OLIVER, 2006). Thus, these frameworks are considered a flexible and easy to use approach, with the ability to help practitioners to reflect upon learning processes and approaches (FREITAS; OLIVER, 2006). However, the frameworks itself do not provide guidance on how to conduct the evaluation, data collection and analysis.

In this regard, the works presented by (FU; SU; YU, 2009; AK, 2012) propose scales providing effective instruments to systematically measure the quality of the games (FU; SU; YU, 2009). However, only the EGameFlow scale (FU; SU; YU, 2009) has been evaluated analysing its validity and reliability as an instrument to evaluate the level of enjoyment provided by e-learning games to their users (FU; SU; YU, 2009). On the other hand, no evaluation of the scale proposed by Ak (2012) has been encountered, thus, leaving its validity and reliability questionable (KITCHENHAM; PFLEEGER; FENTON, 1995; KIMBERLIN; WINTERSTEIN, 2008).

As the most comprehensive support, Mayer (2012) proposes a generic evaluation methodology for serious games. But, although the methodology provides a comprehensive support, including a framework, conceptual models, research designs, evaluation constructs and scales, and data gathering techniques, no information on the applicability and validity of this method have been encountered. On the other hand, the MEEGA (GRESSE VON WANGENHEIM; BORGATTO, 2011) provides an evaluation model that has been systematically developed by using the GQM approach to explicitly define a measurement program. This model has been evaluated in terms of its applicability, usefulness, validity and reliability through a series of case studies. Currently, MEEGA seems to be widely used in practice been reported by several studies from different authors evaluating different games and contexts (CALDERÓN; RUIZ, 2015).

Analysing the quality factors used to evaluate educational games (AQS2), we observed that, there exist a large diversity of factors. However, the learning/knowledge improvement is in fact the most evaluated factor as expected as the main objective of educational games is to potentiate the students' learning. Learning is often evaluated by comparing the competence level after game playing with the competence level beforehand, typically based on a pre/post-test score (MAYER, 2012) or through a self-assessment after game play (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). Besides learning, most approaches also consider several other quality factors, such as challenge, competence, social interaction, fun, usability, etc. also confirming the findings of Calderón and Ruiz, (2015) with respect to educational games in diverse knowledge areas. These factors are evaluated as they are considered important in order to promote a deeper and active learning.

In general, we also observed a lack of methodological support provided in order to operationalize the data collection and analysis (AQS3). Only two approaches (EGameFlow and MEEGA) provide an explicit definition of the data collection instruments, and data analysis methods. In the case of MEEGA, it was the only selected approach that has been developed considering the requirements specifically for the evaluation of games in the context of computing/SE education.

Only one approach (MAYER, 2012) proposes the usage of a more rigorous quasi-experimental research design as a best practice to assess game-based learning (ALL; CASTELLAR; LOOY, 2016). One reason for the lack of further experimental research designs may be the effort required to conduct experiments by not only collecting data after the treatment but also before its application. This may cause a major disruption in the flow of the course and not well accepted by the learners themselves. A more viable alternative in practice may be the conduction of case studies, a non-experimental method, typically using a one-shot post-test only design, as proposed by MEEGA (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). Adopting this research strategy, the evaluation goal is assessed based on the students' perceptions through a standardized questionnaire after the game's application.

However, as a result of our review, we identified as a significant weakness the way how data collection instruments (typically questionnaires) are developed in an ad-hoc manner (AQS4). Yet, in order to obtain valid results, it is imperative to systematically define and operationalize the measures and data collection instruments (KITCHENHAM; PFLEEGER; FENTON, 1995; KIMBERLIN;

WINTERSTEIN, 2008). Only two approaches (EGameFlow and MEEGA) propose systematically developed and evaluated questionnaires. Both, MEEGA and EGameFlow were developed adopting the scale-development guide proposed by DeVellis (2003). MEEGA and EGameFlow are also the only studies (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011) that explicitly report a systematic evaluation (AQS5). The criteria used for validation are defined based on scale-development theory (DEVELLIS, 2003) including applicability, utility, validity, and reliability. The other approaches selected in our review seem to have been evaluated through case or pilot studies only not validating the models/data collection instruments (FREITAS; OLIVER, 2006; CONNOLLY; STANSFIELD; HAINEY, 2009; CARVALHO, 2012; MAYER, 2012).

With respect to data analysis methods, most approaches also do not provide support. Again, only MEEGA and EGameFlow (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011) provide explicit support on how to analyse the collected data. These approaches, typically, use quantitative methods, including descriptive statistics to measure central tendency, dispersion, and measures of dependency. To assist in the understanding results, graphical visualization techniques are used. In addition, EGameFlow also proposes the usage of hypothesis testing to compare two sample means (FU; SU; YU, 2009).

In summary, analysing the state-of-the-art, we observe a lack of systematic, valid, and reliable approaches used for the evaluation of games that cover both the learning assessment and the evaluation of important aspects to provide a positive and engaging player experience. Therefore, based on these results, it becomes obvious that there exists a need for the identification of more consistent and uniform patterns to systematically evaluate educational games in order to obtain valid results that can be used to as a basis for decision on the application of such games and/or their continuous improvement.

3.2 STATE OF THE PRACTICE

In order to elicit the state of the practice on how educational games for computing education are evaluated, we conducted a systematic mapping study, following the steps defined in Figure 12.

3.2.1 Definition

This research aims to elicit the state of the practice on how games for computing education are evaluated. In accordance with this objective, we are focusing on the following analysis questions (AQ), grouped in alignment with the basic phases of an evaluation process based on Wohlin et al. (2012):

Evaluation Scoping/Planning
AQ1: Which analysis factors are evaluated?
AQ2: Which study types/research designs are used in the evaluations?
AQ3: Which evaluation models/methods are used?
AQ4: Which data collection instruments are used?
Operation
AQ5: What are the sample sizes of the reported evaluations?
AQ6: Are the evaluations replicated?
Analysis & Interpretation
AQ7: Which data analysis methods are used?

Inclusion/Exclusion Criteria. We included only articles that presented the evaluation of an educational game (digital or non-digital) for teaching any aspect of computing in a higher education context. We focused on articles written in English (or in Portuguese with an abstract in English), available via digital libraries published between January 1995 and May 2015). We excluded any study not related to computing education or to a higher education context. We also excluded articles that present an educational game or its development, but do not provide information about its evaluation. We also assessed the quality of the reported evaluations, considering only articles that provided information on how the games' evaluation was carried out, and including information on the definition, execution and analysis of the evaluation.

Data Sources and Search String. Data sources have been chosen based on their relevance to the computing domain, including: ACM Digital Library, IEEE Xplore, SpringerLink, ScienceDirect and Wiley Online Library. In addition, we searched via Google Scholar, in order to consider relevant articles published outside the computing domain.

In accordance with our research objective, we defined the search string by identifying core concepts considering also synonyms as indicated in Table 17.

Table 17 - Keywords

Core Concepts	Synonyms
educational game	serious game, game-based learning
evaluation	assessment, validation
computing	computer science, software
education	teaching, instruction

Source: developed by the author.

Using these keywords, the search string was calibrated and customized in conformance with the specific syntax for each of the data sources as presented in Table 18.

Table 18 - Search strings

Data source	Search String
ACM Digital Library	("serious game" OR "educational game" OR "game-based learning") AND (computing OR "computer science" OR software) AND (teaching OR education OR instruction) AND (evaluation OR validation OR assessment) Published since January 1995
IEEE Xplore	(("serious game" OR "educational game" OR "game-based learning") AND (computing OR "computer science" OR software) AND (teaching OR education OR instruction) AND (evaluation OR validation OR assessment) IN metadata) AND (pyr >= 1995 AND pyr <= 2015).
Springer Link	'("serious game" OR "educational game" OR "game-based learning") AND (computing OR "computer science" OR software) AND (teaching OR education OR instruction) AND (evaluation OR validation OR assessment)' published between 1995 - 2015
ScienceDirect	pub-date > 1994 and (("serious game" OR "educational game" OR "game-based learning") AND (computing OR "computer science" OR software) AND (teaching OR education OR instruction) AND (evaluation OR validation OR assessment))
Wiley Online Library	("serious game" OR "educational game" OR "game-based learning") AND (computing OR "computer science" OR software) AND (teaching OR education OR instruction) AND (evaluation OR validation OR assessment) in All Fields between years 1995 and 2015
Google Scholar	("serious game" OR "educational game" OR "game-based learning") AND (computing OR "computer science" OR software) AND (teaching OR education OR instruction) AND (evaluation OR validation OR assessment) Custom range: 1995-2015

Source: developed by the author.

3.2.2 Execution

Table 19 summarizes the returned results per data source. From Google Scholar we selected only the 1,000 most relevant results (100 first pages), from ACM Digital Library the 600 most relevant results and from

Science Direct, we analysed only the 500 most relevant results, observing a lack of relevancy beyond these quantities. From IEEExplore, SpringerLink and Wiley online library all returned articles were analysed. As a result, a total of 3,617 articles were analysed during the first stage.

Table 19 - Selection of primary studies

	Google Scholar	ACM	IEEExplore	SpringerLink	Elsevier	Wiley	Total
Total analyzed during 1st stage	1,000	600	458	526	500	533	**3,617**
Selected after 1st stage	134	11	20	2	17	10	**194**
Selected after 2nd stage	45	3	5	1	3	0	**57**

Source: developed by the author.

During the first stage, the search results were quickly analysed based on their title and short summary. The abstract was read only in case the title did not provide evidence of any exclusion criteria. Irrelevant and duplicate papers were removed. This stage left us with 194 potentially relevant articles. Then, we performed a second stage of selection. In this stage, we analysed the full abstract of the articles and quickly scoped the article for information on the game's evaluation. Articles without information on evaluation were excluded. As a result, 57 articles (describing a total of 58 educational games) were identified as primary studies.

As a result of our search, we also encountered six literature reviews (GRESSE VON WANGENHEIM; SHULL, 2009; GRESSE VON WANGENHEIM; KOCHANSKI; SAVI, 2009; CAULFIELD et al., 2011; IBRAHIM et al., 2011; CONNOLLY et al., 2012; BACKLUND; HENDRIX, 2013), citing other studies not yet identified as part of the encountered primary studies. We, therefore, also analysed the original studies cited in these reviews in accordance with our inclusion/exclusion criteria and included relevant studies as secondary studies. In addition, we also took into consideration another literature review that provides a complete overview on educational games in computing education (BATTISTELLA; GRESSE VON WANGENHEIM, 2014). We, thus, added these relevant studies as secondary studies. Table 20 presents the number of secondary studies selected from each of the identified literature reviews. As a result, a total of 55 secondary studies (on 48 educational games) were included in our SLR.

Table 20 - Selection of secondary studies

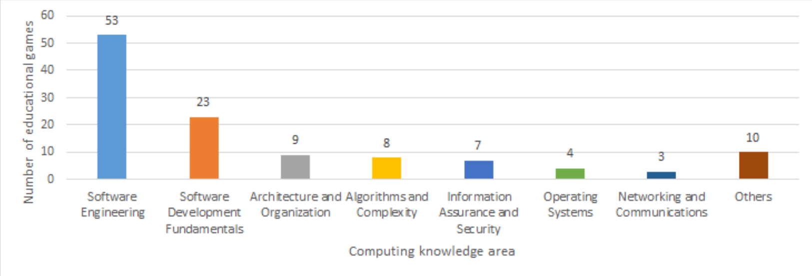
Reference	Number of studies cited in each literature review	Number of studies already selected as primary studies	Number of irrelevant studies	Number of secondary studies
(GRESSE VON WANGENHEIM; SHULL, 2009)	16	16	0	0
(GRESSE VON WANGENHEIM; KOCHANSKI; SAVI, 2009)	13	2	9	2
(CAULFIELD et al., 2011)	35	7	7	21
(IBRAHIM et al., 2011)	11	1	10	0
(CONNOLLY et al., 2012)	1	0	1	0
(BACKLUND; HENDRIX, 2013)	4	2	1	1
(BATTISTELLA; GRESSE VON WANGENHEIM, 2014)	76	24	21	31
Total	**156**	**52**	**49**	**55**

Source: developed by the author.

In total, we identified 112 articles describing 117 studies on the evaluation of educational games for computing education. The complete list and references of the selected articles can be found in the Appendix B.

In general, most of the educational games evaluated by the studies focus on Software Engineering (SE) education (Figure 13). This emphasis on SE games can be explained by the possibility they offer to provide practical experience for students in a safe and controlled environment (BATTISTELA; GRESSE VON WANGENHEIM, 2016). Other computing knowledge areas for which a considerable number of games have been evaluated are Software Development Fundamentals, Algorithms & Complexity, Architecture & Organization and Information Assurance & Security. Further computing knowledge areas seem to adopt fewer games for teaching as shown in Figure 13.

Figure 13 - Distribution of educational games per computing knowledge area



Source: developed by the author.

3.2.3 Data Synthesis and Extraction

In accordance with the defined analysis questions, we systematically extracted information in a spreadsheet from each article selected for analysis. Table 21 lists the information extracted from the selected studies.

Table 21 - Information extracted

General	
Game	Acronym or name of the educational game
Game description	Brief description of the educational game
Computing knowledge area	Computing knowledge area according to Computer Science Curricula (ACM; IEEE-CS, 2013) target by the game
Evaluation Scoping/Planning	
Analysis factor(s)	Analysis factor(s) that are evaluated
Model	Model/method used for evaluation classified in 3 categories: (a) using a specific well-defined model to evaluate educational games; (b) explicitly defining the evaluation (evaluation objective, measures and data collection instruments) applying a systematic approach or (c) ad-hoc in case no systematic definition of the evaluation is presented.
Study type	Study type classified in experimental, non-experimental, quasi-experimental or ad-hoc studies, following common research designs used in education contexts (based on Table 9).
Instrument	Instrument used for data collection, such as questionnaires, interviews, or observations.
Operation	

Sample size categories	Sample size of the reported evaluation(s) classified by the following categories: 1-20, 21-40, 41-60, 61-80, 81-100, 101-120 or >120 participants
Replication	Indicating if the evaluation study has been replicated describing the kind of replication of the evaluation, such as in several courses or different universities.
Analysis & Interpretation	
Data analysis method	Method(s) used for data analysis based on the classification presented in Figure 11. Any reported data analysis not using a defined qualitative or quantitative method is classified as informal.
References	
Reference	Reference of the study.

Source: developed by the author.

3.2.4 Analysis

To answer our analysis questions, we present our findings with respect to each of the analysis questions in alignment with the phases of the evaluation process.

Evaluation scoping/planning

AQ1: Which analysis factors are evaluated?

Analysing the selected studies, we identified a total of 43 different analysis factors that have been used to evaluate educational games for computing education.

Learning is the most frequently evaluated analysis factor (88 studies). This demonstrates that the main concern within these studies is the learning gain resulting from using educational games. Evaluations of this factor often refer to the improvement of competence by comparing the learners' competence level after game playing with their competence level beforehand, typically based on a post-test score (typically through experimental and/or quasi-experimental studies) or a self-assessment (typically through case studies). Few studies have evaluated the learning effect with regard to a systematic definition of learning levels, e.g. based on Bloom's taxonomy.

In general, we observed that in addition to evaluating the learning effect of the games, a wide variety of analysis factors are considered, including: motivation, user experience, usability, and instructional aspects. The factors, however, are defined inconsistently across the studies. The analysis factors described in each selected study have been mapped to a unified definition of the analysis factors. The complete mapping and definition of each analysis factor can be found in the Appendix B.

In summary, analysing the selected studies, we identified that motivation is usually evaluated in terms of confidence, relevance, satisfaction, attention, interest, and concentration. Besides those studies that analyse the concept of motivation in such a detailed way, two studies evaluate motivation in a high-level manner, through open-ended questions, not decomposing the concept into detailed factors.

Many studies also aimed to evaluate the user experience during game play. User experience covers the interaction of individuals with the game, considering thoughts, feelings, pleasure and other perceptions that result from the interaction (TULLIS; ALBERT, 2008). It is usually evaluated in terms of fun, challenge, enjoyability, immersion, social interaction, competence, control, recommendation, engagement, and realism. In this context, we also identified a considerable number of studies evaluating the usability of the games. We grouped the analysis factors with similar concepts of usability, based on ISO/IEC 25010 (ISO/IEC, 2011), in terms of ease of use, learnability, helpfulness, visually pleasing, efficiency, intuitive, and interaction.

Some studies also evaluate instructional aspects in terms of clear goal, sequence, adequacy, retrieval, responding, practical, and feedback. Further analysis factors evaluated include usefulness, correctness, sufficiency, structure, style, systems, strategy, shared values, staff, and skills.

Typically, most studies evaluate more than one analysis factor. The frequency with which the analysis factors were evaluated is shown in Figure 14. A detailed analysis of the identified factors can be found in the Appendix B.

Figure 14 - Frequencies of analysis factors used for the evaluation of games for computing education

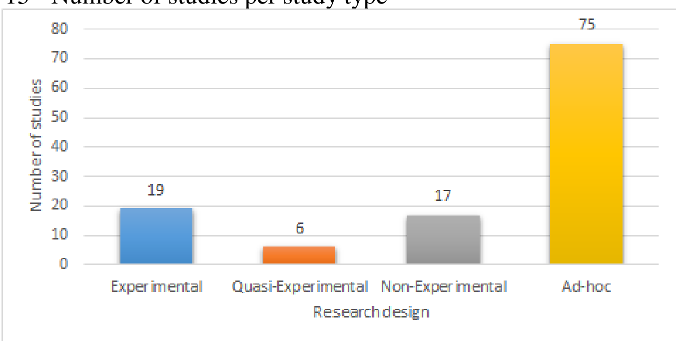
Source: developed by the author.

A considerable number of studies (38) did not define specific analysis factors to be evaluated, but rather evaluated the quality of the games as a generic factor, often using open-ended questions to obtain students' general perceptions after playing the game.

AQ2: Which study types/research designs are used in the evaluations?

In order to answer this question, we classified the research design of the selected studies as described in Section 2.5. The vast majority of the studies was conducted in an ad-hoc manner, analysing learners' informal comments after they played the game or describing some observations of pilot studies (Figure 15).

Figure 15 - Number of studies per study type



Source: developed by the author.

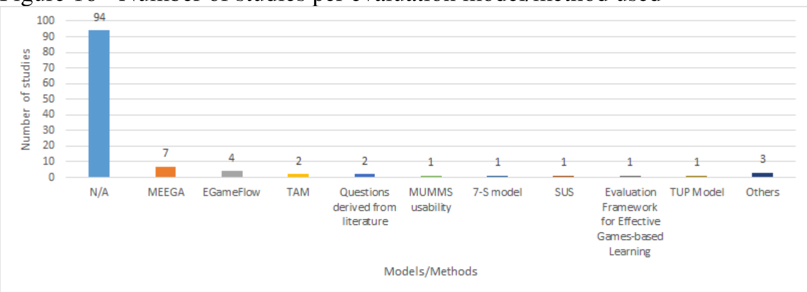
A case study was conducted in 17 studies. In these non-experimental studies, the evaluation was systematically defined, and after the treatment (educational game) data on the learners' perception had been collected, typically through questionnaires.

25 studies adopted a more rigorous research design. 19 studies used an experimental design with random assignment of participants to the experimental and control group. The experimental group typically used the educational game as a treatment and the control group used others instructional strategies such as lectures or readings. Data were typically collected by tests at two points in time, once before the treatment (pre-test) and once afterwards (post-test). Six studies applied a quasi-experimental approach, similar to the experimental designs, but without a random allocation of learners to the experimental or control group.

AQ3: Which evaluation models/methods are used?

As shown in Figure 16, more than 81% (94 studies), did not use any well-defined model or method to conduct the evaluation of the educational game. The evaluations are reported in an informal ad-hoc way, not providing an explicit definition of the evaluation objective, measures or data collection instruments.

Figure 16 - Number of studies per evaluation model/method used



Source: developed by the author.

On the other hand, some studies used well-defined models to evaluate educational games, such as MEEGA (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011), EGameFlow (FU; SU; YU, 2009) or the Evaluation Framework for Effective Games-based Learning (GBL) (CONNOLLY; STANSFIELD; HAINEY, 2009; HAINEY; CONNOLLY; BOYLE, 2010).

MEEGA is a model specifically developed for the evaluation of educational games in terms of motivation, user experience and learning.

Data are collected via a questionnaire completed by the learners after the game has been played. This model has been validated in terms of its applicability, usefulness, validity and reliability through case studies (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011).

EGameFlow is a scale that assesses user enjoyment of e-learning games in terms of: immersion, social interaction, challenge, goal clarity, feedback, concentration, control, and knowledge improvement. Data are collected via a questionnaire after the application of the game. EGameFlow has been validated in terms of its reliability and validity through four games sessions using 4 e-learning games (FU; SU; YU, 2009).

Connolly et al.'s (2009) framework is an approach for GBL evaluation in terms of learner performance, learner/academic motivation, learner/academic perceptions, learner/academic preferences, the GBL environment itself, and collaboration between players. The framework does not provide a defined measurement instrument. So far it has been partially evaluated through two studies (HAINEY et al., 2010; HAINEY et al., 2009), although only one of the studies is relevant to the specific focus of this review.

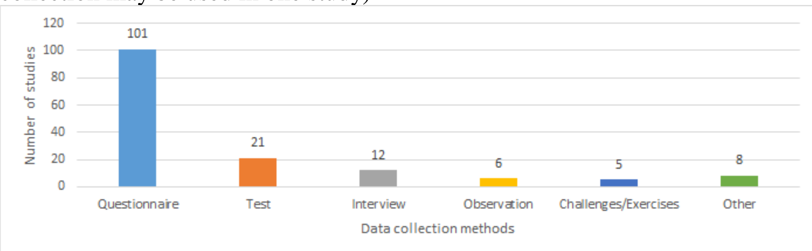
Other selected studies adapt existing evaluation models that have originally been designed for different purposes, such as the Technology Acceptance Model (TAM) (DAVIS; BAGOZZI; WARSHAW, 1989), a model for the evaluation of how users come to accept and use a new technology based on its perceived usefulness and perceived ease of use. The System Usability Scale (SUS) (BROOKE, 1996), which focuses on the subjective assessments of usability through a ten-item questionnaire, or MUMMS (Measuring the Usability of Multimedia Systems) (MUMMS, 2015), a model developed to evaluate the quality of use of new multimedia products/systems by the end user. We also observed the adoption of the 7-S model designed as an organizational analysis tool to assess and monitor changes in the internal situation of an organization with respect to structure, style, systems, strategy, shared values, staff, and skills, and the TUP (Technology, Usability and Pedagogy) model (BEDNARIK et al., 2004), which is a model designed for evaluating educational software in terms of technological aspects, usability aspects and pedagogical aspects. Other approaches used for evaluation include the Felder-Silverman learning style (GRAF et al., 2007), Bloom's Taxonomy (BLOOM, 1956), Gagne's Nine Events of Instruction (GAGNÉ; BRIGGS; WAGER, 1992), Keller's ARCS model (KELLER, 1987), or the Computational Thinking Framework (CTF) (GOUWS; BRADSHAW; WENTWORTH, 2013).

Two studies defined their own evaluation approach by deriving analysis questions from literature. Using the GQM approach (BASILI; CALDIERA; ROMBACH, 1994) both studies systematically derive analysis questions, measures and data collection instruments including questionnaires, observations and the examination of the students' work. The collected data were systematically analysed to answer the analysis questions in order to achieve the evaluation goals.

AQ4: Which data collection instruments are used?

The majority of the studies collects data using a questionnaire (101 studies) (Figure 17). Typically, these questionnaires are developed in an informal manner, without defining a measurement model in order to derive questionnaire items based on theoretical constructs.

Figure 17 - Number of studies per data collection method (more than one data collection may be used in one study)



Source: developed by the author.

Other studies use standardized questionnaires for data collection from well-defined evaluation models/scales, such as MEEGA (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011), EGameFlow (FU; SU; YU, 2009), TAM (DAVIS; BAGOZZI; WARSHAW, 1989) or SUS (BROOKE, 1996), which have been systematically developed and statistically validated.

As part of the questionnaires, 25 studies also included questions on demographic information about the population, such as age, gender, education level and, previous knowledge. In 12 studies, the learners were also interviewed after the game session to capture their perceptions either as the only data collection method (3 studies) or in combination with observations and/or questionnaires.

Only 21 studies reported using tests in order to assess the knowledge of the students. Typically, they were used in studies with an experimental research design, applying them before and after the treatment (13 studies). However, some studies applied tests only after the game session (8 studies). Fewer studies (5) also reported the use of challenges or exercises in order to collect data on the knowledge of the students.

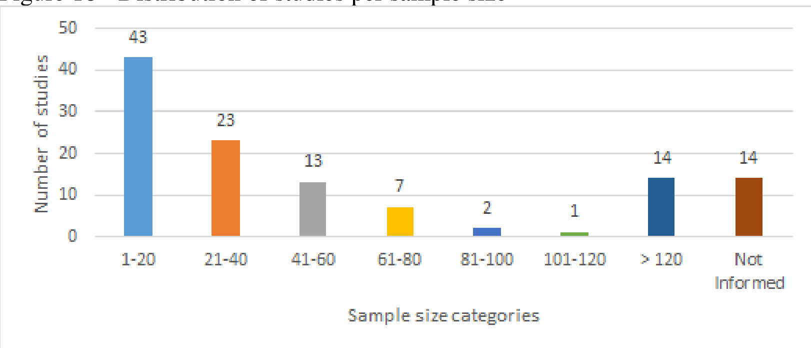
Less frequent types of data collection methods included the recording of verbal communications, or the analysis of learning diaries or students' scores.

Operation

AQ5: What are the sample sizes of the reported evaluations?

As shown in Figure 18, the majority of the evaluations was conducted with very small samples, ranging from 1 to 20 participants. This low number of participants typically corresponds to the size of one class in which the educational game is applied and evaluated. Little more than half of the studies (57%) were performed with less than 40 participants. Several studies (14) did not even report the sample size. However, 15 evaluations ran with more than 100 participants.

Figure 18 - Distribution of studies per sample size

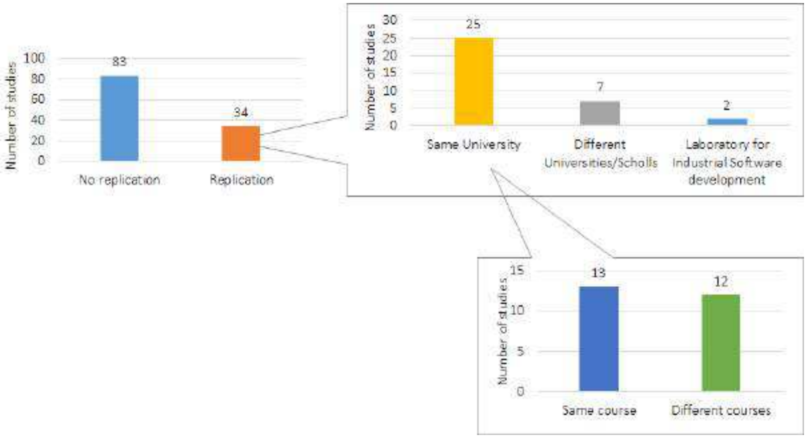


Source: developed by the author.

AQ6: Are the evaluations replicated?

Most of the reported evaluations (83 studies) have not been replicated (Figure 19) in order to increase their external validity. These games were evaluated only through one single study in one specific context, often by the game creators themselves.

Figure 19 - Number of studies per replication type



Source: developed by the author.

On the other hand, 34 studies did replicate the evaluation. Replications occurred in the same university (25 studies) or in different contexts (different laboratories (2 studies) or different universities (7 studies)). Replications at the same university typically occur in different courses (12 studies) or were repeated several times with other populations of the same course (13 studies). Very few longitudinal studies (4) over several years were reported.

Few studies (7) have analysed the data from replicated studies separately in order to compare the results. Normally, data have been analysed in an aggregate manner, to increase sample size and, thus, the validity of the results.

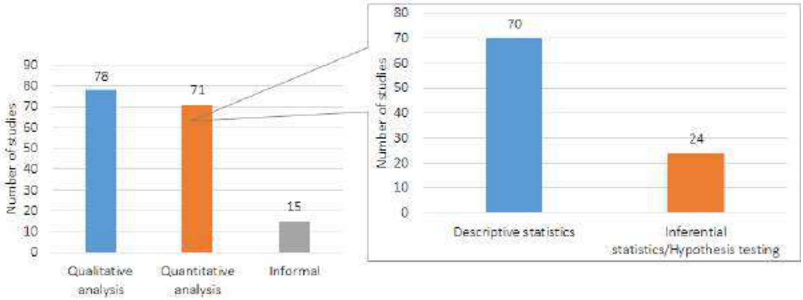
Analysis & Interpretation

AQ7: Which data analysis methods are used?

In most cases (78 studies) qualitative analysis methods were used for analysing comments and students' perceptions, typically collected through open-ended questions (Figure 20).

On the other hand, 71 studies also use quantitative analysis methods ranging from descriptive statistics (70 studies) to inferential statistics, as shown in Figure 20. Hypothesis testing was evident in 24 studies.

Figure 20 - Distribution per data analysis method

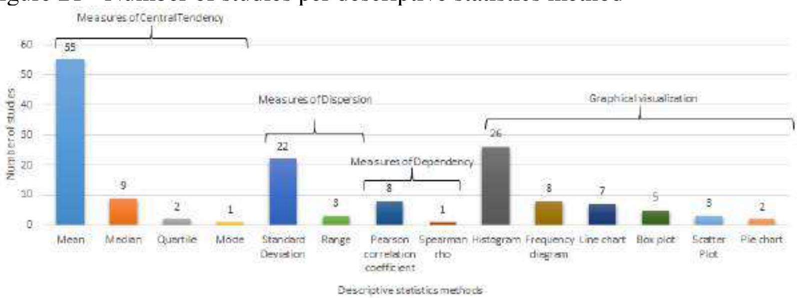


Source: developed by the author.

With respect to descriptive statistics methods (Figure 21), the mean is widely used (55 studies) as measures of central tendency. Other measures of central tendency include median (9 studies), quartile (2 studies) and mode (1 study). This is sometimes complemented by measures of dispersion to measure the degree of variation from the central tendency, including standard deviation (22 studies) and range (3 studies). Measures were occasionally used to examine the dependency between variables, including Pearson correlation coefficient (8 studies) and Spearman rho (1 study).

Typically, measures of central tendency are combined with graphical visualization techniques such as histograms (26 studies), frequency diagrams (8 studies), line charts (7 studies), box plots (5 studies), scatter plots (2 studies) or pie charts (1 study).

Figure 21 - Number of studies per descriptive statistics method

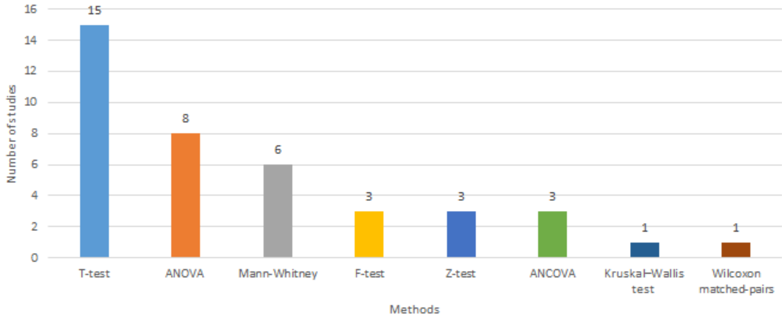


Source: developed by the author.

24 studies also used hypothesis testing in order to reject (or accept) a hypothesis with respect to an analysis factor (Figure 22). Here, the t-test

(15 studies) is the most often used parametric tests to compare two sample means, in a one factor-two treatments design. Six studies used Mann-Whitney, a non-parametric alternative to the t-test. In addition, the F-test, a parametric test, was used to compare two sample distributions in three studies. Z-test (3 studies) is used to refer specifically to the one-sample location test comparing the mean of a set of measurements to a given constant. Wilcoxon (1 study), another non-parametric alternative to the paired t-test, was used in a one factor-two treatments design.

Figure 22 - Distribution per hypothesis testing method



Source: developed by the author.

Studies that analysed one factor with more than two treatments, either used ANOVA (8 studies), Kruskal-Wallis as a non-parametric alternative (1 study) or ANCOVA (3 studies). These methods were used in order to identify significant differences in learning, knowledge levels, and/or rankings of aspects or perceptions among the individual groups (NAVARRO; VAN DER HOEK, 2009; HAINEY et al., 2011).

A considerable number of studies (15) carried out data analysis without using any formal quantitative or qualitative methods.

3.2.5 Discussion

In general, we encountered a considerable number of evaluations of games for computing education, especially for software engineering education.

With respect to the factors evaluated (AQ1), we observed that learning is the most evaluated one as the principal aim is to increase the students' learning. Learning is often evaluated by comparing the competence level after game playing with the competence level beforehand, based on a pre/post-test score or through a self-assessment after game play. In contrast, some studies did not define specific analysis factors to be evaluated. These evaluations typically used open-ended questions to capture students' general perceptions after game play.

Most studies did not only evaluate the learning itself, but also address several other factors such as motivation, user experience and usability, confirming the trend identified by Calderón and Ruiz (2015) with respect to educational games in different knowledge areas. These additional factors are typically evaluated as they are considered important in order to promote a deeper and active learning. Motivation is typically decomposed based on the ARCS model, which defines four factors to represent motivation in instructional design: attention, relevance, confidence and satisfaction, including also concentration and interest. However, user experience is measured quite differently by the studies. Factors that are evaluated with respect to user experience include: fun, social interaction, challenge, immersion, control, competence, recommendation, engagement, realism, and/or enjoyability. Several studies emphasized usability as an important factor rather than, for example, correctness or sufficiency that were evaluated only by two studies. Usability is typically evaluated in terms of ease of use, learnability, helpfulness, visually pleasing, efficiency, intuitive and interaction using standardized questionnaires such as SUS or MUMMS. From this discussion, it becomes clear that there still does not exist a consensus on the definition of analysis factors used to evaluate educational games.

Analysing the scientific rigor applied in the evaluations (AQ2), we observed that only a small number of studies (19 studies) apply an experimental research design. In this respect, we observe a significant weakness in the reported evaluations as most studies were conducted in an ad-hoc manner lacking scientific rigor and, thus, being questionable with respect to the validity of the results obtained. This confirms similar observations of Connolly et al. (2012), and Calderón and Ruiz (2015). Therefore, it seems to be imperative to formalize evaluation studies.

Consistent with the lack of rigorous research designs is the lack of the use of systematic models/methods for the evaluation of the games (AQ3). The majority of the studies (94) we encountered did not indicate the usage of an existing evaluation model/method. Nor did they report a systematic definition of the evaluation objective, measures or data collection instruments. In most studies, the data collection instrument (typically a questionnaire) seems to be developed in an ad-hoc manner (AQ4). Yet, in order to obtain valid results, it is imperative to systematically define and operationalize the measures and data collection instruments (KITCHENHAM; PFLEEGER; FENTON, 1995). Thus, there exists a large threat to the validity of these studies' results (KAZIMOGLU et al., 2012).

On the other hand, however, several studies (11) applied models specifically developed for the evaluation of educational games (MEEGA and EGameFlow). Both MEEGA and EGameFlow have been systematically developed and validated. Both models focus on the evaluation of learning/knowledge improvement and user experience during game play, in the case of MEEGA including the motivation promoted through the game. Currently, MEEGA seems to be widely used in practice as it is reported by several studies from different authors evaluating different games in different contexts, confirming the findings of Calderón and Ruiz (2015). EGameFlow seems to have been proposed by the authors and then discontinued.

Other studies report the usage of standardized questionnaires from other fields, focusing exclusively on very specific analysis factors, including usability questionnaires as well as on the awareness of organizational aspects. A clear limitation in these cases is the exclusive focus on only one specific analysis factor, and, in particular, not covering the learning analysis.

Regarding the significance of the results (AQ5), most evaluations are conducted with small samples, typically corresponding to the size of one class in which the educational game is applied. Consequently, these studies present low statistical power, having a reduced chance of detecting a true effect (FREEDMAN; PISANI; PURVES, 2007). This issue may further reduce the detection of significant results in experiments with the need of a control group (GRESSE VON WANGENHEIM; SHULL, 2009). Thus, if single applications are typically limited by the size of a class, replication of the studies in order to increase sample size becomes imperative. This also increases external validity (FENTON; PFLEEGER, 1998; WOHLIN et al., 2012). However, most of the games were evaluated only once (70%) without any replication either in the same or a different context (AQ6). The replications that have been reported generally aim at increasing the sample size through the repetition of the study by aggregating the data into one set for analysis. So far replications seem not to be used for comparisons between the games, contexts, etc., which could be a first step in the direction of creating guidelines on which game is beneficial in which context.

Analysing the data analysis methods used (AQ7), the majority of the studies performs qualitative data analysis with respect to students' demographic information, comments and perceptions, usually collected after game play. Several studies also use quantitative methods. These typically include descriptive statistics to measure central tendency, dispersion, and dependency. To assist in the understanding of the study results, graphical visualization techniques are often used. Few studies have used inferential statistics requiring an experimental research design. Only 24 studies used hypothesis testing with respect to differences in learning effectiveness between two groups.

In summary, the results of our state of the practice also confirm the findings of related reviews in diverse computing knowledge areas (HAYS, 2005, GRESSE VON WANGENHEIM; SHULL, 2009; CAULFIELD et al., 2011; CALDERÓN; RUIZ, 2015) that most evaluations of educational games for computing education are performed in an ad-hoc manner in terms of research design, measurement and data collection, and analysis, lacking scientific rigor. However, two approaches developed for the evaluation of educational games stand out: MEEGA and EGameFlow, being the MEEGA the only one model which has been exclusively developed for computing/software engineering education (CALDERÓN; RUIZ, 2015).

3.3 THREATS TO VALIDITY OF THE MAPPING STUDIES

As in any systematic review, some threats to the validity of the results exist. We, therefore, identified potential threats and applied mitigation strategies in order to minimize their impact in our mapping studies.

Publication bias. Systematic reviews suffer from the common bias that positive outcomes are more likely to be published than negative ones (KITCHENHAM, 2010). This was evident in this review, in which few papers present insignificant or negative results of the evaluation of the educational game. Nevertheless, we do not consider this a critical threat to our research as rather than focusing on the impact of these games, our aim was to determine how these games had been evaluated as well as to identify approaches used to systematically evaluate games.

Identification of studies. Another risk is the omission of relevant studies. In order to mitigate this risk, we carefully constructed the search string in order to be as inclusive as possible, considering not only core concepts but also synonyms. The risk of excluding relevant primary studies was further mitigated by the use of multiple databases that cover the majority of scientific publications in the field. Further mitigation was achieved through the inclusion of secondary studies that were identified based on other related SLRs in the analysis of the state of the practice.

Study selection and data extraction. Threats to study selection and data extraction were mitigated through providing a detailed definition of the inclusion/exclusion criteria. We defined and documented a rigid protocol for the study selection and both authors (author and his advisor) conducted the selection together, discussing the selection until consensus was achieved. Data extraction was hindered in several cases as many studies were not reported in alignment with common research frameworks. In these cases, information was inferred and discussed by the authors until consensus was achieved. No inter-rater reliability was conducted as the reviews have been carried out by two researchers in constant cooperation, and, thus, such a statistical analysis of the consistency would not provide significant results.

3.4 CHAPTER CONCLUSIONS

The results of the state-of-the-art and the practice indicate that there are only few approaches, which provide a systematic support, for game evaluations. Most of them are frameworks rather than comprehensive evaluation methods, indicating a lack of support on how to conduct such evaluations. In addition, most of the approaches also seem to be developed in a rather ad-hoc manner, not providing an explicit definition of the objective, measures or reliable and valid data collection instruments.

Consistent with this result, is the analysis of the state of the practice, indicating that most games used for computing/SE education are evaluated without explicitly defining an evaluation objective, research design, measurement program, data collection instruments, and data analysis methods (CALDERÓN; RUIZ, 2015; ALL; CASTELLAR; LOOY, 2016; BOYLE et al., 2016; KOSA al., 2016, PETRI; GRESSE VON WANGENHEIM, 2017; KORDAKI; GOUSIOU, 2017). And, often, data are collected only in the form of students' informal comments and/or through questionnaires developed in an ad-hoc manner. Therefore, this lack of scientific rigor leaves the reliability and validity of their results and, thus, the quality and/or effectiveness of such games questionable (CALDERÓN; RUIZ, 2015; ALL; CASTELLAR; LOOY, 2016; BOYLE et al., 2016; KOSA al., 2016, PETRI; GRESSE VON WANGENHEIM, 2017; TAHIR; WANGMAR, 2017).

In this context, two approaches for game evaluations stand out: MEEGA and EGameFlow, being MEEGA the only one evaluation approach that has been exclusively developed for computing/software engineering education. In addition, the MEEGA model is widely used in practice for game evaluations, been reported by several studies from different authors evaluating different games, confirming also the findings of Calderón and Ruiz (2015). Therefore, we identified the MEEGA model (SAVI, 2011) as a prominent approach to the evaluation of educational games. Thus, in this study, we adopted the initial version of the MEEGA model as the basis of our research. However, the reliability and validity evaluation of the initial version of the MEEGA model seems to be limited. The original MEEGA model was evaluated based on a sample of 79 students in the context of only three educational games (SAVI, 2011). Therefore, a question that arises is whether the initial version of the MEEGA model allows to evaluate educational games in a valid and reliable way. Thus, in order to answer this question, we conducted a large-scale evaluation of the initial version of the MEEGA model, presented in the next chapter.

4 A LARGE-SCALE EVALUATION OF THE INITIAL VERSION OF THE MEEGA MODEL

MEEGA is a model developed specifically for the evaluation of educational games (digital and non-digital ones) for computing/software engineering education, which measures the quality of the game through the reaction of the students after they have played the game through the application of a standardized data collection instrument (questionnaire) (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). The initial version of the MEEGA model has been developed using the GQM approach (BASILI; CALDIERA; ROMBACH, 1994) to explicitly define a measurement program to evaluate the quality of games in three factors: motivation, user experience and learning. The quality factor motivation was decomposed into four dimensions: attention, relevance, confidence and satisfaction. The quality factor user experience was decomposed in terms of immersion, challenge, social interaction, fun, competence, and control. The quality factor learning is measured in relation to the first three levels of the revised version of Bloom's taxonomy (remembering, understanding and applying), including two dimensions with respect to short-term and long-term learning (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). Based on the defined quality factors/dimensions a standardized questionnaire was designed in order to operationalize the data collection. The questionnaire adopts a 5-point Likert scale with response alternatives ranging from strongly disagree to strongly agree.

The results of the state-of-the-art and practice show that the initial version of the MEEGA model seems to be an adequate model for the evaluation of games, as well as to being identified as widely used in practice. However, the evaluation of the initial version of the MEEGA model, when it was proposed, was limited, addressing only 3 educational games and a sample of only 79 students (SAVI, 2011).

Therefore, a question that arises is whether the initial version of the MEEGA model allows to evaluate educational games in a valid and reliable way. In this respect, this chapter aims to present a large-scale evaluation of the initial version of the MEEGA model in terms of reliability and validity.

4.1 DEFINITION AND EXECUTION OF THE STUDY

Following the GQM approach (BASILI; CALDIERA; ROMBACH, 1994) the objective of the study was defined: to analyse the initial version of the MEEGA model in order to evaluate its reliability and construct validity from the viewpoint of the researchers in the context of higher computing education and professional Information Technology (IT) training. Thus, in this study the results are interpreted from the researchers' perspective, being the researchers of the Software Quality Group (GQS/INCoD/INE/UFSC), with backgrounds in computing and statistics.

From this objective, we derive the following analysis questions to be analysed based on the evaluation of measurement instruments (CARMINES; ZELLER, 1982; DEVELLIS, 2003; TROCHIM; DONNELLY, 2008):

Reliability

AQ1: Is there evidence for internal consistency of the MEEGA measurement instrument?

Construct Validity

AQ2: Is there evidence of convergent and discriminant validity of the MEEGA measurement instrument?

AQ3: How do underlying factors influence the responses on the items of the MEEGA measurement instrument?

In order to answer these defined analysis questions, we collected data from case studies that evaluated educational games (digital or non-digital) in computing courses in higher education and professional IT trainings using the initial version of the MEEGA model. We identified potential case studies by searching via Google and Google Scholar for articles that reported the usage of the MEEGA model for game evaluation, citing one of the original papers of the model (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011; SAVI; GRESSE VON WANGENHEIM; ULBRICHT; VANZIN, 2010). Then, we contact (via e-mail) the authors requesting the collected data.

As a result, we obtained data from 60 case studies, with 1000 students in 6 different contexts/institutions as summarized in Table 22.

Table 22 - Summary of case studies

Game	**Game type**	**Course/Semester**	**Institution/ Country**	**Sample size**
Computing knowledge area: Software Engineering				
Dealing with difficult people	Non-digital	Project planning and management/2013-1	UFSC/ Brazil	14
		Project management/2013-1		28
		Project planning and management/2015-2		23

DELIVER!	Non-digital	Project planning and management/2010-2	UFSC/ Brazil	15
		Project management/2010-2		13
DOJO	Digital	Project management/2013-1	UDESC/ Brazil	19
EAReqGame	Digital	Software Engineering/2014-2	UFSM/Brazil	14
Paper Tower	Non-digital	Project management/2013-1	UDESC/Brazil	4
PERT-CPM Game	Non-digital	Project management	UNISUL/Brazil	5
PizzaMia	Non-digital	Project management/2013-1	UDESC/Brazil	17
		Project management/2014-1		19
		Project management/2015-1		13
PMMaster	Non-digital	Project planning and management/2010-2	UFSC/Brazil	7
		Project management/2010-2		16
		Project planning and management/2012-1		21
		Project management/2012-1		33
		Project planning and management/2015-1		17
		Project planning and management/2015-2		12
PMQuiz	Digital	Project planning and management/2015-1	UFSC/Brazil	20
		Project management/2015-1		13
		Project planning and management/2015-2		18
		Project management/2015-2		20
Project Detective	Non-digital	Project planning and management/2011-2	UFSC/Brazil	18
		Project management/2011-2		31
		Project planning and management/2013-1		13
Risk Game	Non-digital	Project management/2013-1	UDESC/Brazil	15
Risk Management Game	Non-digital	Project planning and management/2015-2	UFSC/Brazil	18
Schedule and Risk Game	Non-digital	Project management/2014-1	UDESC/Brazil	5
SCRUM'ed	Digital	Project planning and management/2015-1	UFSC/Brazil	23
SCRUMIA	Non-digital	Project planning and management/2010-2	UFSC/Brazil	16
		Project management/2010-2		12
		Project planning and management/2011-1		15
		Project management/2011-1		30
		Project planning and management/2015-1		13
		Project planning and management/2015-2		18

		Agile Methods/2013-1	UDESC/Brazil	23
SCRUM-SCAPE	Digital	Project planning and management/2013-2	UFSC/Brazil	17
ThatPMGame	Digital	Project management/2013-1	UDESC/Brazil	6
		Project management/2013-1		13
TRIVIAL PURSUIT – IFPUG FPA	Non-digital	Training course on IFPUG FPA v4.2	Engineering.IT/ Italy	14
		Training course on IFPUG FPA v4.2		5
XPEnigma	Non-digital	Project management/2013-1	UDESC/Brazil	20
Computing knowledge area: Human-Computer Interaction				
UsabiliCity	Digital	Human-Computer Interaction/2014	Uninorte/Brazil	37
Computing knowledge area: Algorithms and Complexity				
SORTIA (Heapsort)	Non-digital	Data Structure/2012-1	UFSC/Brazil	23
		Data Structure/2014-2		11
		Data Structure 2015-2		21
SORTIA (Heapsort)	Digital and Non-digital	Data Structure/2013-2	UFSC/Brazil	19
		Data Structure/2013-2		9
SORTIA (Quicksort)	Non-digital	Data Structure 2012-1	UFSC/Brazil	27
		Data Structure 2013-2		22
		Data Structure 2013-1		12
		Data Structure 2013-2		5
		Data Structures 2014-2		22
		Data Structures 2014-2		24
		Data Structures/2015-2		4
		Data Structures 2015-1		24
		Data Structures 2015-2		16
SORTIA	Digital	Data Structure 2012-2	UFSC/Brazil	18
		Data Structure 2014-2		17
		Data Structure 2015-2		3
			Total:	**1000**

Source: developed by the author.

Data collected in the selected case studies (Table 22) were pooled in a single sample, thus, used them cumulatively only in order to evaluate the initial version of the MEEGA model (and no a specific game). The pooling of data was possible due to the similarity of the selected case studies and standardization of the data collected. The similarity and standardization in terms of definitions, methods, and measurements are essential aspects for the pooling of data (KISH, 1994). In this respect, the selected studies are similar in terms of definition (with the objective to evaluate an educational game with respect to motivation, user experience and learning), research design (case studies), and context (higher education and professional training on computing/IT). In addition, all selected case studies are standardized in terms of measures (quality factors/dimensions), data collection method (MEEGA measurement instrument), and response format (5-point Likert scale).

4.2 ANALYSIS AND DISCUSSION

In order to answer the defined analysis questions, we perform a statistical evaluation. The evaluation is based on the approach for the construction of measurement instruments as proposed by DeVellis (2003) in alignment with procedures for the evaluation of internal consistency and construct validity of a measurement instrument (TROCHIM; DONNELLY, 2008). Data analysis was conducted using the IBM SPSS Statistics version 23.

Reliability

AQ1: Is there evidence for internal consistency of the MEEGA measurement instrument?

In order to answer this analysis question, we measured the internal consistency of the standardized items of the MEEGA measurement instrument through the Cronbach's alpha coefficient (DEVELLIS, 2016; TROCHIM; DONNELLY, 2008). Cronbach's alpha coefficient (CRONBACH, 1951) indicates indirectly the degree to which a set of items measures a single factor. Thus, here, we want to know whether the MEEGA measurement instrument measures the same quality factor, the reaction of students after they played and educational game. Typically, values of Cronbach's alpha between $0.8 > \alpha \geq 0.7$ are acceptable, between $0.9 > \alpha \geq 0.8$ are good, and $\alpha \geq 0.9$ are excellent (DEVELLIS, 2016), thus, indicating an internal consistency of the instrument.

Analysing the 29 standardized items of the measurement instrument of the initial version of the MEEGA model, the value of Cronbach's alpha for the quality factors of the MEEGA measurement instrument (standardized items) is satisfactory (α=.915). Results of a detailed analysis per quality factor (Table 23) show that Cronbach's alpha is also acceptable on the level of all three quality factors.

Table 23 - Cronbach’s alpha for standardized items per quality factor

Quality factor	Cronbach's alpha
Motivation	.791
User Experience	.875
Learning	.793

Source: developed by the author.

This indicates that there exists an acceptable internal consistency, not only in terms of the reaction of the students to the educational game, but that there exists also an internal consistency of the items related to each of the quality factors (motivation, user experience, and learning). Thus, we can conclude that responses between the items are consistent and precise, indicating the reliability of the standardized items of the MEEGA measurement instrument.

Construct Validity

AQ2: Is there evidence of convergent and discriminant validity of the MEEGA measurement instrument?

Construct validity of a measurement instrument refers to the ability to actually measure what it purports to measure (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008). Convergent and discriminant validity are the two subtypes of validity that make up construct validity (TROCHIM; DONNELLY, 2008). Convergent validity shows that the items that should be related are in reality related. On the other hand, discriminant validity shows that the items that should not be related are in reality not related (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008). In order to obtain evidence of convergent and discriminant validity of the standardized items of MEEGA measurement instrument, the correlations of the items are calculated (DEVELLIS, 2003).

In order to obtain evidence of convergent validity it is expected that the items of the same quality factor (e.g., motivation, user experience and learning) and same dimension (e.g., attention, satisfaction, fun, immersion, etc.) have a large correlation (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008). On the other hand, to obtain evidence of discriminant validity is expected that the items of different quality factors or dimensions should have a small correlation (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008). For example, it is expected that the items of different quality factors (e.g., motivation and user experience) have a small correlation, as in theory, the items are measuring different quality factors.

In order to analyse the correlations between the standardized items of the same quality factor, we used the nonparametric Spearman correlation matrices for each quality factor (Table 24 to 26). A complete matrix including all quality factors is presented in Appendix C. The matrices show the Spearman correlation coefficient, indicating the degree of correlation between two items (item pairs). We used this correlation coefficient as it is the most appropriate correlation analysis for Likert scales (CHEN; POPOVICH, 2002). The correlation coefficients between the items within of the same dimension are coloured. In accordance with Cohen (1988), a correlation between items is considered satisfactory, if the correlation coefficient is greater than 0.29, indicating that there is a medium or large correlation between the items. Satisfactory correlations are marked in bold. The numbers of the items relate to the specification presented in Table 24.

Table 24 - Spearman correlation coefficient of quality factor: Motivation

Item/ Dimension	**1**	**2 Attention**	**3**	**4**	**5 Relevance**	**6**	**7 Confidence**	**8**	**9 Satisfaction**	**10**
1	1.000									
2	**.387**	1.000								
3	**.375**	**.461**	1.000							
4	.228	.258	**.318**	1.000						
5	.289	.271	**.354**	**.360**	1.000					
6	.137	.231	.219	.289	.287	1.000				
7	.198	.164	.154	.177	.222	.276	1.000			
8	.205	.274	**.377**	.271	**.441**	.244	.282	1.000		
9	.255	**.304**	**.388**	**.400**	**.380**	**.301**	.203	**.478**	1.000	
10	.174	.253	.257	.229	.208	.247	.165	.255	**.352**	1.000

Source: developed by the author.

Analysing the correlations of the quality factor motivation (Table 24), we can observe that 15 item pairs are correlated. This indicates that the motivation is fragmented into its various dimensions more independently (attention, relevance, confidence and satisfaction). Yet, items that belong to the same dimension show an acceptable degree of correlation in almost all item pairs, as neither negative values have been detected nor a group of items that consistently did not present a correlation with respect to all evaluated games. We, therefore, can observe a tendency for the items of one dimension to be correlated, thus, indicating evidence of the convergent validity. On the other hand, some item pairs (e.g., 4-9, 5-8, 8-9), which do not belong to the same dimension also show an acceptable degree of correlation. Therefore, we cannot establish discriminant validity.

With respect to the quality factor user experience, 72 item pairs are correlated (Table 25). This also indicates that the user experience is fragmented in its individual dimensions (immersion, social interaction, challenge, fun, competence, and digital game). Again, items belonging to the same dimension show a satisfactory correlation with respect to all pairs of items. However, items 25 and 26 (control dimension) present negative correlation values, showing that these items do not have a correlation with other dimensions, and, therefore, seem not to be measuring user experience. Although, within the dimension (control), the items showed a satisfactory correlation. Thus, indicating that these items are measuring the characteristics of control, yet, they seem not to be related to the quality factor user experience.

Table 25 - Spearman correlation coefficient of quality factor: User Experience

	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
	Immersion			Social Interaction			Challenge		Fun				Competence		Control	
11	1.00															
12	**.677**	1.00														
13	**.624**	**.666**	1.00													
14	.257	.279	.269	1.00												
15	**.399**	**.419**	**.397**	**.635**	1.00											
16	**.333**	**.352**	**.312**	**.540**	**.557**	1.00										
17	**.304**	**.322**	**.314**	.238	**.301**	**.329**	1.00									
18	**.418**	**.450**	**.459**	.285	**.421**	**.345**	**.472**	1.00								
19	**.454**	**.523**	**.460**	**.339**	**.562**	**.415**	**.393**	**.574**	1.00							
20	**.386**	**.464**	**.467**	.158	**.330**	.214	**.325**	**.400**	**.469**	1.00						
21	**.380**	**.408**	**.411**	.200	**.339**	.271	**.395**	**.467**	**.520**	**.478**	1.00					
22	**.377**	**.457**	**.424**	.184	**.317**	.272	**.374**	**.433**	**.493**	**.504**	**.679**	1.00				
23	.262	.275	**.315**	.144	.175	.192	.272	**.312**	**.320**	.261	**.315**	**.313**	1.00			
24	**.331**	**.343**	**.353**	.185	**.291**	.276	**.377**	**.396**	**.452**	**.364**	**.431**	**.424**	**.489**	1.00		
25	-.064	-.039	-.052	-.183	-.083	-.086	-.121	-.145	-.034	.080	-.052	-.009	-.022	-.061	1.00	
26	-.067	-.039	-.060	-.167	-.075	-.059	-.101	-.144	-.011	.070	-.058	-.012	-.032	-.091	**.785**	1.00

Source: developed by the author.

In general, the items belonging to the same dimension showed a medium degree of correlation, thus, we can establish a convergent validity. However, several item pairs of different dimensions also showed a medium degree of correlation. Thus, again we cannot establish discriminant validity in the quality factor of user experience.

Analysing the correlations of the items of the quality factor learning, all three item pairs of the questionnaire are correlated (Table 26). This quality factor demonstrated a large correlation between items than the other two quality factors.

Table 26 - Spearman correlation coefficient of quality factor: Learning

	27	28	29
	Short-term Learning		Long-term Learning
27	1.00		
28	.636	1.00	
29	.490	.453	1.00

Source: developed by the author.

Summarizing, we can observe that in general, there exists a correlation between items within a single dimension with respect to all three quality factors. This indicates that, considering the dimensions, a convergent validity can be established. On the other hand, few item pairs present small or negative correlation, which indicate that these items need to be revised or re-grouped to other quality factors or dimensions. In general, items of different dimensions within a quality factor also present a medium or large correlation, thus, no discriminant validity could be established. This also can be observed when analysing the matrix that shows the correlations between items of all quality factors (Appendix C). The degree of correlation between the items shows whether the items measure (or not) the same quality factor/dimension, thus indicating evidence (or not) of convergent and discriminant validity. However, these results do not determine how many quality factors underlie the set of the MEEGA measurement instrument items. With this objective, we performed a factor analysis, answering our analysis question AQ3.

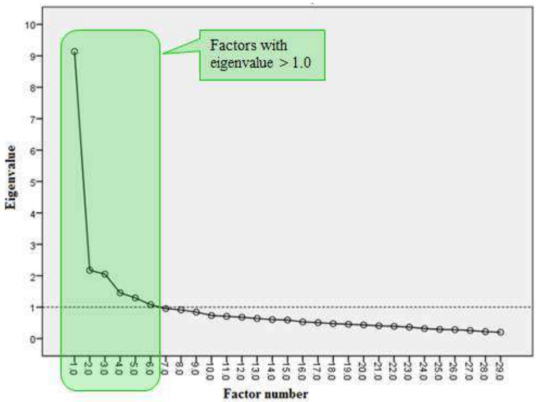
AQ3: How do underlying factors influence the responses on the items of the MEEGA measurement instrument?

In order to identify the number of factors (quality factors or dimensions) that represents the responses of the set of the 29 standardized items of the measurement instrument of the initial version of the MEEGA model, we performed a factor analysis. Based on the original definition of the MEEGA model we assume that it is influenced by three underlying factors (motivation, user experience and learning).

In order to analyse whether the items of the MEEGA measurement instrument can be submitted to factor analysis process (BROWN, 2006), we used the Kaiser-Meyer-Olkin (KMO) index and Bartlett's test of sphericity being the most commonly used ones (BROWN, 2006). These methods indicate how much the realization of the factor analysis is appropriate for a specific set of items (BROWN, 2006). The KMO index measures the sampling adequacy with values between 0 and 1. An index value near 1.0 supports a factor analysis and anything less than 0.5 is probably not amenable to useful factor analysis (DZIUBAN; SHIRKEY, 1974). Bartlett's sphericity test also indicates whether the factor analysis is appropriate with the values of a significance level < 0.05 are considered acceptable (DZIUBAN; SHIRKEY, 1974). Analysing the set of items of the MEEGA measurement instrument, we obtained a KMO index of .907 and a significance level of 0.000. Consequently, indicating that factor analysis is appropriate to analyse the number of factors that represents the responses of the MEEGA measurement instrument.

Applying the factorial analysis, the number of factors retained in the analysis is decided (GLORFELD, 1995; BROWN, 2006). Here we used the Kaiser-Guttman criterion for this decision. The Kaiser-Guttman criterion is a widely used method of determining the number of factors. This method states that the number of factors is equal to the number of eigenvalues greater than 1 (GLORFELD, 1995). The eigenvalue refers to the value of the variance of the all the items which is explained by a factor (GLORFELD, 1995). Following the Kaiser-Guttman criterion, our results show that 6 factors should be retained, explaining 59.29% of the data. The scree plot (Figure 23) shows the eigenvalue for each factor number (representing each item). The dotted line illustrates the Kaiser criteria (eigenvalues>1), justifying the retention of 6 factors. With respect to MEEGA, this means that the responses of the measurement instrument are representing six underlying concepts. Thus, the results of the factorial analysis indicate a different decomposition than the assumed one into three quality factors (motivation, user experience and learning) as proposed in the original MEEGA model.

Figure 23 - Scree Plot



Source: developed by the author.

Once identified the number of underlying factors, another issue is to determine which items are loaded into which factor. In order to identify the factor loadings of the items, a rotation method is used (BROWN, 2006; TABACHNICK; FIDEL, 2007). Here we used the Varimax with Kaiser Normalization rotation method being the most widely accepted and used rotation method (TABACHNICK; FIDEL, 2007). Table 27 shows the factor loadings of the items associated with the 6 retained factors. The highest factor loading of each item, indicating to which factor the item is most related, is marked in bold.

Table 27- Factor loadings of the MEEGA measurement instrument items

Factor/ Dimension		No	Description	Factor					
				1	2	3	4	5	6
Motivation	Attention	1	The game design is attractive.	-.171	**.857**	-.087	-.031	-.075	.084
		2	There was something interesting at the beginning of the game that captured my attention.	-.227	**.749**	.073	-.011	.062	-.109
		3	The variation (form, content or activities) helped me to keep attention to the game.	-.016	**.468**	.157	.069	.086	-.107
	Relevance	4	The game content is relevant to my interests.	.281	**.457**	-.259	.000	.116	.014
		5	The way the game works suits my way of learning.	**.452**	.270	-.140	-.068	.157	-.069
		6	The game content is connected to other knowledge I already had.	-.157	.238	-.117	-.092	**.646**	-.024
	Confidence	7	It was easy to understand the game and start using it as study material.	-.084	.283	-.213	.132	**.472**	.227
		8	Passing through the game, I felt confident that I was learning.	**.376**	.064	-.066	.011	.348	-.135

	Satisfaction	**9**	I am satisfied because I know I will have opportunities to use in practice things I learned playing this game.	**.426**	.183	-.151	.119	.264	-.003
		10	It is due to my personal effort that I manage to advance in the game.	-.025	-.302	.229	.055	**.776**	.061
User Experience	**Immersion**	**11**	Temporarily I forgot about my daily; I have been fully concentrated on the game.	-.063	-.057	**.861**	.025	.082	-.006
		12	I did not notice the time pass while playing; when I saw the game had already ended.	-.005	.035	**.847**	.022	.002	.022
		13	I felt myself more in the game context than real life, forgetting what was around me.	-.009	.034	**.813**	-.010	.059	.003
	Social Interaction	**14**	I was able to interact with others during the game.	-.010	-.100	-.050	**.925**	.014	-.082
		15	I had fun with other people.	.005	.074	.128	**.826**	-.074	.046
		16	The game promotes cooperation and/or competition among the players.	.060	.024	-.021	**.811**	-.018	.044
	Challenge	**17**	This game is appropriately challenging for me, the tasks are not too easy nor too difficult.	.083	**.317**	.101	.107	.122	-.103
		18	The game progresses at an adequate pace and does not become monotonous - offers new obstacles, situations or variations in its tasks.	-.017	**.434**	.318	.070	.015	-.152
	Fun	**19**	I had fun with the game.	-.017	**.549**	.236	.186	-.016	.057
		20	When interrupted at the end of the class, I was disappointed that the game was over.	.023	**.440**	.427	-.110	-.027	.133
		21	I would recommend this game to my colleagues.	.287	**.645**	.076	-.074	-.127	.033
		22	I would like to play this game again.	.318	**.550**	.187	-.051	-.190	.070
	Competence	**23**	I achieved the goals of the game applying my knowledge.	.109	-.178	.226	-.096	**.739**	.008
		24	I had positive feelings on the efficiency of this game.	.282	.190	.174	-.103	**.300**	-.021
	Control (digital games)	**25**	The controls to perform actions in the game responded well.	.054	-.018	.037	-.024	.039	**.928**
		26	It's easy to learn how to use the interface and game controls.	-.046	.042	-.002	.014	.057	**.930**
Learning	**Short-term learning**	**27**	The game contributed to my learning in this course.	**.976**	-.160	-.089	.050	-.032	.034
		28	The game was efficient for my learning, comparing it with other activities of the course.	**.985**	-.285	.081	.024	-.080	.028
	Long-term learning	**29**	The experience with the game will contribute to my professional performance in practice.	**.738**	.087	.086	-.024	-.143	-.005

Source: developed by the author.

Analysing the factor loadings of the items (Table 27), we can observe that, the first factor (factor 1), includes a set of 6 items (5, 8, 9, 27, 28 and 29). This result seems to suggest that these items are related to quality factor learning. Although the items 5, 8 and 9 are originally related to other three dimensions (relevance, confidence, and satisfaction), the description of these items also involves learning and, thus, seem to be justifying this relation.

With respect to factor 2, it includes a set of 10 items (1, 2, 3, 4, 17, 18, 19, 20, 21 and 22). These items refer to four different dimensions of the original questionnaire (attention, relevance, challenge, and fun), measuring two different quality factors (motivation and user experience). Similarly, factor 5, includes a set of 5 items (6, 7, 10, 23 and 24. Originally the items of factor 5 are related to four different dimensions (relevance, confidence, satisfaction, and competence), and two different quality factors (motivation and user experience). Thus, the results of factor 2 and 5 suggest that the original classification of MEEGA model may not be the most appropriate and needs to be revised. The allocation of the items in different factors, may be explained due to the free translation (to Brazilian Portuguese) and adaptation of the original items to the context of educational games. For example, the item ('The wording of feedback after the exercises, or of other comments in this lesson, helped me feel rewarded for my effort') from the original ARCS questionnaire (KELLER, 2009) is represented through item 10 in the MEEGA questionnaire ('It is due to my personal effort that I manage to advance in the game'). Another possible indication may in fact be the need for a re-grouping of these items as also pointed out by other empirical studies evaluating the standardized questionnaires used in the development of MEEGA (HUANG et al., 2006; JOHNSON, 2012).

Analysing the results of factor 3, we can observe that this factor is composed by a set of three items (11, 12, 13), all three of the same dimension (immersion) as proposed by the original MEEGA model. The same can be observed with respect to factor 4 being mainly composed by a set of three items (14, 15 and 16), all of which are referring to the original dimension of social interaction. Factor 6 also is composed of a set of two items (25, 26) both related to the dimension of digital game, as originally proposed in MEEGA model.

In summary, although the initial version of the MEEGA model presents an internal consistency, indicating its reliability, there are several limitations in terms of its validity. The results of the exploratory factor analysis, clearly, indicate the conceptual overlap between the factors of motivation and user experience.

4.2.1 Threats to validity

Due to the characteristics of this type of research (case study), this work is subject to various threats to validity. We, therefore, identified potential threats and applied mitigation strategies in order to minimize their impact on our research.

Construct validity. Some threats are related to the design of the study (WOHLIN et al., 2012). In order to mitigate this threat, we defined and documented a systematic methodology for our case study. In the definition of the case study, we used the GQM approach (BASILI; CALDIERA; ROMBACH, 1994) to systematically define the objective of this study and to decompose the objective into analysis questions. Another risk is related to the omission of existing data sets related to the evaluation of the MEEGA model. In order to mitigate this risk, we searched for existing evaluation studies using the MEEGA model via Google and Google Scholar representing broad search engines. We included data sets from all studies that we encountered and for which we received the collected data. Another risk refers to the quality of the data pooled into a single sample, in terms of standardization of data (response format) collected and adequacy to MEEGA model. As our study is limited exclusively to evaluations that used the MEEGA model the risk is minimized as in all studies the same data collection instrument has been used. Another issue refers to the pooled data from different contexts. To mitigate this threat, we selected studies considering only the context of higher education and professional IT training with respect to only one knowledge area: computing.

External validity. In terms of external validity, a threat to the possibility to generalize the results is related to the sample size and diversity of the data used for the evaluation. In respect to sample size, our evaluation used data collected from 60 case studies evaluating 24 different digital and non-digital games, involving a population of 1000 students. In terms of statistical significance, this is a satisfactory sample size, allowing the generation of significant results (SITZMANN et al., 2010). The data has been obtained from game applications in 6 different institutions/contexts. However, as the data collection was restricted to evaluation that used the MEEGA measurement instrument for data collection, the majority of the data came from Brazil, where it is used more prominently, with only one application from an organization in Italy.

Reliability. In terms of reliability, a threat refers to what extent the data and the analysis are dependent on the specific researchers. In order to mitigate this threat, we documented a systematic method, defining clearly the study objective, the process of data collection, and the statistics methods used for data analysis. Another issue refers to the correct choice of statistical tests for data analysis. To minimize this threat, we performed a statistical evaluation based on the approach for the construction of measurement scales as proposed by DeVellis (2016), which is aligned with procedures for the evaluation of internal consistency and construct validity of a measurement instrument (TROCHIM; DONNELLY, 2008).

4.3 CHAPTER CONCLUSIONS

This chapter presents a large-scale evaluation of the initial version of the MEEGA model. In summary, the results of the evaluation, analysing data from 60 case studies involving 1000 students, indicate a satisfactory reliability (Cronbach's alpha α=.915). However, in terms of validity, improvement opportunities have been identified. The results of its validity evaluation clearly indicate the need for restructuring the initial version of the MEEGA model, not only in terms of the number of quality factors, but also in relation to the grouping of items of the dimensions and quality factors. Especially, items related to the quality factors motivation and user experience need to be revised, as the results of the factor analysis indicate that they may not be related to a different quality factors as proposed in the original model.

Therefore, based on the results, these two quality factors (motivation and user experience) seem to overlap conceptually and, thus, need to be revised with respect to their wording and classification to a quality factor/dimension.

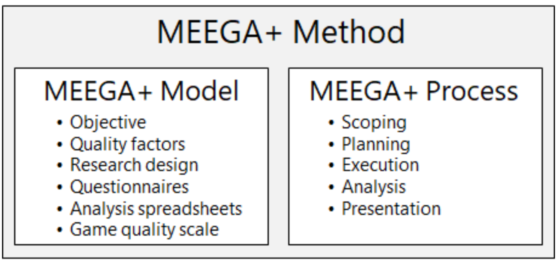
In this regard, the next chapter presents the MEEGA+ method, developed based on the results of this large-scale evaluation of the initial version of the MEEGA model and the results of the state-of-the-art and the practice.

5 MEEGA+: A METHOD FOR THE EVALUATION OF THE QUALITY OF GAMES FOR COMPUTING EDUCATION

This chapter presents the MEEGA+ method, which has been systematically developed for the evaluation of the quality of games used as an instructional strategy for computing/SE education, based on the results of the literature reviews (PETRI; GRESSE VON WANGENHEIM, 2016; PETRI; GRESSE VON WANGENHEIM, 2017) and the large-scale analysis of the initial version of the MEEGA model (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017).

In this study, we understand a method as a systematic approach to achieve a certain objective or result and, which describes the characteristics of an ordered process or a procedure used in the engineering of a product (IEEE, 2002; IEEE 2010). Based on this definition, the MEEGA+ method aims to provide a systematic support for the evaluation of games for computing education. It is composed of an evaluation model (MEEGA+ Model) (section 5.1) 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) (section 5.2) defining phases, activities and work products, guiding researchers on how to plan, execute and analyse the results of game evaluations. In this study, we consider that a quality model defines a set of characteristics and/or sub-characteristics and their relationships, providing a basis for the specification of quality requirements to be evaluated (ISO/IEC, 2011). A process is defined as a sequence of interrelated steps and activities that are performed to achieve a determined purpose (IEEE, 2002). Figure 24 presents the composition of the MEEGA+ method.

Figure 24 - The MEEGA+ method



Source: developed by the author.

5.1 THE MEEGA+ MODEL

The objective of the MEEGA+ model is: to evaluate the quality of educational games in terms of usability and player experience from the students' perspective in the context of computing education (PETRI; GRESSE VON WANGENHEIM, BORGATTO, 2018a).

Following the GQM approach, this objective is systematically decomposed into factors to be measured. The initial version of the MEEGA model evaluates games in terms of motivation, user experience, and learning (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011). However, the results of its large-scale evaluation show that there is a conceptual overlap between the factors of motivation and user experience (PETRI; GRESSE VON WANGENHEIM, 2017). In addition, analysing the results of the literature review (PETRI; GRESSE VON WANGENHEIM, 2017), we identified a trend, covering a set of factors used for the games' evaluation mainly related to motivation, user experience, usability, engagement, enjoyment and perceived learning (PETRI; GRESSE VON WANGENHEIM, 2017). In this respect, the majority of these factors (motivation, user experience, usability, engagement, enjoyment, and perceived learning) is still fragmented in dimensions (Table 28) that can overlap conceptually with other factor (e.g., immersion and focused attention). Thus, some dimensions of different factors are similar or related to another dimension of another factor. Therefore, we mapped the dimensions, analysing their conceptual definitions and similarities as presented in Table 28.

Table 28 - Mapping of evaluation factors

Factor/ Dimension	**Motivation** (KELLER, 1987; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011)	**User Experience** (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011)	**Usability** (ISO/IEC, 2011)	**Engagement** (WIEBE et al., 2014)	**Enjoyment** (FU; SU; YU, 2009)	**Perceived Learning** (SINDRE; MOODY, 2003; GRESSE VON WANGENHEIM; BORGATTO, 2011)
Attention	X			X (Focused Attention)	X (Concentration)	
Relevance	X					
Confidence	X					
Satisfaction	X			X		
Immersion		X		X (Focused Attention)	X	
Social Interaction		X			X	

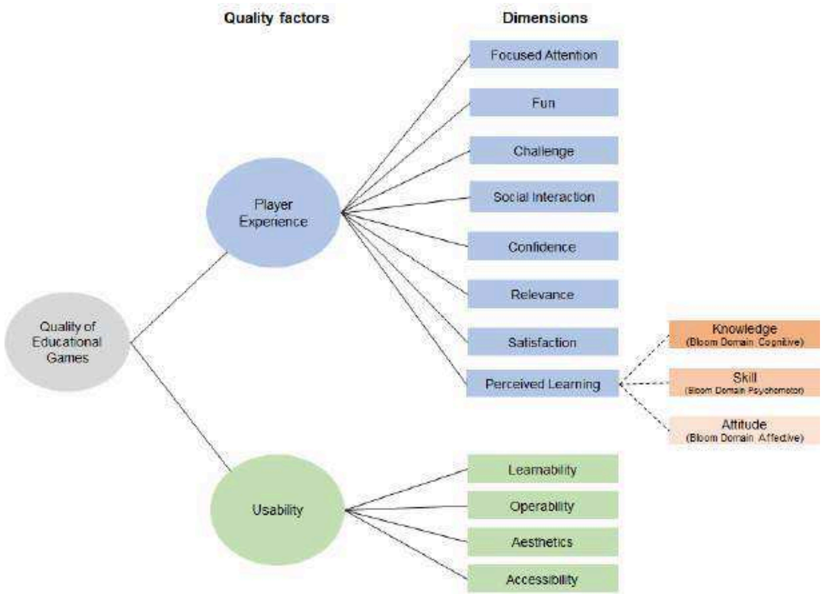
Challenge		X			X	
Fun		X		X (Satisfaction)		
Competence		X				
Control/ Operability/ Autonomy		X (Control)	X (Operability)		X (Autonomy)	
Learnability			X			
Aesthetics			X	X		
Accessibility			X			
Perceived usability				X		
Goal clarity					X	
Feedback					X	
Knowledge Improvement					X	X

Source: developed by the author.

Based on the results of the mapping, analysing conceptual definitions and similarities of the factors/dimensions, and on the results of a preliminary statistical analysis of the MEEGA+ model (exploratory factor analysis), analysing a sample of 718 students from 40 case studies (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2018a; PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2018b), we defined that the MEEGA+ model is decomposed into two quality factors and their dimensions (Figure 25).

In this study, we define the 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 (ISO/IEC, 2011; DAVIS, 1989; MOHAMED; JAAFAR, 2010). And, the player experience is a quality factor that covers a deep involvement of the student in the gaming task, including its perception of learning, feelings, pleasures, and interactions with the game, environment and other players (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011; O'BRIEN; TOMS, 2010; WIEBE et al., 2014; SWEETSER; WYETH, 2005; FU; SU; YU, 2009; TULLIS; ALBERT, 2008; KELLER, 1987; ISO/IEC, 2011; SINDRE; MOODY, 2003; POELS; KORT; IJSSELSTEIJN, 2007; GÁMEZ, 2009; TAKATALO et al., 2010; BROOKE, 1996; DAVIS, 1989).

Figure 25 - Decomposition of the MEEGA+ model



Source: developed by the author.

The definitions of these dimensions are presented in Table 29.

Table 29 - Definition of the dimensions

Quality factor	**Dimension**	**Definition**
Usability	Aesthetics	Evaluating, if the game interface enables pleasing and satisfying interaction for the user (ISO/IEC, 2011).
Usability	Learnability	Evaluating, if the game can be used by specified users to achieve specified goals of learning to use the game with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use (ISO/IEC, 2011).
Usability	Operability	Evaluating the degree to which a game has attributes that make it easy to operate and control (ISO/IEC, 2011).
Usability	Accessibility	Evaluating, if the game can be used by people with low/moderate visual impairment and/or color blindness (ISO/IEC, 2011).
Player experience	Focused Attention	Evaluating the attention, focused concentration, absorption and the temporal dissociation of the students (KELLER, 1987; WIEBE et al., 2014; SAVI;

		GRESSE VON WANGENHEIM; BORGATTO, 2011).
	Fun	Evaluating the students' feeling of pleasure, happiness, relaxing and distraction (POELS; KORT; IJSSELSTEIJN, 2007, SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011).
	Challenge	Evaluating how much the game is sufficiently challenging with respect to the learner's competency level. The increase of difficulty should occur at an appropriate pace accompanying the learning curve. New obstacles and situations should be presented throughout the game to minimize fatigue and to keep the students interested (SWEETSER; WYETH, 2005; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011).
	Social Interaction	Evaluating, if the game promotes a feeling of a shared environment and being connected with others in activities of cooperation or competition (FU; SU; YU, 2009; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011).
	Confidence	Evaluating, if students are able to make progress in the study of educational content through their effort and ability (e.g., through tasks with increasing level of difficulty) (KELLER, 1987; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011).
	Relevance	Evaluating, if students realize that the educational proposal is consistent with their goals and that they can link content with their professional or academic future (KELLER, 1987; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011).
	Satisfaction	Evaluating, if students feel that the dedicated effort results in learning (KELLER, 1987; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011).
	Perceived Learning	Evaluating the perceptions of the overall effect of the game on students' learning in the course (SINDRE; MOODY, 2003; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011).

Source: developed by the author.

In this study, we focus on the evaluation of the perceived learning, not evaluating the learning effectiveness and/or impact of the game. This is justified due to the practical limitations in to conduct evaluation studies in the computing/SE context. In this regard, different studies provide evidence that evaluations based on students' perceptions (including perceptions of their learning) provide reliable and valid data, mainly, when collected through reliable and valid measurement instruments (BROWN; ANDRADE; CHEN, 2015).

In order to operationalize the measurement of these defined quality factors/dimensions, a research design is defined.

Definition of the research design. Evaluations of educational games are typically conducted through empirical studies, which aim to measure if the target audience has achieved the defined objectives (BRANCH, 2009). These studies may range from experimental to non-experimental studies (WOHLIN et al., 2012). In order to define a research design for an empirical study, its practical limitations and objective(s) need to be considered. In the context of this study, related to the evaluation of games for computing education, it is expected that the evaluation can be conducted quickly, in a non-intrusive way in order to not interrupt the normal flow of a class and to not impair the participants involved in the study. Therefore, we chose a case study design, which allows an in-depth research of an individual, group or event (WOHLIN et al., 2012; YIN, 2017). The study is conducted as a one-shot post-test only design, in which the case study begins with the application of the treatment (educational game) and then a measurement instrument (questionnaire) is answered by the students (self-assessment) in order to collect data on their perceptions about the game.

The conduction of case studies in this context is justified by typical restrictions applying the games in the classroom. While experiments adopt more rigorous research designs, they may have significant limitations when conducted in educational contexts, such as computing/SE education (SCHANZENBACH, 2012). For example, regarding the feasibility of implementing an experiment, students may feel impaired by random allocation in control and experimental group(s). In addition, the different interventions used in the experiment may be considered inferior (SITZMANN et al., 2010). Threats can also be introduced through differences between pre/post tests and/or the impact of additional causal factors on the test results. In addition, in order to obtain significant statistical results from such experiments, a considerable sample size is required (SITZMANN et al., 2010; WOHLIN et al., 2012). However, this may not be feasible due to the small number of students commonly enrolled in computing courses (BOWMAN, 2018). Thus, even when undertaking this substantial amount of effort, the study may not yield significant results (ALL; CASTELLAR; LOOY, 2016). Therefore, a case study may be a more suited research design, providing a deep understanding of the object of study (educational game) in its real context (computing education) (WOHLIN et al., 2012) while minimizing interruptions of the normal flow of a class (PETRI; GRESSE VON WANGENHEIM, 2017).

Collecting data via self-assessment is well accepted for measuring diverse factors, such as engagement, motivation, or usability (JOHNSON; CHRISTENSEN, 2016). However, its reliability and validity may still be questionable, if used to measure learning (BROWN; ANDRADE; CHEN, 2015). In this respect, although no consensus is reached, several studies provide evidence that self-assessment can provide reliable and valid information (ROSS, 2006; ANDRADE; VALTCHEVA, 2009; THOMAS; MARTIN; PLEASANTS, 2011; BROWN; ANDRADE; CHEN, 2015; SHARMA et al., 2016), especially when using reliable and valid measurement instruments (BROWN; ANDRADE; CHEN, 2015). Therefore, a compromise may be the development of standardized and statistically validated measurement instruments increasing the validity and reliability of the data being collected (KASUNIC, 2005; DEVELLIS, 2016).

Definition of the MEEGA+ measurement instrument. In accordance with the research design and based on the defined dimensions, we generate a set of items, customizing and unifying existing standardized questionnaires (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011; KELLER, 1987; TULLIS; ALBERT, 2008; SINDRE; MOODY, 2003; SWEETSER; WYETH, 2005; POELS; KORT; IJSSELSTEIJN, 2007; GÁMEZ, 2009; TAKATALO et al., 2010, O'BRIEN; TOMS, 2010; WIEBE et al., 2014; FU; SU; YU, 2009; MOHAMED; JAAFAR, 2010; ZAIBON; SHIRATUDDIN, 2010; ZAIBON, 2015; BROOKE, 1996; DAVIS, 1989; ISO/IEC, 2011).

Table 30 shows the items of the MEEGA+ measurement instrument for each quality factor/dimension and their sources/references.

Table 30 - MEEGA+ measurement instrument items and their references

Quality factor	Dimension	Item No.	Description
Usability	Aesthetics (ISO/IEC, 2011)	1	The game design is attractive (interface, graphics, cards, boards, etc.).
		2	The text font and colors are well blended and consistent.
	Learnability (ISO/IEC, 2011)	3	I needed to learn a few things before I could play the game.
		4	Learning to play this game was easy for me.

		5	I think that most people would learn to play this game very quickly.
	Operability (ISO/IEC, 2011)	6	I think that the game is easy to play.
		7	The game rules are clear and easy to understand.
	Accessibility (ISO/IEC, 2011)	8	The fonts (size and style) used in the game are easy to read.
		9	The colors used in the game are meaningful.
Player experience	Confidence (KELLER, 1987; SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011)	110	The contents and structure helped me to become confident that I would learn with this game.
	Challenge (SWEETSER; WYETH, 2005; GRESSE VON WANGENHEIM; BORGATTO, 2011)	11	This game is appropriately challenging for me.
		12	The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace.
		13	The game does not become monotonous as it progresses (repetitive or boring tasks).
	Satisfaction (KELLER, 1987; GRESSE VON WANGENHEIM; BORGATTO, 2011)	14	Completing the game tasks gave me a satisfying feeling of accomplishment.
		15	It is due to my personal effort that I managed to advance in the game.
		16	I feel satisfied with the things that I learned from the game.
		17	I would recommend this game to my colleagues.
	Social Interaction (FU; SU; YU, 2009; GRESSE VON WANGENHEIM; BORGATTO, 2011)	18	I was able to interact with other players during the game.
		19	The game promotes cooperation and/or competition among the players.
		20	I felt good interacting with other players during the game.
	Fun (POELS; KORT; IJSSELSTEIJN; 2007, GRESSE VON WANGENHEIM; BORGATTO, 2011)	21	I had fun with the game.
		22	Something happened during the game (game elements, competition, etc.) which made me smile.
	Focused Attention	23	There was something interesting at the beginning of the game that captured my attention.

	(KELLER, 1987; WIEBE et al., 2014; GRESSE VON WANGENHEIM; BORGATTO, 2011)	24	I was so involved in my gaming task that I lost track of time.
		25	I forgot about my immediate surroundings while playing this game.
	Relevance (KELLER, 1987; GRESSE VON WANGENHEIM; BORGATTO, 2011)	26	The game contents are relevant to my interests.
		27	It is clear to me how the contents of the game are related to the course.
		28	This game is an adequate teaching method for this course.
		29	I prefer learning with this game to learning through other ways (e.g. other teaching methods).
	Perceived Learning (SINDRE; MOODY, 2003; GRESSE VON WANGENHEIM; BORGATTO, 2011)	30	The game contributed to my learning in this course.
		31	The game allowed for efficient learning compared with other activities in the course.

Source: developed by the author.

The standardized items presented in Table 30 compose the MEEGA+ measurement instrument in order to evaluate both digital and non-digital games for computing education.

Furthermore, items related to the learning goals of each game are included in the measurement instrument to be customized in accordance with the specific learning goals of each educational game. Typically, games for computing education are used to improve the knowledge on the cognitive levels of remembering, understanding, and application (ACM; IEEE-CS, 2013) in accordance with the revised version of Bloom's taxonomy (ANDERSON; KRATHWOHL; BLOOM, 2001). However, the MEEGA+ model is also flexible to cover goals related to higher cognitive levels, such as analysing, evaluating and creating. Besides technical knowledge, games can also contribute to skill development, such as problem-solving, communication, teamwork, leadership, etc. (ACM; IEEE-CS, 2013). Such learning goals can be classified in accordance with the taxonomy of the psychomotor domain (SIMPSON, 1972). In addition, games can also contribute to develop professional attitudes, such as a mature behaviour of the student, considering professional and legal issues as well as an ethical attitude in the profession (ACM; IEEE-CS, 2013). Learning goals related to this typically target the awareness and growth in attitudes, emotion, and feelings and can be classified in accordance to the taxonomy of affective domain (KRATHWOHL, BLOOM; MASIA, 1973). Thus, for each learning goal of the game, the following statement should be customized in the MEEGA+ measurement instrument: *The game contributed to <verb related to the level of the learning goal (cognitive, psychomotor, and affective)> <goal/concept>*. For example, in accordance with the learning goals of SCRUMIA (GRESSE VON WANGENHEI; SAVI; BORGATTO, 2013), a game to reinforce the understanding of SCRUM concepts and to practice the SCRUM process, such statement would be: "*The game contributed to recall concepts related to Sprint Planning.*" These statements related to the learning goals of the game also compose the MEEGA+ measurement instrument that is applied after the game session in order to capture the students' perceptions about their level of agreement (or disagreement) in the achieving of the learning objective(s) of the game.

Additionally, the MEEGA+ measurement instrument also supports the identification of the sample characteristics, collecting demographic information of the students that compose the sample of the study. Know the characteristics of the sample are important in order to interpret and analyse the data collected (WOHLIN et al., 2012). Thus, the sample is characterized through the MEEGA+ measurement instrument in terms of age group, gender, and frequency that the students' play games (digital and/or non-digital games).

A comparison between the measurement instruments of the initial version of the MEEGA model (SAVI, 2011) and the MEEGA+ model is presented in Appendix D.

Response format. As response format, we adopt a 5-point Likert scale with response alternatives ranging from strongly disagree to strongly agree (DEVELLIS, 2016; MALHOTRA; BIRKS, 2008). The use of a Likert scale, in its original 5-point format, allows to express the opinion of the individual (student) under the object of study (educational game) with precision, besides allowing the individual being comfortable to express their opinion, using a neutral point and, thus, contributing to the quality of the answers (DAWES, 2008).

Analysis of the data collected. In order to provide a support in how to analyse the data collected through the MEEGA+ measurement instrument, we propose a definition of the data analysis, adopting the GQM approach (BASILI; CALDIERA; ROMBACH, 1994).

Following the MEEGA+ model, the objective of the evaluation is: to evaluate the quality of educational games in terms of usability and player experience from the students' perspective in the context of computing education (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2018a).

Based on the objective defined for the evaluation, following the MEEGA+ model and the GQM approach (BASILI; CALDIERA; ROMBACH, 1994), the objective is decomposed into quality aspects and analysis questions to be analysed:

Usability

AQ1: Does the <*name of the evaluated game*> have a good usability?

Player Experience

AQ2: Does the <*name of the evaluated game*> provide a positive player experience?

In addition to this analysis questions, complementary questions may address the identification of the characteristics of the sample in terms of age, gender and frequency that the students play games:

AQ3: How old are the students that compose the sample of the study?

AQ4: What is the gender of the students that compose the sample of the study?

AQ5: What is the frequency that the students play digital and/or non-digital games?

In order to answer these defined analysis questions, data collected through the MEEGA+ measurement instrument are analysed through descriptive statistical methods (TROCHIM; DONNELLY, 2008; WOHLIN et al., 2012). Descriptive statistical methods are used to describe, and graphically present interesting aspects of the data collected (WOHLIN et al., 2012). Table 31 presents the descriptive statistical methods used to answer each analysis question.

Table 31 - Descriptive statistical methods used to answer the analysis questions

Analysis questions	**Descriptive statistical methods**	**Data analysed**
AQ1: Does the <*name of the evaluated game*> have a good usability?	Measures of central tendency (median, average and frequency of responses); Graphical visualization (frequency charts)	Data collected through the MEEGA+ measurement instrument on the usability quality factor.
AQ2: Does the <*name of the evaluated game*> provide a positive player experience?	Measures of central tendency (median, average and frequency of responses); Graphical visualization (frequency charts)	Data collected through the MEEGA+ measurement instrument on the player experience quality factor.

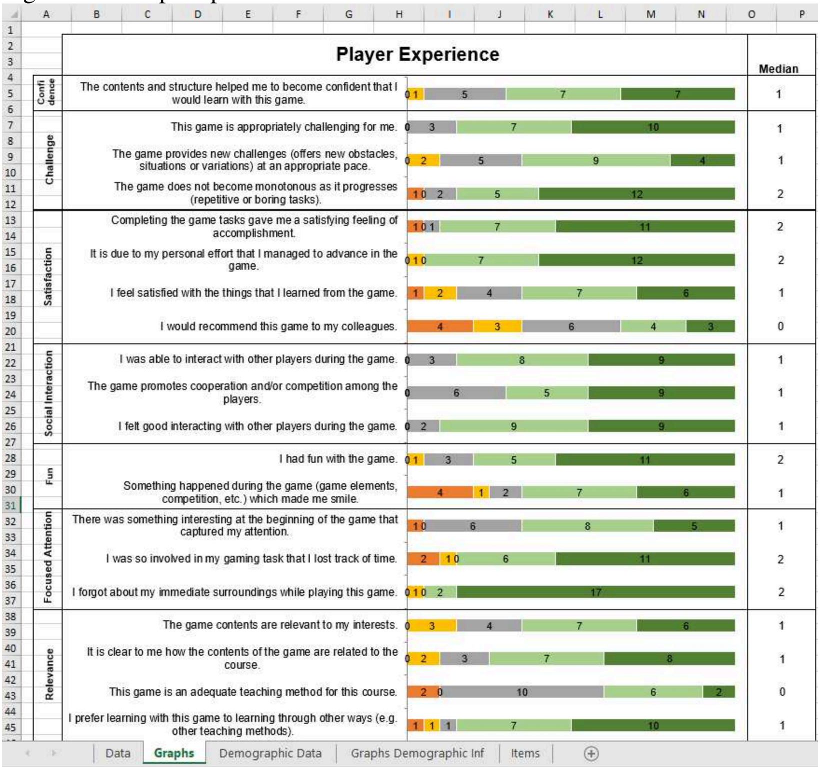
AQ3: How old are the students that compose the sample of the study?	Measures of central tendency (frequency of responses); Graphical visualization (frequency charts)	Data collected through the MEEGA+ measurement instrument on demographic information of the sample.
AQ4: What is the gender of the students that compose the sample of the study?	Measures of central tendency (frequency of responses); Graphical visualization (frequency charts)	Data collected through the MEEGA+ measurement instrument on demographic information of the sample.
AQ5: What is the frequency that the students play digital and/or non-digital games?	Measures of central tendency (frequency of responses); Graphical visualization (frequency charts)	Data collected through the MEEGA+ measurement instrument on demographic information of the sample.

Source: developed by the author.

In order to operationalize the analysis of the data collected and assist researchers to answer the defined analysis questions using descriptive statistical methods, a standardized analysis spreadsheet was prepared using Microsoft Excel (desktop version). The spreadsheet contains five internal spreadsheets (Data; Demographic data; Graphs, Demographic graphs; and Items).

In order to answer the analysis questions AQ1 and AQ2, frequency charts are automatically generated in the analysis spreadsheet, presenting the frequency of the responses and the median for each measurement instrument item. The graphics are presented in accordance with the quality factors and dimensions of the MEEGA+ model (Figure 26).

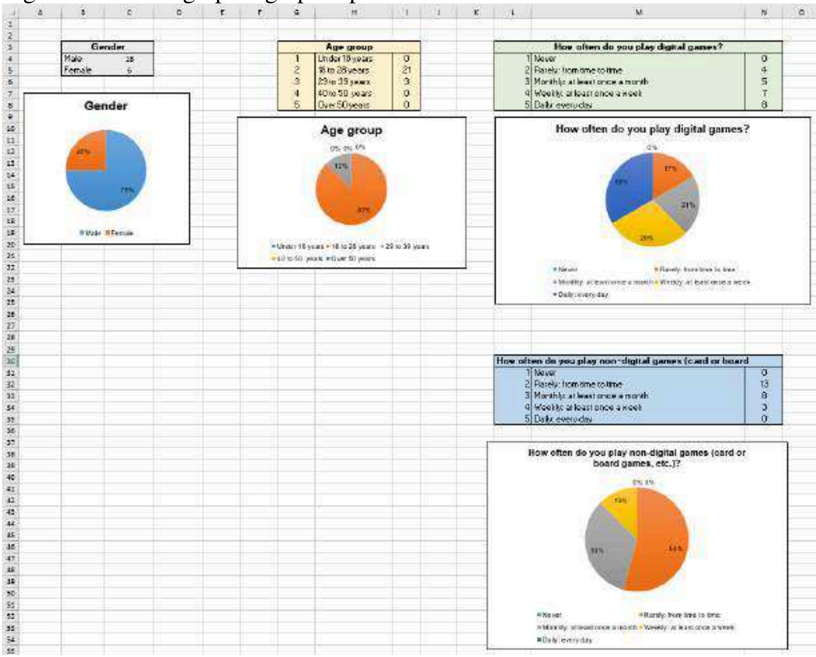
Figure 26 - Graphs spreadsheet



Source: developed by the author.

Similarly, the graphics presenting the demographic information are generated and presented in the Demographic graph spreadsheet (Figure 27), providing information to answer the analysis questions AQ3, AQ4, and AQ5.

Figure 27 - Demographic graphs spreadsheet



Source: developed by the author.

In order to assist instructors/researchers in the use of the MEEGA+ data analysis spreadsheet, an instruction guide was prepared. This guide describes the step by step for the use of the MEEGA+ data analysis spreadsheet, describing how the data should be inserted and prepared to support the automatic generation of the descriptive analysis.

5.1.1 Game Quality Scale

The MEEGA+ scale aims to classify the evaluated game in a quality level, based on the students' perception, using the MEEGA+ measurement instrument.

The scale has been developed by adopting the statistical technique Item Response Theory (IRT) (PASQUALI; PRIMI, 2003), which allows to express through mathematical models the relationship between observable variables (questionnaire items) and latent traits (game's quality) based on the students' perceptions (PASQUALI; PRIMI, 2003).

Based on data collected in the case studies applying the MEEGA+ model (Table 40), the parameters of discrimination (*a*) and difficulty (*b*) were determined for each MEEGA+ measurement instrument item (Table 32), adopting the probabilistic model proposed by Samejima (1969). The parameter *a* is associated with how much the questionnaire item discriminates (differentiates) the students in relation to the latent trait (game's quality), where the higher its value is, the more associated with the latent trait is the questionnaire item. The discrimination parameters range from 0.752 to 2.987, indicating that all items have a satisfactory discrimination parameter (ANDRADE; TAVARES; VALLE, 2000). The parameter *b* is associated with the degree of difficulty of the questionnaire item, where the higher its value is, the more difficult the students to agree with the questionnaire item in relation to the game's quality. Thus, based on the probabilistic model (SAMEJIMA, 1969), the probability of a student *j* chose a category *k* (*k* = 1, 2, 3, 4, 5) for a measurement instrument item *i* is given by:

$$P_{i,k}(\theta) = \frac{1}{1+e^{-a_i(\theta_j - b_{i,k})}} - \frac{1}{1+e^{-a_i(\theta_j - b_{i,k+1})}}$$

where an item that has 5 categories (ranging to strongly disagree to strongly agree) assumes the restriction where *b1*=0, presenting 4 values of the difficulty parameter in the scale (*b2*, *b3*, *b4*, *b5*), besides of the discrimination parameter (*a*), which determines how much the item discriminates (differentiates) the students in relation to the latent trait (game's quality) (θ_j).

In accordance with Tezza, Bornia and Andrade (2011), a measurement instrument item to be considered satisfactory in a measurement scale, should have the discrimination value $a >= 0.7$. Thus, measurement instrument items that have a discrimination parameter $a < 0.7$ are disregarded from the analysis, as they may not be correctly differentiating the quality level of an evaluated game (PASQUALI; PRIME, 2003). Thus, items 2, 3, 4, 8, and 9, were disregarded in the construction of the scale. Table 32 shows the parameters of the remaining items for the scale definition.

Table 32 - Parameters of discrimination (*a*) and difficulty (*b*)

Item No.	*a*	*b1*	*b2*	*b3*	*b4*
1	0.937	-5.274	-2.820	-0.768	1.667
5	0.752	-5.477	-2.720	- 1.381	1.072
6	0.811	-6.389	-3.599	-1.996	0.598
7	1.076	-5.045	-3.095	-1.700	0.593
10	1.598	-3.582	-2.187	-0.642	1.368
11	1.558	-2.665	-1.565	-0.469	1.479
12	1.637	-2.669	-1.584	-0.501	1.264
13	1.438	-2.676	-1.247	-0.273	1.396
14	2.232	-2.199	-1.379	-0.434	0.947
15	1.291	-3.161	-1.931	-0.566	1.380
16	2.517	-2.535	-1.823	-0.712	0.923
17	2.602	-2.069	-1.246	-0.494	0.852
18	1.397	-3.060	-2.350	-1.373	0.303
19	1.678	-2.654	-1.979	-1.364	0.243
20	2.049	-2.569	-1.978	-0.894	0.450
21	2.987	-2.262	-1.559	-0.750	0.423
22	1.776	-2.680	-1.724	-0.952	0.478
23	1.643	-2.730	-1.450	-0.259	1.368
24	1.842	-1.955	-0.965	-0.120	1.043
25	1.621	-2.047	-0.932	0.113	1.380
26	1.578	-3.083	-2.129	-0.778	0.906
27	1.189	-5.093	-3.779	-2.345	-0.119
28	1.752	-3.269	-2.476	-1.317	0.165
29	1.284	-3.355	-2.303	- 0.725	0.722
30	1.878	-3.230	-2.329	-1.143	0.549
31	1.575	-3.678	-2.269	-0.856	0.917

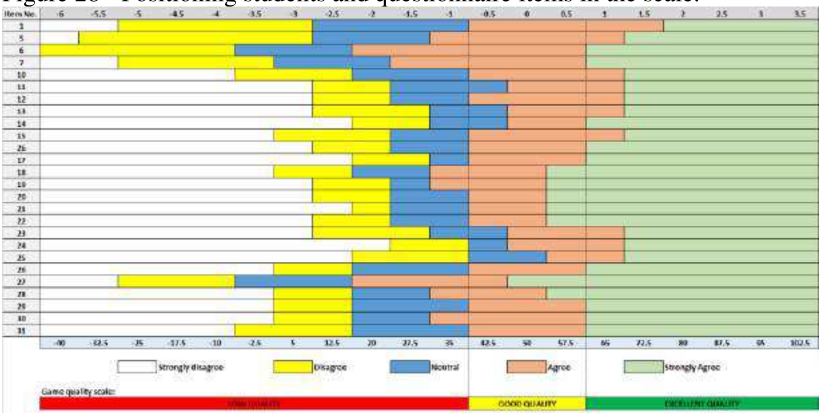
Source: developed by the author.

The parameter degree of difficulty (*b*) indicates the category of the scale in which the item has more information. Items 13 (The game does not become monotonous as it progresses (repetitive or boring tasks)), 11 (This game is appropriately challenging for me) and 1 (The game design is attractive) have a large degree of difficulty. Thus, when a student strongly agrees with these items, there is a high latent trait, indicating a large quality level of the evaluated game. On the other hand, items 5 (I think that most people would learn to play this game very quickly) and 6 (I think that the game is easy to play) have a low degree of difficulty. Thus, these items would be answered with agreement by the majority of the students, including in the evaluation of games that have a low-quality level.

Based on the parameters of discrimination and difficulty it is possible to interpret how the measurement instrument items contribute to the definition of a measurement scale. The scale is defined from anchor items, which determine the categories of the latent trait (game's quality).

Thus, to position the items on the scale and identify the categories of the scale (quality levels) is considered the probability parameter $P_{i,k}(\theta)>=0.50$ (ANDRADE; TAVARES; VALLE, 2000). Based on this definition, the items are positioned on the scale and the anchor items are identified, shown in Figure 28.

Figure 28 - Positioning students and questionnaire items in the scale.



Source: developed by the author.

Based on the positioning of items throughout the scale, three levels of quality are defined: low quality, good quality, and excellent quality (Table 33). The quality levels are defined using an average 50 and standard deviation 15, scale (50.15), applying the formula $\theta_{50.15}=50+15*\theta_{0.1}$ (ANDRADE; TAVARES; VALLE, 2000).

Table 33 - Game quality levels

Quality level	Level description
Low quality ($\theta < 42.5$)	At this level, the game rarely provides social interaction and hardly ever produces moments of fun among the players. The game does not capture the student's focused attention, does not arouse the confidence that he/she will learn from the game, nor does it produce feelings of satisfaction. The game rarely presents challenges, has monotonous tasks and does not contribute to student learning. Although a game at this level has a low relevance to the students' interests, a student recognizes that the game's content is related to the course. In terms of usability, a game at this level sometimes exhibits operability features, which may have some clear rules and be easy to play.
Good quality ($42.5 <= \theta < 65$)	At this level, the game sometimes presents challenging activities, offering new challenges for students. It provides moderately focused attention to the players, although students do not forget about their surroundings. Sometimes the game also provides feelings of

	confidence and satisfaction in the players. Frequently the game presents moments of social interaction and fun among the players. Often the game is considered relevant to the students' interests and, usually, the students recognize that the game's content is related to the course. Frequently the game contributes efficiently to student learning. In terms of usability, the game usually has the clear rules and is easy to play, although, usually does not present a fully attractive design.
Excellent quality ($\theta >= 65$)	At this level, the game is challenging for students and has no monotonous activities. It is highly relevant to students' interests and provides excellent focused attention, satisfaction, fun, and social interaction. It allows the student to be confident that he/she will learn from the game and contribute to and efficient student learning. In terms of usability, the game presents excellent operability and learnability, that is, it has clear rules and is easy to learn to play. Even so, a game at this level may present improvements in terms of aesthetics, not presenting a fully attractive design.

Source: developed by the author.

The quality level of an evaluated game is determined based on the data collected using the MEEGA+ measurement instrument and analysing them through an R script which applies the defined scale scores in the collected data. The scripts and the instruction guide to determine the quality level of a game are available at <http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/>.

5.2 THE MEEGA+ PROCESS

In order to guide the application of the model, the MEEGA+ method also contains a systematic process for the execution of the evaluation of educational games. The process provides a systematic support, detailing the steps and interrelated activities, guiding researchers and instructors in the conduction of game evaluations.

The MEEGA+ process has been designed in a prescriptive way, defining how the process should be executed, establishing rules and procedures (ACUÑA; FERRÉ, 2001). Thus, the MEEGA+ process is described through phases, activities and work products (ACUÑA et al., 2000; BENALI; DERNIAME, 1992; FINKELSTEIN; KRAMER; NUSEIBEH, 1994):

- **Phase** is a set of activities grouped in steps, presenting a logical and structured sequence.
- **Activity** is the stage of a process that produces externally visible changes of state in the product. An activity can have inputs,

outputs, intermediate results (work products). The activity implements procedures, rules and objectives to transform a product.

- **Work products** are the inputs and outputs of an activity from a process. They can be produced and consumed throughout the process and can have long life cycles, being created, accessed and modified (changing its status) during the process. Each work product has a status (initial, in preparation, and ready to use). The status of a work product indicates its position in the process, determining which actions can be performed on the work product (MÜNCH et al., 2012). Table 34 presents the status of the work products adopted in the MEEGA+ process, defined based in Münch et al. (2012).

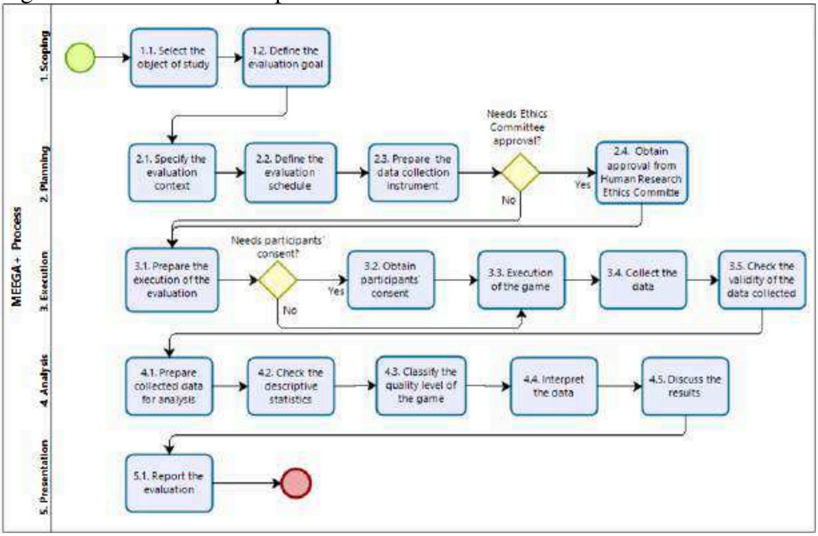
Table 34 - Status of the work products

Status	Description
Initial	Work products are initially set to this status. Typically, there is not an action taken on the work product. Typically, a work product with this status is a work product created in the process or an existing product that needs a customization, used as an input to an activity such as, a game developed, a template of a document/spreadsheet, etc. Example: MEEGA+ data collection instruments
In Preparation	Indicating that the work product is incomplete, not ready for use. Examples: the game needs to be installed, the data collection instruments need to be printed.
Ready to use	Indicating that all activities associated with the work product has been completed and the work product is ready to be used. Examples: the game is installed and ready to play, the questionnaires are printed.

Source: developed by the author.

Phases and activities of the MEEGA+ process are defined based on the process of empirical studies proposed by Wohlin et al. (2012), and in practical experiences of our research group (GQS/INCoD/INE/UFSC) in the conduction of game evaluations (http://www.gqs.ufsc.br/software-engineering-education/). Figure 29 presents the MEEGA+ process modelling adopting the BPM notation.

Figure 29 - The MEEGA+ process



Source: developed by the author.

The MEEGA+ process may be conducted either manually supported by tools, using the standardized questionnaires, spreadsheets and scripts provided by the MEEGA+ model (available in English, Spanish, and Brazilian Portuguese at http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/). Or semi-automated, adopting the AssistantMEEGA+ v0.9 (SOARES et al., 2018) (http://assistantmeega.com.br/), developed by our research partners from USES/UFAM (Usability and Software Engineering Research Group[3], Federal University of Amazonas). However, the current version (v0.9) of the AssistantMEEGA+ provides only a partial support to the MEEGA+ process, not supporting all activities of the MEEGA+ process, e.g. not supporting the classification of the quality level of the evaluated game, using the MEEGA+ scale, and, currently, it is available in Brazilian Portuguese only.

Typically, game evaluations are conducted by researchers. However, the MEEGA+ process is described in order to be applied by any student or instructor, not requiring advanced knowledge in education, measurement or statistics. Thus, in this process, we use the generic role

[3] http://uses.icomp.ufam.edu.br/

researcher representing any student/instructor/etc. responsible for the game evaluation.

The MEEGA+ process is organized into phases and activities:

Phase 1. Scoping. In the first phase of the MEEGA+ process, the evaluation scope is defined. The scope of the evaluation is set by defining its goals and the object of study (Table 35).

As this process does not support the development of a new game, a premise for the process is that the game that will be evaluated must have already been developed.

Table 35 - Activities of the Scoping phase

Phase 1. Scoping	
Activity 1.1. Select the object of study (educational game)	
Description	The objective of this activity is to select the object of study (educational game). An educational game is classified as an instructional strategy that typically involves competition and is organized by rules and restrictions to achieve a certain educational goal. Typically, an educational game is characterized by various elements, such as goals, rules, restrictions, interaction, challenge, competition, rewards and feedback (ABT, 2002; PRENSKY, 2007; RITTERFELD; CODY; VORDERER, 2010; DJAOUTI et al., 2011). The selected game may be a digital game (computer game) or a non-digital game (board, cards game, etc.). Digital games are electronic games that involve human interaction with a user interface to generate visual feedback on an electronic device such as smartphones, computers, tablets, etc., whereas non-digital games are played with non-digital resources such as game boards, cards, pencils and papers, etc. (ABT, 2002; PRENSKY, 2007; RITTERFELD; CODY; VORDERER, 2010; DJAOUTI et al., 2011). The selected game should be specific to the development of a skill, knowledge, or attitude, for any knowledge area in the context of computing education. It is expected that the selected game will not be offensive or violent. Does not discriminate against students in terms of race, sex, religion, nationality, disability, sexual orientation or age. And, it should attend the game's conditions of use (copyright). As a suggestion, the game may be selected from game repositories: https://pmiteach.org/teaching-pm/resources/ https://www.infoq.com/news/2008/10/agile-games https://list.ly/list/CL-management-games-and-simulations-for-itsm http://games-factory-online.nl/seriousgames-english/seriousgamescatalogue/ http://www.semq.eu/leng/proimplo.htm http://tastycupcakes.org/pt http://www.gqs.ufsc.br/igr In addition, the game may also be selected from the results of systematic reviews that report several games for computing education: Battistella, P.; Gresse von Wangenheim, C. Games for Teaching Computing in Higher Education – A Systematic Review. IEEE Technology and

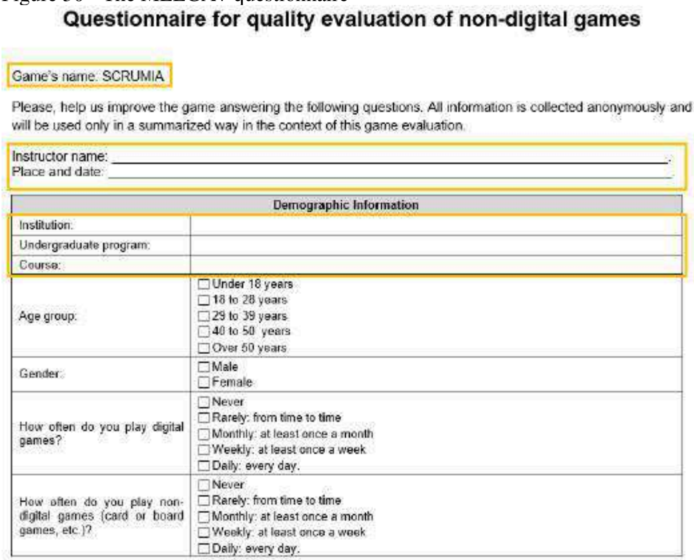
	Engineering Education (ITEE) Journal, 9(1), 8-30, 2016. Available at: http://www.gqs.ufsc.br/wp-content/uploads/2011/11/ITEE-Games-for-Teaching-Computing-in-Higher-Education_Vdraft.pdf Petri, G.; Gresse von Wangenheim, C. How games for computing education are evaluated? A systematic literature review. Computers & Education, 107, 68-90, 2017. Available at: http://dx.doi.org/10.1016/j.compedu.2017.01.004
Work products	**Input:** P1.1.1 – Educational game [Initial]
	Output: P1.1.1 – Educational game [In Preparation]
Activity 1.2. Define the evaluation goal	
Description	This activity aims to define the evaluation goal in terms of the object of study, purpose, quality aspects, perspective, and context. The evaluation goal is defined considering the definitions proposed by Wohlin et al. (2012) and Basili et al. (1994): ▪ The object of study is the entity that is analysed in the evaluation. ▪ The purpose defines what the intention of the evaluation. ▪ The quality aspects define the factors under analysis in the evaluation. ▪ The perspective tells the viewpoint from which the evaluation results are interpreted. ▪ The context is the environment in which the evaluation is run. Thus, based on the definition of the MEEGA+ model, which follows the GQM goal template (BASILI; CALDIERA; ROMBACH, 1994), the evaluation objective is defined: Analyse the <*name of the selected game*> for the purpose of *evaluate the quality* in terms of *usability and player experience* from the *students' point of view* in the context of *higher computing education*.
Work products	**Input:** P1.1.1 – Educational Game [In Preparation]; P1.2.1 – MEEGA+ evaluation goal [Ready to use] The detailed description of the MEEGA+ model is available at <http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/>
	Output: P1.2.2 - Evaluation goal [Ready to use]

Source: developed by the author.

Phase 2. Planning. After scoping of the evaluation, it is planned (Table 36).

Table 36 - Activities of the Planning phase

Phase 2. Planning	
Activity 2.1. Specify the evaluation context	
Description	This activity aims to specify the context where the evaluation will be conducted. Typically, an educational game is used in an educational context. Thus, in this activity, the context is specified in terms of educational institution, program and course, in which the game will be applied.
Work products	**Input:** P1.1.1 – Educational Game [In Preparation]; P1.2.2 - Evaluation goal [Ready to use]
	Output: P2.1.1 – Evaluation context [Ready to use]
Activity 2.2. Define the evaluation schedule	

Description	This activity aims to plan the evaluation schedule, defining the date, hour and place of the application of the game in the defined course. If the selected game is a digital game and needs a computer lab (with computers, tablets, etc.) available to play it. In this activity, the lab must be defined and scheduled, considering the expected population (P2.1.1 – Evaluation context).
Work products	**Input:** P1.1.1 – Educational Game [In Preparation]; P2.1.1 – Evaluation context [Ready to use]
	Output: P2.2.1 – Evaluation schedule [Ready to use]
Activity 2.3. Prepare the data collection instrument	
Description	This activity aims to prepare/customize the data collection instrument, provided by the MEEGA+ model, for the evaluation of the selected game. If the evaluation is conducted manually, it is necessary to choose the questionnaire language (English, Spanish, and Brazilian Portuguese). Information about the evaluation context (P2.1.1 – Evaluation context) can be previously inserted in the questionnaire in the Demographic Information section, including the game's name, researcher, place and date, institution, program, and course. Figure 30 presents an example based in the non-digital game SCRUMIA (GRESSE VON WANGENHEIM; SAVI; BORGATTO, 2013): Figure 30 - The MEEGA+ questionnaire  **Questionnaire for quality evaluation of non-digital games** Game's name: SCRUMIA Please, help us improve the game answering the following questions. All information is collected anonymously and will be used only in a summarized way in the context of this game evaluation. Instructor name: _____ Place and date: _____ **Demographic Information** Institution: Undergraduate program: Course: Age group: ☐ Under 18 years ☐ 18 to 28 years ☐ 29 to 39 years ☐ 40 to 50 years ☐ Over 50 years Gender: ☐ Male ☐ Female How often do you play digital games? ☐ Never ☐ Rarely: from time to time ☐ Monthly: at least once a month ☐ Weekly: at least once a week ☐ Daily: every day. How often do you play non-digital games (card or board games, etc.)? ☐ Never ☐ Rarely: from time to time ☐ Monthly: at least once a month ☐ Weekly: at least once a week ☐ Daily: every day. Source: developed by the author. Another customization is in terms of the learning objectives of the selected game. For each learning objective of the game, the following statement should be customized in the questionnaire: The game contributed to <*verb related to the level of the learning goal (cognitive, psychomotor, and affective)*> <goal/concept>. An example, in accordance with the learning goals of SCRUMIA (GRESSE VON WANGENHEIM; SAVI; BORGATTO, 2013), would be:

<table>
<tr><td></td><td>
“The game contributed to recall concepts related to Sprint Planning.”

Figure 31 presents an example of the statement regarding the learning objective of SCRUMIA.

Figure 31 - Customized items of the MEEGA+ questionnaire
<table>
<tr><td rowspan="2">Statements</td><td colspan="5">Select an option as your evaluation</td></tr>
<tr><td>Strongly disagree</td><td>Disagree</td><td>Neither disagree nor agree</td><td>Agree</td><td>Strongly agree</td></tr>
<tr><td>The game contributed to my learning in this course.</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td></tr>
<tr><td>The game allowed for efficient learning compared with other activities in the course</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td></tr>
<tr><td>The game contributed to recall the concepts related to Sprint Planning.</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td></tr>
</table>
Source: developed by the author.

The MEEGA+ method is also flexible to include additional items if the researcher needs to evaluate another quality factor(s). However, as this factor is outside the scope of the MEEGA+, it is not possible to assure its validity and reliability in the measurement of new factors included.
</td></tr>
<tr><td rowspan="2">Work products</td><td>Input: P1.1.1 – Educational Game [In Preparation]; P2.3.1 – MEEGA+ data collection instruments [Initial]; P2.1.1 – Evaluation context [Ready to use]; P2.2.1 – Evaluation schedule [Ready to use]
The MEEGA+ questionnaires are available in English, Spanish, and Brazilian Portuguese at http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/.</td></tr>
<tr><td>Output: P2.3.1 – MEEGA+ data collection instruments [In Preparation]</td></tr>
<tr><td colspan="2">Activity 2.4. Obtain approval from the Human Research Ethics Committee (optional)</td></tr>
<tr><td>Description</td><td>Although the evaluation of a game offers minimal risk to the participants, some educational institutions require that all research involving humans be approved by the ethics committee. Thus, before the conduction of the evaluation is necessary to verify the requirements of the ethics committee of the institution that the evaluation will be conducted.
To approve a research involving humans, an ethics committee, typically, requires the declaration of a coordinator, a research project, data collection instruments, and a consent form.
This activity is considered a risk in the process, as typically the approval of the research by the ethics committee takes a considerable time.</td></tr>
<tr><td rowspan="2">Work products</td><td>Input: P1.1.1 – Educational Game [In Preparation]; P2.3.1 – MEEGA+ data collection instruments [In Preparation]; P1.2.2 - Evaluation goal [Ready to use]; P2.1.1 – Evaluation context [Ready to use]; P2.2.1 – Evaluation schedule [Ready to use]</td></tr>
<tr><td>Output: P2.4.1 – Approval of the Ethics Committee (optional) [Ready to use]</td></tr>
</table>

Source: developed by the author.

Phase 3. Execution. When an evaluation has been planned, it is carried out in order to collect the data to be analysed. This third phase aims to organize and define the execution of the game evaluation for the selected participants. In this phase, data are collected in order to achieve the evaluation objective (Table 37).

Table 37 - Activities of the Execution phase

Phase 3. Execution	
Activity 3.1. Prepare the execution of the evaluation	
Description	This activity aims to prepare the materials required for the execution of the game. In this activity, if necessary, the data collection instruments and consent forms should be printed, based on the number of the participants described in P2.1.1 – Evaluation context. To non-digital games: organize and/or print (if necessary) the materials of non-digital games, such as boards, cards, etc. To digital games: organize and install (if necessary) the game. Additional materials should be prepared, such as projector, mobile devices for digital games. And, tables and chairs for non-digital games.
Work products	**Input:** P1.1.1 – Educational Game [In Preparation]; P2.1.1 – Evaluation context [Ready to use]; P2.2.1 – Evaluation schedule [Ready to use]; P2.3.1 – MEEGA+ data collection instruments [In Preparation]
	Output: P2.3.1 – MEEGA+ data collection instruments [Ready to use]
	Output: P3.1.2 – Consent form (optional) [Ready to use]
	Output: P1.1.1 – Educational Game [Ready to use]
Activity 3.2. Obtain participants' consent (optional)	
Description	Before the execution of the game, if necessary, the consent form should be signed by all participants, indicating that they agree and accept to participate in the research.
Work products	**Input:** P3.1.2 – Consent form [Ready to use]
	Output: P3.2.1 – Participants' consent [Ready to use]
Activity 3.3. Execution of the game	
Description	During this activity the game is applied in the participants, using the game materials and considering the evaluation schedule (P2.2.1 – Evaluation schedule) and context (P2.1.1 – Evaluation context). Before the game execution, the instructor presents the game and explain its rules. Then, the participants can start playing the game.
Work products	**Input:** P1.1.1 – Educational Game [Ready to use]; P2.1.1 – Evaluation context [Ready to use]; P2.2.1 – Evaluation schedule [Ready to use]; P2.3.1 – MEEGA+ data collection instruments [Ready to use]
	Output: P3.3.1 – Game executed [Ready to use]
Activity 3.4. Collect the data	
Description	After the execution of the game, data collection takes place. If the evaluation is being conducted using the AssistantMEEGA+, this activity is performed through an electronic form of the AssistantMEEGA+. If the evaluation is being conducted manually, the data collection instrument should be distributed to the participants in order to them fill out the questionnaire based on their perceptions about the game.
Work products	**Input:** P2.3.1 – MEEGA+ data collection instruments [Ready to use]; P3.3.1 – Game executed [Ready to use]
	Output: P3.4.1 – Data collected [Initial]
Activity 3.5. Check the validity of the data collected	
Description	When data has been collected, the researcher must check if the data are reasonable and that it has been collected correctly. This covers aspects such

	as if the participants have understood the items and therefore answered them correctly. Thus, based on Wohlin et al. (2012), Yin (2017) and Izquierdo and Pedrero (2014), criteria are defined in order to check if the data collected are valid: ▪ Check if all participants have participated seriously of the evaluation (e.g. not answering all the questionnaire items using the same response category). ▪ Check if all questionnaire items have only one response (one response category). ▪ Check if each questionnaire was answered completely and correctly (not missing to answer more than 10% of the questionnaire items (4 or more items) (IZQUIERDO; PEDRERO, 2014)). Questionnaires answered incompletely (missing to answer 4 or more questionnaire items) and that not meet the defined criteria, should be disregarded of the analysis and the participant removed from the sample.
Work products	**Input:** P3.4.1 – Data collected [Initial]
	Output: P3.4.1 – Data collected [In Preparation]

Source: developed by the author.

Phase 4. Analysis. After collecting data in the execution phase, conclusions are drawn based on this data. To be able to draw valid conclusions, evaluation data must be analysed. Thus, the fourth phase of the MEEGA+ process aims to interpret and analyse the data collected in the execution phase (Table 38).

Table 38 - Activities of the Analysis phase

Phase 4. Analysis	
Activity 4.1. Prepare collected data for analysis	
Description	If the evaluation was conducted using the AssistantMEEGA+, this activity is performed in an automatic way. If the evaluation was conducted manually, the data collected using the MEEGA+ questionnaire must be prepared for the data analysis process by introducing them in the data analysis spreadsheet provided by the MEEGA+ model. Select and download the spreadsheet on the website <http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/>. The researcher needs to introduce the data collected in the spreadsheet and customize the learning objectives, following the detailed instructions for the use of the MEEGA+ data analysis spreadsheet. As presented in Figure 32, each row of the data analysis spreadsheet represents an answer of one student for the game evaluation (one questionnaire). The columns represent the MEEGA+ measurement instrument items and are organized in the spreadsheet in the same order that are presented in the questionnaire. Each answer must be introduced in the spreadsheet, following the definition as presented in Table 38.1.

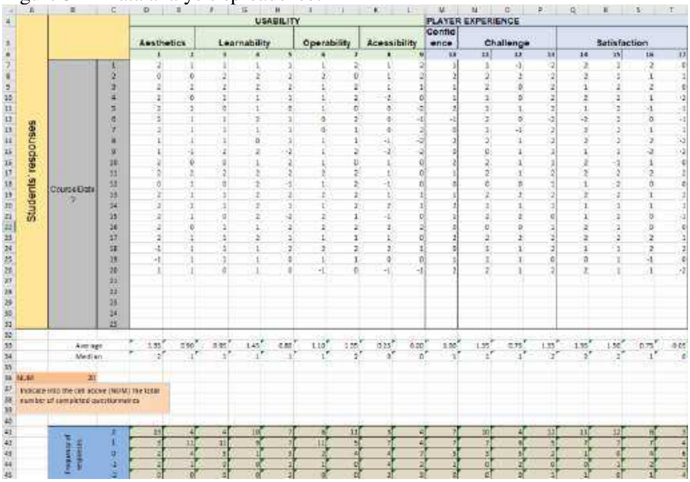
Table 38.1. Introducing an answer in the spreadsheet

Answer	Introduce into the spreadsheet
Strongly disagree	-2
Disagree	-1
Neither disagree nor agree	0

Agree	1
Strongly Agree	2

Source: developed by the author.

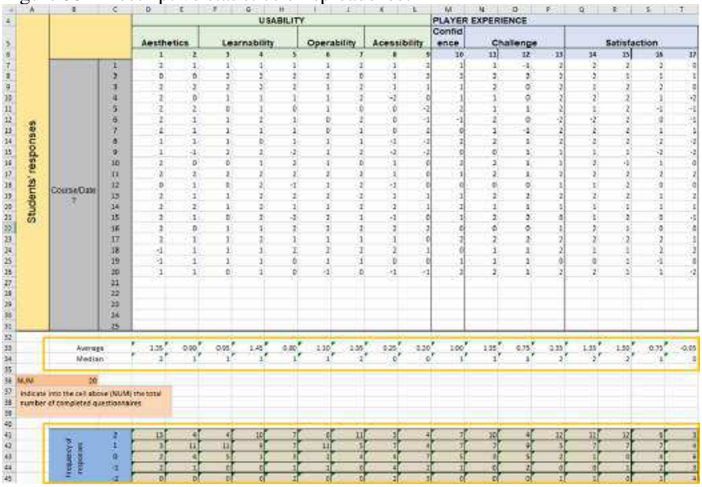
Figure 32 - Data analysis spreadsheet



Source: developed by the author.

Work products	**Input:** P3.4.1 – Data collected [In Preparation]; P4.1.1 – MEEGA+ data analysis spreadsheets and instructions guide [Initial]; The spreadsheets are available in English, Spanish, and Brazilian Portuguese at <http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/> Instructions also available in English, Spanish, and Brazilian Portuguese at <http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/>
	Output: P3.4.1 – Data collected [Ready to use]
	Output: P4.1.1 – MEEGA+ data analysis spreadsheets [In Preparation]
Activity 4.2. Check the descriptive statistics	
Description	Descriptive statistics deal with the presentation and numerical processing of a data set (WOHLIN et al., 2012). After the data collected are prepared and organized in the analysis spreadsheets, the data are characterized by automatically calculating descriptive statistics. Thus, in this activity, it is necessary only to check if the descriptive statistics were generated correctly. The spreadsheet automatically calculates the frequency of responses, average and median for each measurement instrument item, as presented in Figure 33.

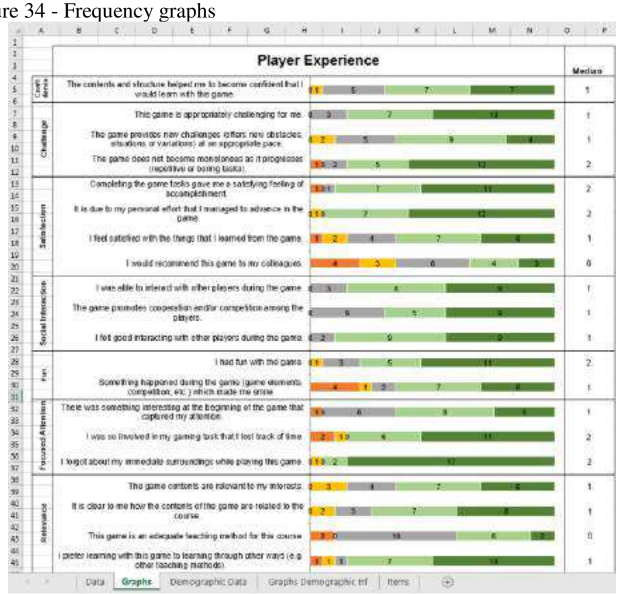
Figure 33 - Descriptive statistics on spreadsheet



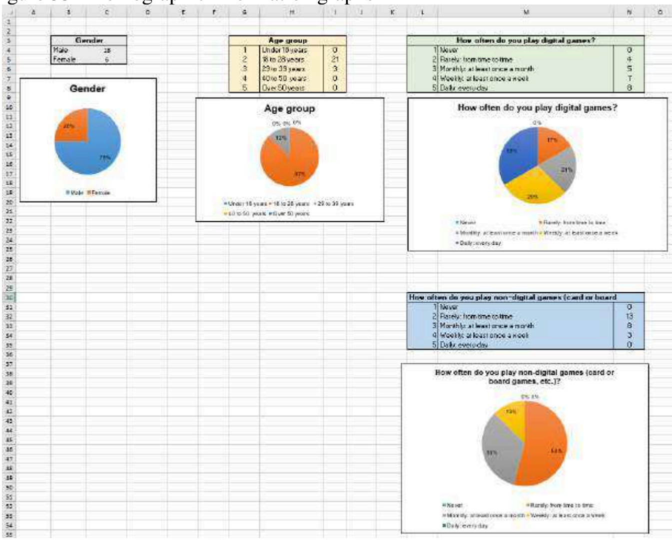
Source: developed by the author.

Graphics of frequency are also automatically generated and presented in the Graphs spreadsheet (Figure 34). The graphics are presented in accordance with the quality factors and dimensions of the MEEGA+ model.

Figure 34 - Frequency graphs



Source: developed by the author.

	Similarly, the graphics presenting the demographic information are generated and presented in the Demographic graph spreadsheet (Figure 35). Figure 35 - Demographic information graphs  Source: developed by the author.
Work products	**Input:** P3.4.1 – Data collected [Ready to use]; P4.1.1 – MEEGA+ data analysis spreadsheets [In Preparation]
	Output: P4.2.1 – Descriptive statistics results [Ready to use]
	Output: P4.1.1 – MEEGA+ data analysis spreadsheets [Ready to use]
Activity 4.3. Classify the quality level of the game	
Description	If the evaluation was conducted using the AssistantMEEGA+, this activity is not supported in the current version (v0.9). Thus, this activity needs to be performed manually. After the data collected are prepared and organized in the analysis spreadsheets, these data may be used to determine the quality level of the evaluated game, classifying it based on the game quality scale, defined in the MEEGA+ model. In order to classify the game using the MEEGA+ scale, it is necessary to follow these steps: **1. Download the files**: it is necessary to download the files of the scale (available at: http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/) and organize them into a single directory on your computer. **2. Prepare the files**: Prepare an auxiliary file (named EXTRA) in csv (comma-separated values) extension with the data collected for the items (1, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31). This file is an input for an R script which applies the defined scale scores in the data collected. This auxiliary file must be in the same directory that the other files. **3. Use RStudio:** Using the statistical software RStudio (https://www.rstudio.com/), load the defined directory and execute the R script (SCRIPT_TRI_SCORE), which calculates the scores of the data collected, applying the Item Response Theory. As a result, on the console of the RStudio, the

scores will be presented, and a new auxiliary file (SCORE_TRI_EXTRA.csv) will be generated in the directory, presenting the calculated scores and the standard error, as shown the Figure 36.

Figure 36 - Scale scores

	A	B	C	D
1		ID	SCORE_TRI	ERRO_PADRAO_TRI
2	1	N1	-1,3648	0,6962
3	2	N2	-1,351	0,6132
4	3	N3	-0,8182	0,5252
5	4	N4	1,836	0,3176
6	5	N5	1,1933	0,3155

Source: developed by the author.

4. Analyse the scores and classifying the game: The Item Response Theory calculates the score (column SCORE_TRI) (in a (0.1) scale) of an individual and positions it on the defined scale. However, we are interested in the classification of the game (and not of an individual). Thus, we must calculate the average of the provided scores (column SCORE_TRI) of all participants that evaluated the game. Based on the scores presented in Figure 36, the average of these scores is θ=-0.10. However, in order to provide a better understanding of these values, we transform these scores in a (50.15) scale, applying the following formula $\theta_{50.15}=50+15*\theta_{0.1}$. Thus, applying this formula in the average score, we obtained a value of θ=48.5. Based on this final value, we may classify the evaluated game in the MEEGA+ game scale. With a score of θ=48.5, this game is classified as a game with good quality (42.5<=θ<65). Therefore, this game typically presents the characteristics of its quality level, as described in Table 38.2.

Table 38.2. Game quality levels

Quality level	Level description
Low quality (θ < 42.5)	At this level, the game rarely provides social interaction and hardly ever produces moments of fun among the players. The game does not capture the student's focused attention, does not arouse the confidence that he/she will learn from the game, nor does it produce feelings of satisfaction. The game rarely presents challenges, has monotonous tasks and does not contribute to student learning. Although a game at this level has a low relevance to the students' interests, a student recognizes that the game's content is related to the course. In terms of usability, a game at this level sometimes exhibits operability features, which may have some clear rules and be easy to play.
Good quality (42.5 <= θ < 65)	At this level, the game sometimes presents challenging activities, offering new challenges for students. It provides moderately focused attention to the players, although students do not forget about their surroundings. Sometimes the game also provides feelings of confidence and satisfaction in the players. Frequently the game presents moments of social interaction and fun among the players. Often the game is considered relevant to the students' interests and, usually, the students recognize that the game's content is related to the course. Frequently the game contributes efficiently to student learning. In terms of usability, the game usually has the clear rules and is easy to play, although, usually does not present a fully attractive design.
Excellent quality	At this level, the game is challenging for students and has no monotonous activities. It is highly relevant to students' interests

	(θ >= 65) — and provides excellent focused attention, satisfaction, fun, and social interaction. It allows the student to be confident that he/she will learn from the game and contribute to and efficient student learning. In terms of usability, the game presents excellent operability and learnability, that is, it has clear rules and is easy to learn to play. Even so, a game at this level may present improvements in terms of aesthetics, not presenting a fully attractive design. Source: developed by the author.
Work products	**Input:** P3.4.1 – Data collected [Ready to use]; P4.1.1 – MEEGA+ data analysis spreadsheets [Ready to use]; P4.3.1 – MEEGA+ game quality scale [Ready to use] The files of the scale are available at http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/.
	Output: P4.3.2 – Game quality level [Ready to use]
Activity 4.4. Interpret the data	
Description	When the data collected are organized and characterized by descriptive statistics, it is important interpret them in order to achieve the evaluation goal, defined in the scoping phase. The analysis is performed following the quality aspects determined in the evaluation goal. Following the MEEGA+ model, the game is evaluated in terms of usability and player experience. Thus, we suggest that the data be analysed for each quality factor (usability and player experience). For each quality factor, analyses the frequency of each measurement instrument item, identifying the degree of agreement and/or disagreement of the students. Based on the degree of agreement and/or disagreement, strengths and weaknesses of the evaluated game may be also identified and reported. In addition, the quality level of the evaluated game, defined by the MEEGA+ scale, may be reported, describing the characteristics of the game in respect of its quality level.
Work products	**Input:** P3.4.1 – Data collected [Ready to use]; P4.1.1 – MEEGA+ data analysis spreadsheets [Ready to use]; P4.2.1 – Descriptive statistics results [Ready to use]; P4.3.2 – Game quality level [Ready to use]; P1.2.2 - Evaluation goal [Ready to use]
	Output: P4.4.1 – Evaluation results [Ready to use]
Activity 4.5. Discuss the results	
Description	This activity aims to discuss the findings identified in the evaluation results, indicating the main contributions of the use of the evaluated game as an instructional strategy for computing education, as well as its improvement opportunities. In addition, the results of the evaluated game may be compared to results of related studies and game evaluations, indicating the similarities and differences to other games that have similar learning objectives. Furthermore, it is important identifying threats to the study validity, as well as report mitigation strategies adopted in order to minimize the impact in the study.
Work products	**Input:** P4.4.1 – Evaluation results [Ready to use]
	Output: P4.5.1 – Discussion [Ready to use]

Source: developed by the author.

Phase 5. Presentation. In this phase the evaluation is reported (Table 39).

Table 39 - Activities of the Presentation phase

Phase 5. Presentation	
Activity 5.1. Report the evaluation	
Description	This activity aims to produce an evaluation report describing, in detail, how the evaluation of the selected game was defined, planned, executed and analysed. We focus on the academic reporting to journals and conferences, based on the definition of Jedlitschka, Ciolkowski, and Pfahl (2008) and Wohlin et al. (2012).
Work products	**Input:** P1.1.1 – Educational game [Ready to use]; P4.1.1 – MEEGA+ data analysis spreadsheets [Ready to use]; P1.2.2 - Evaluation goal [Ready to use]; P2.1.1 – Evaluation context [Ready to use]; P2.2.1 – Evaluation schedule [Ready to use]; P2.4.1 – Approval of the Ethics Committee [Ready to use]; P4.2.1 – Descriptive statistics results [Ready to use]; P4.3.2 – Game quality level [Ready to use]; P4.4.1 – Evaluation results [Ready to use]; P4.5.1 – Discussion [Ready to use]
	Output: P5.1.1 – Evaluation report [Ready to use]

Source: developed by the author.

A template for the work products (outputs) of the MEEGA+ process is available in Appendix E.

5.3 CHAPTER CONCLUSIONS

In this chapter, we present the design of the MEEGA+ method, which aims to provide a systematic support for the evaluation 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, guiding researchers on how to plan, execute and analyse the results of game evaluations.

The MEEGA+ method, presented in this chapter, is evaluated by experts and in terms of reliability and validity of its evaluation model. The next chapter presents details of the evaluation of the MEEGA+ method.

6 APLICATION AND EVALUATION OF THE MEEGA+ METHOD

This chapter presents the evaluation of the MEEGA+ method. As defined in the research method, the evaluation of the MEEGA+ method is performed in two steps. The first step (section 6.1) aims to evaluate the reliability and validity of the measurement instrument of the MEEGA+ model based on data collected through a series of case studies conducted. The second step (section 6.2) aims to evaluate the quality of the MEEGA+ method based on the experts' perspective. Section 6.3 presents the conclusions obtained in the chapter.

6.1 EVALUATION OF THE MEEGA+ MODEL

In order to evaluate the MEEGA+ model in terms of reliability and construct validity, with respect to its measurement instrument, we conduct a series of case studies. Reliability and construct validity are important issues with respect to measurement instruments such as questionnaires (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008; DEVELLIS, 2016). Reliability refers to the degree of consistency or stability of the instrument items on the same construct. Internal consistency reliability is measured in order to evaluate the consistency of results across items within a questionnaire (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008) through Cronbach's alpha coefficient (CRONBACH, 1951). Construct validity of a questionnaire is explained as its ability to actually measure what it purports to measure, including convergent and discriminant validity, which is measured through the degree of correlation between the instrument items (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008).

6.1.1 Definition of the evaluation

The study objective is defined using the GQM goal template (BASILI; CALDIERA; ROMBACH, 1994): to analyse the MEEGA+ measurement instrument in order to evaluate its reliability and construct validity from the viewpoint of the researchers in the context of computing education. Results of this study are interpreted from the researchers' perspective, the researchers are members of the Software Quality Group (GQS/INCoD/INE/UFSC), with backgrounds in computing and statistics.

Following the GQM approach, the study objective is decomposed into quality aspects and analysis questions to be analysed based on the evaluation of measurement instruments (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008; DEVELLIS, 2016):

Reliability

AQ1: Is there evidence for internal consistency of the MEEGA+ measurement instrument?

Construct Validity

AQ2: How do underlying factors influence the responses on the items of the MEEGA+ measurement instrument?

AQ3: Is there evidence of convergent and discriminant validity of the MEEGA+ measurement instrument?

6.1.2 Execution

From July 2016 to July 2018, we conducted a series of 62 case studies (44 studies using non-digital games and 18 studies using digital games), evaluating 24 different educational games (7 digital games and 17 non-digital games) using the MEEGA+ model in five computing knowledge areas: Software Engineering, Software Development Fundamentals, Algorithms & Complexity, Information Management, and Human-Computer Interaction.

The majority of the case studies (41) used games for teaching Software Engineering (13 studies using digital games and 28 studies using non-digital games). Thirteen case studies are related to teaching Algorithms & Complexity via non-digital games only. In addition, the sample includes data from four case studies for teaching Software Development Fundamentals (two studies using non-digital games and two studies using digital games), three studies using a digital game for teaching Information Management, and one study using a non-digital game for teaching Human-Computer Interaction.

In each of these case studies, the MEEGA+ measurement instrument was used after the game session (treatment) in order to collect the students' perceptions about the game. In total, responses from 1048 students in eight different educational institutions were collected as summarized in Table 40. The execution of the case studies has been formally approved by the Human Research Ethics Committee of the Federal University of Santa Catarina (Certificate No. 1.601.297/2016). At the beginning of each case study, participating students signed an informed consent, and an authorization for image use in academic publications.

Table 40 - Summary of the case studies conducted using the MEEGA+ model

Game	Game type	Game session time	Context	Course/Semester	Institution/ Country	Sample size (M - Men, W - Women, NI – Not Informed)
Computing knowledge area: Software Engineering						
PMQuiz	Digital	30 minutes	Undergraduate course in computing program	Project planning and management/2016-2	UFSC/Brazil	29 (26M, 3W)
				Project management/2016-2		20 (19M, 1W)
				Project management/2016-2	UFSM/Brazil	11 (9M, 2W)
				Project planning and management/2017-1	UFSC/Brazil	17 (16M, 1W)
				Project management/2017-1		24 (18M, 6W)
				Project planning and management/2017-2		18 (17M, 1W)
				Project planning and management/2018-1		19 (17M, 2W)
Use cases	Non-digital	30 minutes	Graduate course in computing program	SE Instructional Strategies/2016-2	UFSC/Brazil	6 (4M, 2W)
Comunica	Non-digital	30 minutes	Graduate course in computing program	SE Instructional Strategies/2016-2	UFSC/Brazil	6 (4M, 2W)
EAReqGame	Digital	30 minutes	Undergraduate course in computing program	Software Engineering/2016-2	IFC/Brazil	6 (5M, 1W)
				Software Engineering/2017-1	IFC/Brazil	41 (32M, 9W)
Unified Process Game	Non-digital	30 minutes	Graduate course in computing program	SE Instructional Strategies/2016-2	UFSC/Brazil	6 (4M, 2W)
The Classes Game	Non-digital	30 minutes	Graduate course in computing program	SE Instructional Strategies/2016-2	UFSC/Brazil	6 (4M, 2W)
PMDome	Non-digital	120 minutes	Graduate course in computing program	IT Management/2016-2	IFFar/Brazil	20 (17M, 3W)
			Undergraduate course in	PM Workshop/2016-2	IFSC/Brazil	27 (18M, 9W)

			computing program			
PMMaster	Non-digital	90 minutes	Undergraduate course in computing program	Project planning and management/2016-2	UFSC/Brazil	24 (23M, 1W)
				Project management/2016-2		21 (20M, 1W)
				Project planning and management/2017-1		17 (16M, 1W)
				Project management/2017-1		18 (13M, 5W)
				sProject planning and management/2017-2		16 (15M, 1W)
				Project planning and management/2018-1		17 (16M, 1W)
Project Detective	Non-digital	90 minutes	Undergraduate course in computing program	Project planning and management/2016-2	UFSC/Brazil	26 (23M, 3W)
				Project planning and management/2017-1		17 (16M, 1W)
				Project planning and management/2017-2		17 (16M, 1W)
QuizGame Moodle	Digital	30 minutes	Undergraduate course in computing program	Project management/2017-1	UFSC/Brazil	21 (20M, 1W)
Risk Management Game	Non-digital	90 minutes	Undergraduate course in computing program	Project management/2016-2	UFSC/Brazil	31 (28M, 3W)
				Project management/2016-2		23 (22M, 1W)
				Project management/2017-1		21 (20M, 1W)
				Project management/2017-1		21 (16M, 5W)
				Software Engineering/2017-01	IFSC/Brazil	36 (16M, 18W, 2NI)
				Software Engineering/2017-01	IFSC/Brazil	31 (18M, 12W, 1NI)
SCRUM’ed	Digital	30 minutes	Undergraduate course in computing program	Software Engineering/2016-2	ULBRA/Brazil	18 (18M)
SCRUMIA	Non-digital	90 minutes	Undergraduate course in	Project planning and management/2016-2	UFSC/Brazil	26 (24M, 2W)

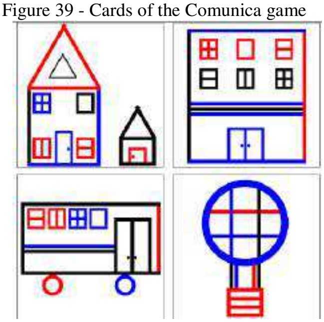
			computing program	Project planning and management/2017-1		19 (19M)
				Project planning and management/2017-2		17 (15M, 2W)
				Project planning and management/2018-1		16 (14M, 2W)
Ball Point Game	Non-digital	30 minutes	Undergraduate course in computing program	Games Workshop/2017-2	UCA/Spain	10 (9M, 1W)
Dealing with difficult people	Non-digital	30 minutes	Undergraduate course in computing program	Games Workshop/2017-2	UCA/Spain	10 (9M, 1W)
MediSoft	Non-digital	60 minutes	Undergraduate course in computing program	Project planning and management/2018-1	UFSC/Brazil	21 (20M, 1W)
ProDec	Digital	60 minutes	Undergraduate course in computing program	Software Project Management/2017-1	UCA/Spain	20 (20M)
			Graduate course in computing program	Software Project Management/2016-2		4 (3M, 1W)
Computing knowledge area: Algorithms & Complexity						
SORTIA (Heapsort)	Non-digital	90 minutes	Undergraduate course in computing program	Data Structures/2016-2	UFSC/Brazil	17 (17M)
				Data Structures/2016-2		7 (7M)
				Data Structures/2017-1		23 (22M, 1W)
				Data Structures/2017-1		6 (4M, 2W)
				Data Structures/2017-2		17 (17M)
				Data Structures/2017-2		11 (10M, 1W)
SORTIA (Quicksort)	Non-digital	90 minutes	Undergraduate course in computing program	Data Structures/2016-2	UFSC/Brazil	21 (20M, 1W)
				Data Structures/2016-2		6 (6M)
				Data Structures/2017-1		25 (23M, 2W)
				Data Structures/2017-1		6 (4M, 2W)
				Data Structures/2017-2		18 (18M)

				Data Structures/2017-2		11 (10M, 1W)
				Data Structures/2018-1		27 (25M, 2W)
Computing knowledge area: Software Development Fundamentals						
Logical Castle	Non-digital	60 minutes	Graduate course in computing program	SE Instructional Strategies/2016-2	UFSC/Brazil	6 (5M, 1W)
Save the King	Non-digital	60 minutes	Graduate course in computing program	SE Instructional Strategies/2016-2	UFSC/Brazil	6 (4M, 2W)
Kahoot! C Language	Digital	90 minutes	Undergraduate course in computing program	Programming fundamentals /2017-2	UCA/Spain	35 (30M, 5W)
			Undergraduate course in computing program	Workshop C Language/2018-1	UCA/Spain	16 (15M, 1W)
Computing knowledge area: Human-Computer Interaction						
Memorizing Heuristics	Non-digital	30 minutes	Graduate course in computing program	SE Instructional Strategies/2016-2	UFSC/Brazil	6 (5M, 1W)
Computing knowledge area: Information Management						
SQLIsland	Digital	60 minutes	Undergraduate course in computing program	Database/2018-1	UFSC/Brazil	3 (2M, 1W)
			Undergraduate course in computing program	Data Mining/2018-1	UFSC/Brazil	2 (2M)
			Graduate course in computing program	Database Tuning/2018-1	UNISUL/Brazil	5 (5M)
					Total:	1048 (910M, 135W, 3NI)

Source: developed by the author.

An overview of each of the evaluated game and its objectives are presented in Table 41.

Table 41 - Overview of the evaluated games

Computing knowledge area: Software Engineering	
PMQuiz (PETRI et al., 2016) is a digital quiz game to review Software Project Management (SPM) knowledge. Players answer each question within a time limit using their smartphone. The player who scores most points by answering more questions correctly in the shortest time is the winner. The learning objective of this game is to review SPM concepts such as scope, time, and cost. The game is available at <http://www.gqs.ufsc.br/kahoot-pm-quiz/>.	Figure 37 - Applications of the game PMQuiz 
Use cases is a non-digital game that aims to develop use cases diagrams using cards that represent the objects of diagrams such as actors, use cases, relationships, etc. Students are divided into two groups in order to develop diagrams corresponding to each system (virtual library, virtual bookstore and virtual air travel agency). The first group that complete a use case diagram is the winner. The learning objective of this game is to reinforce and apply concepts related to use case diagrams.	Figure 38 - Cards of the Use cases game 
Comunica is a card, paper & pencil game that aims to show the importance of communication among the stakeholders of a project. It simulates the communication of customer, requirements analyst and programmer. The customer must verbally and gestically transmit to the requirements analyst the message he received in the form of a drawing. The requirements analyst should try to understand and write what the customer transmitted and then pass on to the programmer what he understood. The programmer should draw the final product based on what was informed by the requirements analyst. At the end of the game, the original is compared with the final product. The learning objective is to how the importance of communication in a project and improve the communication skills.	Figure 39 - Cards of the Comunica game 
EAReqGame (PETRI; CHIAVEGATTI, 2015) is an RPG (Role Playing Game) in which the player acts as a Requirements Engineer, collecting and prioritizing requirements in a simulated scenario, seeking to progress in the phases of the game to elicit and prioritize the essential requirements for the development of an academic system. The learning objectives of the game are to apply requirements elicitation using the scenario analysis technique; and to understand and apply the prioritization of the requirements collected in	Figure 40 - Interfaces of the EAReqGame 

the scenario, classifying them as mandatory, desirable and out of scope.

Unified Process (UP) Game is a board game to teach the principles, phases and disciplines of the UP. The game includes questions about UP principles, phases and disciplines and for each correct answer the player receives a card. The player who collects cards from all UP principles, phases and disciplines will win the game by going through the beginning again. Its learning objective is to identify and recognize the principles, phases and disciplines of the Unified Process.

Figure 41 - Gameboard of the UP Game

The Class Game is a board game that involves concepts related to Class Diagram. In the game, each player receives a (different) goal, which is to build a certain Class Diagram. In order to achieve the objective, the player has to collect on the board the items that make up his diagram to be built. The player who first collects the items that make up his class diagram wins. The learning objective of the game is to understand and apply concepts related to class diagrams.

Figure 42 - Gameboard of the Class Game

PMDome (PMDOME, 2017) is a simulation game with the learning objective to motivate the importance of planning in a project. The game simulates Project Management (PM) planning and execution phases asking students to plan time and resources and then executing the project, constructing a Geodesic Dome, using pens, and sheets of paper, etc.

Figure 43 - Applications of the game PMDome

PM Master (GRESSE VON WANGENHEIM, 2012) is a board game with questions about different PM knowledge areas, such as scope, time, and quality management. The player who first correctly answers a question from each of the knowledge areas wins the game. It aims to review and reinforce basic PM concepts in accordance with the PMBOK (4th edition), focusing specifically on SPM.

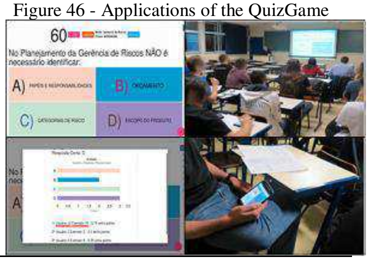
Figure 44 - Applications of the game PM Master

Detective Game – what killed the project? (GRESSE VON WANGENHEIM et al., 2014) is a deductive table top game with the objective to monitor and control a software project using earned value management. The game takes place in the context of a hypothetical software company that finished a software project that. Players receive a set of project documents to analyse and revise its monitoring and control by completing calculations of schedule and cost variances as well as performance indexes. The learning objective of the game is to reinforce concepts and to teach the competency to apply knowledge on earned value management.

Figure 45 - Applications of the game Detective

QuizGame (TONUSSI: HAUCK, 2017) is an educational question and answer game integrated into the Moodle platform. The learning objective of the game is to review and reinforce SPM concepts. The students answer questions using their personal mobile devices. The player who answers more questions correctly in the shortest time is the winner.

Figure 46 - Applications of the QuizGame



Risk Management Game is a board game and aims to motivate the importance of risk planning in PM. In the game, players must arrive at the delivery of a hypothetical project (end of the board), planning and passing through the project phases with the allocated resources. Starting the game, players need to perform a risk analysis, identifying how unexpected risks may affect the project’s resources. The player who reaches the end of the game first with financial resources left is the winner.

Figure 47 - Applications of the Risk Management Game

SCRUM’ed (SCHNEIDER, 2015) is a 3D role-playing game with a narrative based on Scrum concepts (roles, ceremonies, and artefacts). The learning objective is to reinforce and exemplify Scrum concepts. Players assume the role of a Scrum Master and need to help “knights” in planning and executing the project for their client, the King. Players are taken on a journey through a Daily Scrum, helping the team to define and update the Sprint Backlog and the task-board based on the Product Backlog, running a Sprint Review Meeting, with the goal to keep the project on schedule. The game is available at <http://www.gqs.ufsc.br/scrumed-a-3d-role-playing-game-to-learn-scrum/>.

Figure 48 - Applications of the game SCRUM’ed

SCRUMIA (GRESSE VON WANGENHEI; SAVI; BORGATTO, 2013) is a group simulation exercise with the purpose of planning and executing sprints of a hypothetical project by applying SCRUM as part of an SPM course. Its learning objective is to reinforce the concepts and to teach the competency to apply agile project management using SCRUM.

Figure 49 - Applications of the game SCRUMIA

Ball Point Game (GLOGER, 2017) aims to illustrate the dynamics of a team working iteratively focusing on continuous improvement. In the game, teams must pass balls by all participants until they return to the starting point to score points. The balls cannot be passed to the neighbours on the right and left, they cannot touch the ground and the balls must pass through the air to the other player. At the end of the game, the team that scored the most points is the winner.

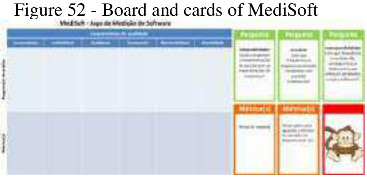
Figure 50 - Applications of the Ball Point Game

Dealing with difficult people (GRESSE VON WANGENHEIM; CARVALHO; BATTISTELLA, 2013) is a simulation that aims to illustrate difficulties related to team management in software projects. In small groups, the game simulates a kick-off meeting of a software project. One of the students assumes the role of the project manager, who conducts the meeting and at the end must have received the agreement of all other group members. Each member of the group assumes a difficult personality (e.g., no person, know-it-all, grenade), while the project manager must react accordingly in order to manage these difficult personalities.

Figure 51 - Applications of the game Dealing with difficult people

MediSoft is a board and card game to illustrate the measurement planning in the context of quality control in SPM courses. At the beginning of the game, players receive a board indicating quality characteristics, defined based on the ISO9126 and ISO9421, and two cards indicating analysis questions. During the game, players need to complete the board by correctly decomposing the quality characteristics, their analysis questions and respective metrics, in order to define a measurement program. The player that first completes the board, win the game. The learning objective of the game is to reinforce and understand the process of measurement planning.

Figure 52 - Board and cards of MediSoft



ProDec (CALDERÓN; RUIZ, 2013) is a simulation-based game to teach SPM concepts, as well as supporting the comprehension and knowledge acquisition of SPM lifecycle processes. Its aim is that learners are able to successfully manage a software project. Player win when s/he is able to complete the project within the time and cost. It allows learners to create their own project plans from scratch, execute the simulation of their own project scenarios, practice their decision-making skills by controlling and monitoring the progress of the project execution, and close the project analysing the results of their performance during gameplay to get lessons learned.

Figure 53 - Applications of the game ProDec

Computing knowledge area: Algorithms & Complexity

SORTIA (Quicksort/Heapsort) (BATTISTELLA et al., 2017) is a board game that can be used to represent the memory model of a sequence of recursive calls of the Quicksort algorithm and to simulate the execution of the steps of the Heapsort algorithm. The game is designed to be played in groups of 4 to 6 students. In the game, students need to order 21 values (random numbers from 0-100) in an array, accordingly to the Quicksort/Heapsort algorithms. The learning objective of the game is to understand and apply the Quicksort/Heapsort algorithms.

Figure 54 - Applications of the game SORTIA

Computing knowledge area: Software Development Fundamentals

Logical Castle is a board game developed to teach programming in logic in a fun way. The game aims to teach the structure of a computer program, the sequence of procedures, Boolean logic, conditions, repetitions, and functions. The learning objective of the game is to reinforce and understand concepts related to the fundamentals of programming.

Figure 55 - Application of the game Logical Castle

Save the King is a non-digital game that mixes the board and chess pieces with the fun and dynamism of the UNO card game. Each player moves their tower as they play the number cards. The objective is to describe and execute a code, using the board houses, to save the king and to add more points than the opponents collecting the pieces of the board. The learning objective is to reinforce and understand algorithm concepts.

Figure 56 - Board and cards of the game Save the King

Kahoot! C Language is a quiz game to review knowledge of the C programming language. Using smartphones players answer questions within a time limit to complete C commands in a determined algorithm. The player who scores most points by answering more questions correctly in the shortest time is the winner. The learning objective of this game is to reinforce and review concepts of the C programming language.

Figure 57 - Applications of the game Kahoot! C Language

Computing knowledge area: Human-Computer Interaction

Memorizing Heuristics is an adaptation of the traditional memory game. 40 cards are used, with 10 Nielsen' usability heuristics and 30 usability problems. The cards are shuffled and placed on the table face down. When a problem is flipped together with their heuristic they can be stacked by the player. The game ends when there are no more cards to turn. As learning objectives, students should be able to remember heuristics and associate usability problems with their heuristics.

Figure 58 - Cards of the game Memorizing Heuristics

Computing knowledge area: Information Management

SQLIsland (SCHILDGEN; DEßLOCH, 2015) is a digital adventure game. In the beginning of the game, the player is marooned on an island and encounters the inhabitants that only understand the database language SQL. Luckily, they are in the company of a pilot who could help you escape, but he is being held captive. The player's task is to earn money, sell items and buy a sword which will help him to free the pilot from the fetters of the kidnapper. The player controls a fictional character by using SQL queries in order to achieve the game goal. The learning objective of the game is to understand and apply concepts of the SQL language. The game is available at: <http://wwwlgis.informatik.uni-kl.de/extra/game/>.

Figure 59 - Interface of the game SQLIsland

Source: developed by the author.

Data collected in the case studies were pooled in a single sample (*n*=1048), using them cumulatively in order to evaluate the MEEGA+ model (and not a specific game). The pooling of data was possible due to the similarity of the case studies and standardization of the data collected using the MEEGA+ model. The case studies are similar in terms of definition (with the objective to evaluate an educational game), research design (case studies), and context (computing education). In addition, all case studies are standardized in terms of measures, data collection method

(MEEGA+ measurement instrument), and response format (5-point Likert scale).

6.1.3 Analysis

In order to evaluate the MEEGA+ model, we performed a statistical evaluation using the IBM SPSS Statistics trial version 23 for data analysis. The evaluation follows the approach for the scale development as proposed by DeVellis (2016) in alignment with procedures for the evaluation of internal consistency and construct validity of a measurement instrument (TROCHIM; DONNELLY, 2008).

Reliability

AQ1: Is there evidence for internal consistency of the MEEGA+ measurement instrument?

Evaluation of the standardized items. We measured the internal consistency of the MEEGA+ measurement instrument through the Cronbach's alpha coefficient (DEVELLIS, 2016; TROCHIM; DONNELLY, 2008). Cronbach's alpha coefficient (CRONBACH, 1951) indicates the degree to which a set of items measure a single factor. Thus, here, we want to know whether the MEEGA+ measurement instrument measures the same quality factor, the quality of the educational game. Typically, values of Cronbach's alpha between $0.8 > \alpha \geq 0.7$ are acceptable, between $0.9 > \alpha \geq 0.8$ are good, and $\alpha \geq 0.9$ are excellent (DEVELLIS, 2016), thus, indicating an internal consistency of the instrument.

Analysing the 31 standardized items of the MEEGA+ measurement instrument, the value of Cronbach's alpha is excellent (α = .927). Results of analysis of each of the quality factor show that Cronbach's alpha is also satisfactory (Table 42).

Table 42 - Cronbach’s alpha for standardized items per quality factor

Quality factor	Cronbach's alpha
Usability	.841
Player Experience	.932
Total	.927

Source: developed by the author.

Furthermore, we also analysed the Cronbach's alpha if an item was deleted (Table 43), expecting that no item (if deleted) should cause a substantial increase in the Cronbach's alpha (DEVELLIS, 2016).

Table 43 - Cronbach's alpha coefficients if item deleted

Item No.	Description	Cronbach's alpha, if item was deleted
1	The game design is attractive (interface, graphics, cards, boards, etc.).	.925
2	The text font and colors are well blended and consistent.	.926
3	I needed to learn a few things before I could play the game.	.927
4	Learning to play this game was easy for me.	.926
5	I think that most people would learn to play this game very quickly.	.926
6	I think that the game is easy to play.	.925
7	The game rules are clear and easy to understand.	.925
8	The fonts (size and style) used in the game are easy to read.	.928
9	The colors used in the game are meaningful.	.927
10	The contents and structure helped me to become confident that I would learn with this game.	.923
11	This game is appropriately challenging for me.	.924
12	The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace.	.924
13	The game does not become monotonous as it progresses (repetitive or boring tasks).	.924
14	Completing the game tasks gave me a satisfying feeling of accomplishment.	.922
15	It is due to my personal effort that I managed to advance in the game.	.925
16	I feel satisfied with the things that I learned from the game.	.922
17	I would recommend this game to my colleagues.	.922
18	I was able to interact with other players during the game.	.925
19	The game promotes cooperation and/or competition among the players.	.924
20	I felt good interacting with other players during the game.	.923
21	I had fun with the game.	.921
22	Something happened during the game (game elements, competition, etc.) which made me smile.	.924
23	There was something interesting at the beginning of the game that captured my attention.	.924
24	I was so involved in my gaming task that I lost track of time.	.923
25	I forgot about my immediate surroundings while playing this game.	.924
26	The game contents are relevant to my interests.	.924

27	It is clear to me how the contents of the game are related to the course.	.926
28	This game is an adequate teaching method for this course.	.924
29	I prefer learning with this game to learning through other ways (e.g. other teaching methods).	.925
30	The game contributed to my learning in this course.	.923
31	The game allowed for efficient learning compared with other activities in the course.	.924

Source: developed by the author.

Based on the results of the reliability analysis, we can conclude that the answers to the items are consistent and precise, indicating the reliability of the standardized items of the MEEGA+ measurement instrument.

Evaluation of the Customized Items. Results of the reliability analysis of the items that are customized in accordance with the specific learning goals of each evaluated educational game, show that Cronbach's alpha is excellent also for the customized items (α = .953). This also indicates an excellent reliability of this part of the MEEGA+ measurement instrument.

Construct Validity

AQ2: How do underlying factors influence the responses on the items of the MEEGA+ measurement instrument?

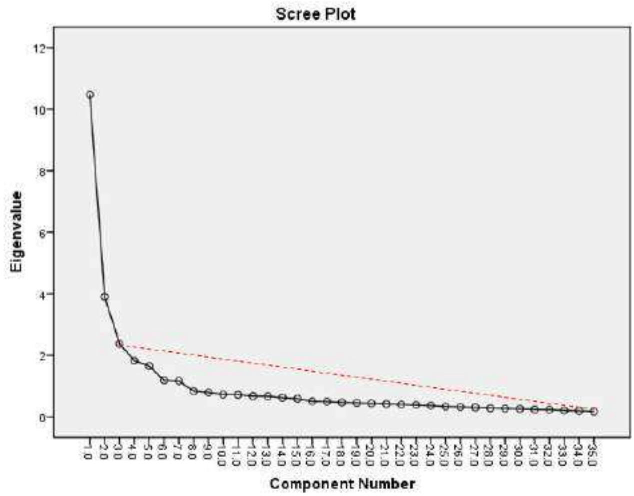
In order to confirm the number of quality factors that represent the responses of the 31 standardized items of the MEEGA+ measurement instrument, we performed an exploratory factor analysis. Based on the definition of the MEEGA+ model we assume that it is influenced by two underlying factors (usability and player experience).

In order to analyse whether the MEEGA+ measurement instrument items can be submitted to factor analysis process (BROWN, 2006), we used the Kaiser-Meyer-Olkin (KMO) index and Bartlett's test being the most commonly used ones (BROWN, 2006). These methods indicate how much the realization of the exploratory factor analysis is appropriate for a specific set of items (BROWN, 2006). The KMO index measures the sampling adequacy with values between 0 and 1. An index value near 1.0 supports a factor analysis and anything less than 0.5 is probably not amenable to useful factor analysis (DZIUBAN; SHIRKEY, 1974). The Bartlett's test also indicates whether the factor analysis is appropriate, considering acceptable values of a significance level <0.05 (DZIUBAN; SHIRKEY, 1974). Analysing the MEEGA+ measurement instrument items, we obtained a KMO index of .928 and a significance level of 0.000.

Therefore, indicating that factor analysis is appropriate to analyse the number of factors that represents the responses of the MEEGA+ measurement instrument.

Applying the factorial analysis, the number of factors retained in the analysis is decided (GLORFELD, 1995; BROWN, 2006). Here we used the Cattell's scree test (CATTELL, 1966; CATTELL, 1978) for this decision, as is one of the most used strategies to determine the number of components to retain (RAICHE et al., 2013). The Cattell's scree test plots the components (items) as the X-axis and the corresponding eigenvalues as the Y-axis. The Cattell's scree test involves plotting the eigenvalues in descending order of their magnitude against their component numbers and determining where they level off. The break between the steep slope and a levelling off (named the elbow) indicates the number of meaningful factors. Thus, the Cattell's scree test disregards all further components after the one starting the elbow (CATTELL, 1966; CATTELL, 1978). Based on this definition, the scree plot (Figure 60) shows that the quick change of the slope of the curve (the elbow) appears in the third factor. In addition, the red dotted line (Figure 60) shows that the drop of the curve is less abrupt (levelling off) from the third factor. This justifies the retention of two factors in our analysis.

Figure 60 - Scree plot



Source: developed by the author.

However, sometimes the Cattell's scree test is criticized as a researcher bias may be introduced due to the subjectivity involved in determining the number of components to retain using a visual inspection device (scree test) (GARSON, 2013). In order to minimize this bias, we also analysed the percent of cumulative variance of the factors as a complementary method for the decision on the number of factors retain in the analysis. In this respect, the two factors retained in our analysis explain 43.60% of the cumulative variance. This variance is considered as adequate for the decision in the retention of the factors, considering 40% the threshold of the variance as acceptable (CUESTA, 1996). With respect to the MEEGA+, this means that the responses of the measurement instrument are representing two underlying concepts (quality factors).

Once identified the number of underlying factors, another issue is to determine which items are loaded into which factor. In order to identify the factor loadings of the items, we used the Varimax with Kaiser Normalization rotation method, being the most widely used rotation method (TABACHNICK; FIDELL, 2007). Table 44 shows the factor loadings of the items associated with the two retained factors. The highest factor loading of each item, indicating to which factor the item is most related, is marked in bold.

Table 44 - Factor loadings

Quality factor	**Item No.**	**Description**	**Factors**	
			1	**2**
Usability	**1**	The game design is attractive (interface, graphics, cards, boards, etc.).	.326	**.508**
	2	The text font and colors are well blended and consistent.	.164	**.525**
	3	I needed to learn a few things before I could play the game.	.028	**.684**
	4	Learning to play this game was easy for me.	.056	**.796**
	5	I think that most people would learn to play this game very quickly.	.105	**.791**
	6	I think that the game is easy to play.	.111	**.798**
	7	The game rules are clear and easy to understand.	.260	**.668**
	8	The fonts (size and style) used in the game are easy to read.	.012	**.515**
	9	The colors used in the game are meaningful.	.067	**.506**
Player experience	**10**	The contents and structure helped me to become confident that I would learn with this game.	**.549**	.353
	11	This game is appropriately challenging for me.	**.657**	-.009

	12	The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace.	**.697**	-.037
	13	The game does not become monotonous as it progresses (repetitive or boring tasks).	**.585**	.095
	14	Completing the game tasks gave me a satisfying feeling of accomplishment.	**.747**	.108
	15	It is due to my personal effort that I managed to advance in the game.	**.492**	.194
	16	I feel satisfied with the things that I learned from the game.	**.735**	.194
	17	I would recommend this game to my colleagues.	**.736**	.224
	18	I was able to interact with other players during the game.	**.574**	-.038
	19	The game promotes cooperation and/or competition among the players.	**.610**	.036
	20	I felt good interacting with other players during the game.	**.706**	.052
	21	I had fun with the game.	**.772**	.202
	22	Something happened during the game (game elements, competition, etc.) which made me smile.	**.617**	.150
	23	There was something interesting at the beginning of the game that captured my attention.	**.634**	.118
	24	I was so involved in my gaming task that I lost track of time.	**.686**	.067
	25	I forgot about my immediate surroundings while playing this game.	**.658**	.058
	26	The game contents are relevant to my interests.	**.573**	.199
	27	It is clear to me how the contents of the game are related to the course.	**.404**	.239
	28	This game is an adequate teaching method for this course.	**.595**	.249
	29	I prefer learning with this game to learning through other ways (e.g. other teaching methods).	**.519**	.171
	30	The game contributed to my learning in this course.	**.661**	.178
	31	The game allowed for efficient learning compared with other activities in the course.	**.609**	.182

Source: developed by the author.

Analysing the factor loadings of the items (Table 44), we can observe that the first factor (factor 1), consists of a set of 22 items (items no. 10 to 31), including items from the dimensions of focused attention, fun, challenge, social interaction, confidence, relevance, satisfaction, and perceived learning. Thus, this result seems to confirm that these items (and their original dimensions), in fact, are related to measuring the quality of educational games in terms of player experience. Therefore, this result indicates that the quality factor of player experience, when evaluating educational games for computing education, is composed of the dimensions of confidence, challenge, satisfaction, social interaction, fun, focused attention, relevance, and perceived learning.

With respect to factor 2, a set of 9 items (items no. 1 to 9) is related. This result, clearly, suggests that factor 2 is related to the concept of usability, measuring the aesthetics, learnability, operability, and accessibility of educational games. Again, this result seems to confirm the original structure of the MEEGA+ model also regarding to the quality factor of usability.

In summary, the results of the exploratory factor analysis seem to be confirming the quality factors of the original structure of the MEEGA+ model, indicating that the quality of educational games is evaluated in terms of usability and player experience.

AQ3: Is there evidence of convergent and discriminant validity of the MEEGA+ measurement instrument?

In order to obtain evidence of convergent and discriminant validity of the standardized items of the MEEGA+ measurement instrument, the correlations of the items are calculated (DEVELLIS, 2016). Convergent and discriminant validity are considered two subcategories of construct validity (TROCHIM; DONNELLY, 2008). Convergent validity shows that the items that should be related are, in fact, related. On the other hand, discriminant validity shows that the items that should not be related are, in fact, not related (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008).

In order to obtain evidence of convergent validity it is expected that the items of the same quality factor (e.g., usability or player experience) demonstrate a large correlation (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008). On the other hand, to obtain evidence of discriminant validity it is expected that items of different quality factors demonstrate a small correlation (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008). For example, it is expected that the items of different quality factors (e.g., usability and player experience) demonstrate a small correlation, as in theory, the items are measuring different quality factors.

In order to analyse the correlations between the standardized items, we used the nonparametric Spearman correlation matrices for each quality factor (Tables 45 and 46). A complete matrix including all quality factors is presented in Appendix F. The matrices show the Spearman correlation coefficient, indicating the degree of correlation between two items (item pairs). We used this correlation coefficient, as it is the most appropriate correlation analysis for Likert scales (CHEN; POPOVICH, 2002). The correlation coefficients between the items within of the same dimension are coloured. In accordance with Cohen (COHEN, 1988), a correlation between items is considered satisfactory, if the correlation coefficient is greater than 0.29, indicating that there is a moderate correlation, or a large correlation, if the coefficient is greater than 0.50 (COHEN, 1988). A coefficient of about 0.10 indicates a small correlation between the items (COHEN, 1988). Satisfactory correlations are marked in bold.

Analysing the correlations between the items of the quality factor usability (Table 45), we can observe that most of the item pairs present a moderate or large correlation. This result indicates that, although usability is fragmented into its dimensions, the items present a satisfactory correlation in order to measure what the factor purports to measure (the usability of the educational games). On the other hand, few items pairs (7) presented an unsatisfactory correlation according to Cohen's coefficient. Even so, these item pairs have a correlation coefficient very close to the level of moderate correlation. Therefore, based on the correlation coefficients of the quality factor usability, we can establish a convergent validity.

Table 45 - Spearman correlation coefficient of the quality factor Usability

Item/ Dimension	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	Aesthetics		**Learnability**			**Operability**		**Accessibility**	
1	**1.00**								
2	**.58**	**1.00**							
3	**.29**	.24	**1.00**						
4	**.31**	.28	**.68**	**1.00**					
5	**.36**	.25	**.59**	**.71**	**1.00**				
6	**.36**	**.31**	**.55**	**.70**	**.72**	**1.00**			
7	**.36**	**.29**	**.47**	**.54**	**.54**	**.61**	**1.00**		
8	**.38**	**.50**	.27	**.33**	**.30**	**.38**	**.38**	**1.00**	
9	**.42**	**.63**	.20	.28	.27	**.34**	**.33**	**.65**	**1.00**

Source: developed by the author.

With respect to the quality factor of player experience (Table 46), we can observe again that the majority of the item pairs presents a moderate or a large correlation coefficient. This result indicates that, in fact, the player experience in educational games is measured in terms of confidence, challenge, satisfaction, social interaction, fun, focused attention, relevance, and perceived learning as suggested by the large correlation coefficients between the items. Therefore, considering the majority of moderate and large correlation coefficients, we can observe that there is evidence of convergent validity also in the quality factor of player experience.

Table 46 - Spearman correlation coefficient of the quality factor Player Experience

Item	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	Confidence	Challenge			Satisfaction				Social Interaction			Fun		Focused Attention			Relevance				Perceived	
10	1.00																					
11	0.40	1.00																				
12	0.36	0.59	1.00																			
13	0.27	0.46	0.50	1.00																		
14	0.40	0.48	0.47	0.46	1.00																	
15	0.38	0.36	0.31	0.27	0.46	1.00																
16	0.48	0.44	0.44	0.35	0.56	0.48	1.00															
17	0.47	0.44	0.43	0.46	0.57	0.39	0.68	1.00														
18	0.28	0.32	0.34	0.34	0.37	0.20	0.37	0.40	1.00													
19	0.29	0.37	0.39	0.37	0.45	0.26	0.39	0.42	0.62	1.00												
20	0.36	0.39	0.40	0.40	0.48	0.26	0.45	0.49	0.69	0.69	1.00											
21	0.41	0.39	0.43	0.48	0.58	0.31	0.49	0.62	0.50	0.56	0.63	1.00										
22	0.31	0.32	0.30	0.39	0.45	0.22	0.38	0.47	0.43	0.44	0.55	0.67	1.00									
23	0.38	0.36	0.38	0.34	0.43	0.26	0.40	0.50	0.31	0.33	0.42	0.53	0.46	1.00								
24	0.38	0.36	0.38	0.43	0.49	0.25	0.38	0.49	0.35	0.40	0.48	0.58	0.49	0.58	1.00							
25	0.28	0.31	0.35	0.38	0.47	0.24	0.35	0.48	0.39	0.40	0.48	0.53	0.44	0.51	0.76	1.00						
26	0.39	0.38	0.29	0.29	0.45	0.36	0.52	0.51	0.27	0.30	0.34	0.44	0.29	0.36	0.34	0.34	1.00					
27	0.29	0.25	0.20	0.20	0.27	0.30	0.39	0.34	0.22	0.25	0.27	0.34	0.29	0.22	0.20	0.18	0.48	1.00				
28	0.43	0.33	0.31	0.27	0.39	0.37	0.50	0.48	0.22	0.29	0.34	0.45	0.37	0.34	0.29	0.30	0.48	0.65	1.00			
29	0.29	0.26	0.24	0.26	0.31	0.29	0.38	0.38	0.20	0.25	0.30	0.38	0.31	0.28	0.26	0.32	0.32	0.41	0.62	1.00		
30	0.46	0.37	0.39	0.26	0.43	0.37	0.58	0.49	0.26	0.30	0.34	0.44	0.33	0.36	0.33	0.30	0.53	0.53	0.65	0.53	1.00	
31	0.41	0.34	0.37	0.28	0.41	0.36	0.53	0.47	0.22	0.26	0.33	0.44	0.32	0.34	0.31	0.28	0.44	0.43	0.56	0.55	0.73	1.00

Source: developed by the author.

In order to obtain evidence of discriminant validity, we analysed the correlation coefficients of the different quality factors. Here, we are evaluated, if the items of different quality factors demonstrate a small correlation (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008). As shown in the Spearman correlation matrix (Appendix F), the majority of the item pairs of different quality factors (e.g., 3-18 and 9-25) presented a small correlation coefficient, thus, indicating evidence of discriminant validity.

In summary, we can observe that, in general, there is a correlation between items within a quality factor. This indicates that convergent validity can be established for the two quality factors (usability and player experience). In addition, in general, items of different quality factor presented a small correlation, thus, we also can identify evidence of discriminant validity. Based on these results, we can conclude that the MEEGA+ measurement instrument actually measures what it purports to measure (the quality of games for computing education in terms of usability and player experience).

6.1.4 Discussion

The results of the statistical analysis of the MEEGA+ measurement instrument show sufficient evidence to consider the reliability and construct validity of MEEGA+ acceptable as a model for the evaluation of games used for computing education.

In terms of reliability (AQ1), the results of the analysis indicate a satisfactory Cronbach's alpha for all quality factors (Cronbach's alpha α=.927), indicating the internal consistency of the MEEGA+ measurement instrument. The reliability of the MEEGA+ model increased in comparison to the initial version of the MEEGA model (Cronbach's alpha α=.915) (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017). This indicates that the items of the MEEGA+ measurement instrument are consistent and precise with respect to the evaluation of games' quality and that the evolution of the model, in fact, presents an improvement with respect to its reliability. Also, comparing our results with the reliability reported by some of the related approaches, the evaluation of MEEGA+ demonstrated a greater reliability than that presented in each factor of the EGameFlow scale (Cronbach's alpha α=.80) (FU; SU; YU, 2009).

In terms of construct validity, based on the results of an exploratory factor analysis (AQ2), we identified that the responses of the MEEGA+ measurement instrument are explained by two underlying concepts (quality factors), representing 43.60% of the cumulative variance. This result indicates that the quality of games for computing education is evaluated through two quality factors, confirming the original structure of the MEEGA+ model, in terms of usability and player experience. The factor measuring the player experience consists of a set of 22 items covering the dimensions focused attention, fun, challenge, social interaction, confidence, relevance, satisfaction, and perceived learning. Items related to the factor measuring the game's usability, measure the dimensions aesthetics, learnability, operability, and accessibility of educational games.

Different to the results of the factor analysis of the initial version of the MEEGA model (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017), which indicated a conceptual overlap in terms of motivation and user experience, our results indicate and confirm a well-defined structure of the MEEGA+ model composed of two quality factors (player experience and usability) in order to evaluate games for computing education. And, in comparison to the related approaches, which typically use a wide variety of factors and some of them overlap conceptually, our results provide evidence of a well-defined conceptual structure. The dimensions have been defined following a trend identified in a previous literature review (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017), and a mapping study, thus, minimizing the overlap of concepts.

Analysing the correlation coefficients between the items of the two quality factors of the MEEGA+ model (AQ3), we can observe that, in general, there is a moderate and large correlation between the majority of the items within each quality factor. This indicates that a convergent validity can be established for the two quality factors (usability and player experience). In the same way, items of different quality factors present a small correlation, and, thus, provide evidence of discriminant validity. And, compared with the initial version of the MEEGA model, the MEEGA+ model again presents an improvement, as, although the MEEGA model demonstrated evidence of convergent validity, it did not present evidence of discriminant validity, indicating a conceptual overlap (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017). In contrast, with respect to the MEEGA+ model, we can identify evidence of convergent and discriminant validity, presenting a well-defined and correlated structure, composed of two quality factors (player experience and usability).

Therefore, based on our results, we can conclude that the MEEGA+ measurement instrument has a high reliability and measures what it purports to measure: the quality of games used for computing education in terms of usability and player experience.

Threats to validity. As any kind of research, this study has limitations and it is subject to threats to validity. We, therefore, identified potential threats and applied mitigation strategies in order to minimize their impact on our research.

Some threats are related to the design of the study. In order to mitigate this threat, we defined and documented a systematic research method. The MEEGA+ model has been defined adopting the GQM approach, systematically decomposing the evaluation objective. The measuring instrument has been developed following the scale and questionnaire development guides. In addition, for the evaluation of the MEEGA+ measuring instrument, a case study has been systematically defined and documented.

Another threat refers to the quality of the data pooled into a single sample, in terms of standardization of data (response format) and adequacy of the MEEGA+ model. As our study is limited exclusively to evaluations that used the MEEGA+ model this risk is minimized as in all studies the same data collection instrument has been used. Another issue refers to the pooled data from different contexts. To mitigate this threat all case studies have been conducted in similar contexts (higher computing education).

A limitation of our study refers to the evaluating the learning. Adopting a non-experimental research design (case studies), only a post-test using self-assessment has been applied in order to evaluate the students' perceived learning. A pre-test has not been applied and, therefore, it was not possible to accurately the learning difference promoted by the games. However, regarding the self-assessment, although there is no consensus, there is evidence that self-assessment provides reliable, valid and useful information for this type of study (SITZMANN et al., 2010; THOMAS; MARTIN; PLEASANTS, 2011; SHARMA et al., 2016), mainly when using a systematic, reliable, and valid evaluation model as MEEGA+.

In terms of external validity, a threat to the possibility to generalize the results is related to the sample size and diversity of the data used for the evaluation. In respect to sample size, our evaluation used data collected from 62 case studies evaluating 24 different educational games, involving a population of 1048 students from 8 different institutions. In terms of statistical significance, this is a satisfactory sample size, allowing the generation of significant results (WOHLIN et al., 2012).

In terms of reliability, a threat refers to what extent the data and the analysis are dependent on the specific researchers. In order to mitigate this threat, we systematically documented the evaluation of the MEEGA+ model, defining clearly the study objective, the process of data collection, and the statistical methods used for data analysis. Another issue refers to the correct choice of statistical tests for data analysis. To minimize this threat, we performed a statistical evaluation following the guide for the construction of measurement scales as proposed by DeVellis (2016), which is aligned with procedures for the evaluation of internal consistency and construct validity of a measurement instrument (TROCHIM; DONNELLY, 2008).

6.2 EVALUATION OF THE MEEGA+ METHOD

This section presents the quality evaluation of the MEEGA+ method based on the experts' perspective. In this study, quality is determined as the degree to which a model, component, and/or process meets specified requirements and user needs to a specific objective (IEEE 2002; 2010). Thus, in the context of this evaluation, it is expected that the MEEGA+ method provides a systematic support, covering the user (researchers) needs to conduct a quality evaluation of games for computing education.

6.2.1 Definition

Adopting the GQM goal template (BASILI; CALDIERA; ROMBACH, 1994) the study objective is defined: to analyse the MEEGA+ method in order to evaluate its quality from the viewpoint of experts in educational games in the context of computing education.

Following the GQM approach, the study objective is decomposed into quality aspects, analysis questions and metrics to be analysed. The quality aspects are defined based on the quality characteristics of a model and/or process such as: validity (JOHNSON; CHRISTENSEN, 2016), correctness (RITTGEN, 2010), authenticity (RITTGEN, 2010), consistency (IEEE, 2010), understandability (MATOOK; INDUSKA, 2009; RITTGEN, 2010), unambiguousness (IEEE 2010), completeness (MATOOK; INDUSKA, 2009; RITTGEN, 2010), flexibility (MATOOK; INDUSKA, 2009), and usability (DAVIS, 1989; MATOOK; INDUSKA, 2009; RITTGEN, 2010). Table 47 presents the definition of these quality characteristics.

Table 47 - Definitions of the quality characteristics

Quality characteristics	Definition
Usability	Refers to the degree of understanding, ease of use and applying the method in an effective and efficient way (DAVIS, 1989; MATOOK; INDUSKA, 2009; RITTGEN, 2010).
Validity	Is defined as the appropriateness of the interpretations, inferences, and actions that we make based on the results of a study (JOHNSON; CHRISTENSEN, 2016).
Authenticity	Refers to the degree to which the method can realistically represent the domain it was defined (RITTGEN, 2010).
Correctness	Refers to how correct is the method, what is the extent of existing errors (RITTGEN, 2010).
Completeness	Refers to the degree of coverage of the method, if the method is sufficiently complete (MATOOK; INDUSKA, 2009; RITTGEN, 2010).
Consistency	Refers to the degree of uniformity, standardization, and freedom from contradiction among the components of the method (IEEE, 2010).
Understandability	Refers to the degree to which the purpose, concepts, and structure of the method is clear to the researchers (MATOOK; INDUSKA, 2009; RITTGEN, 2010).
Unambiguousness	Refers to the degree to which a definition/statement is described in terms that only allow a single interpretation (IEEE, 2010).
Flexibility	Refers to the degree to which the model and/or process can be adapted to changes, allowing it to be applied in contexts other than the one defined (MATOOK; INDUSKA, 2009).

Source: developed by the author.

In order to obtain an overall evaluation of the MEEGA+ method, and a detailed evaluation of the MEEGA+ model and the MEEGA+ process, analysis questions and metrics are defined for each quality characteristic (Table 48).

Table 48 - Analysis questions and metrics

Overall Evaluation of the MEEGA+ method	
Quality characteristic: Validity	
Analysis question	**AQ01:** Does the MEEGA+ method provide valid feedback on the game's quality?
Metrics	**M01.01.** Median of the experts' perception regarding the validity of the MEEGA+ method.
Quality characteristic: Usability	
Analysis question	**AQ02:** Does the MEEGA+ method have usability?
Metrics	**M02.01.** Median of the experts' perception regarding the effectiveness of the MEEGA+ method. **M02.02.** Median of the experts' perception regarding the efficiency of the MEEGA+ method. **M02.03.** Median of the experts' perception regarding the learnability of the MEEGA+ method.

	M02.04. Median of the experts' perception regarding the ease of use of the MEEGA+ method. **M02.05.** Median of the experts' perception regarding the utility of the MEEGA+ method.
Quality characteristic: Authenticity	
Analysis question	**AQ03:** Is the MEEGA+ method authentic?
Metrics	**M03.01.** Median of the experts' perception regarding the authenticity of the MEEGA+ method.
Detailed Evaluation of the MEEGA+ model	
Quality characteristic: Correctness	
Analysis question	**AQ04:** Is the MEEGA+ model correct?
Metrics	**M04.01.** Number of positive responses regarding the correctness of the MEEGA+ model. **M04.02.** Number of negative responses regarding the correctness of the MEEGA+ model. **M04.03.** Number of errors identified.
Quality characteristic: Completeness	
Analysis question	**AQ05:** Is the MEEGA+ model complete?
Metrics	**M05.01.** Number of positive responses regarding the completeness of the MEEGA+ model. **M05.02.** Number of negative responses regarding the completeness of the MEEGA+ model. **M05.03.** Number of incomplete items.
Quality characteristic: Consistency	
Analysis question	**AQ06:** Is the MEEGA+ model consistent?
Metrics	**M06.01.** Number of positive responses regarding the consistency of the MEEGA+ model. **M06.02.** Number of negative responses regarding the consistency of the MEEGA+ model. **M06.03.** Number of inconsistencies identified.
Quality characteristic: Understandability	
Analysis question	**AQ07:** Is the MEEGA+ model understandable?
Metrics	**M07.01.** Number of positive responses regarding the understandability of the MEEGA+ model. **M07.02.** Number of negative responses regarding the understandability of the MEEGA+ model. **M07.03.** Number of no understandable items.
Quality characteristic: Unambiguousness	
Analysis question	**AQ08:** Is the MEEGA+ model unambiguous?
Metrics	**M08.01.** Number of positive responses regarding the unambiguousness of the MEEGA+ model. **M08.02.** Number of negative responses regarding the unambiguousness of the MEEGA+ model. **M08.03.** Number of ambiguous items.
Quality characteristic: Flexibility	
Analysis question	**AQ09:** Is the MEEGA+ model flexible?

Metrics	**M09.01.** Median of the experts' perception regarding the flexibility of the MEEGA+ model.
Detailed Evaluation of the MEEGA+ process	
Quality characteristic: Correctness	
Analysis question	**AQ10:** Is the MEEGA+ process correct?
Metrics	**M10.01.** Number of positive responses regarding the correctness of the MEEGA+ process. **M10.02.** Number of negative responses regarding the correctness of the MEEGA+ process. **M10.03.** Number of errors identified.
Quality characteristic: Completeness	
Analysis question	**AQ11:** Is the MEEGA+ process complete?
Metrics	**M11.01.** Number of positive responses regarding the completeness of the MEEGA+ process. **M11.02.** Number of negative responses regarding the completeness of the MEEGA+ process. **M11.03.** Number of incomplete items.
Quality characteristic: Consistency	
Analysis question	**AQ12:** Is the MEEGA+ process consistent?
Metrics	**M12.01.** Number of positive responses regarding the consistency of the MEEGA+ process. **M12.02.** Number of negative responses regarding the consistency of the MEEGA+ process. **M12.03.** Number of inconsistencies identified.
Quality characteristic: Understandability	
Analysis question	**AQ13:** Is the MEEGA+ process understandable?
Metrics	**M13.01.** Number of positive responses regarding the understandability of the MEEGA+ process. **M13.02.** Number of negative responses regarding the understandability of the MEEGA+ process. **M13.03.** Number of no understandable items.
Quality characteristic: Unambiguous	
Analysis question	**AQ14:** Is the MEEGA+ process unambiguous?
Metrics	**M14.01.** Number of positive responses regarding the unambiguousness of the MEEGA+ process. **M14.02.** Number of negative responses regarding the unambiguousness of the MEEGA+ process. **M14.03.** Number of ambiguous items.
Quality characteristic: Flexibility	
Analysis question	**AQ15:** Is the MEEGA+ process flexible?
Metrics	**M15.01.** Median of the experts' perception regarding the flexibility of the MEEGA+ process.

Source: developed by the author.

In order to collect data on the defined metrics based on the experts' perspective, we conducted an expert panel (BEECHAM et al., 2005). An expert panel consists of bringing together experts from a specific knowledge area in order to get their opinion on some aspect of the research object (BEECHAM et al., 2005). As research strategy we conducted a survey, being the most suitable strategy for collecting information from people about a new method (WOHLIN et al., 2018). A questionnaire was defined as the data collection instrument, as it is best suited to the nature and type of data that we need to analyse (BEECHAM et al., 2005).

The questionnaire items are determined based on the defined metrics (Table 48). Table 49 presents the relationships between the metrics and the questionnaire items.

Table 49 - Relationship between metrics and questionnaire items

Metrics	Questionnaire items	Response format
Quality characteristic: Validity		
M01.01	Even though it is a self-assessment I consider the results on the quality of the game valid.	Likert scale (ranging from strongly disagree to strongly agree)
	I think that the classification of the game on a quality level using the MEEGA+ scale is valid.	
Quality characteristic: Usability		
M02.01	The MEEGA+ method allows to evaluate the quality of the educational game.	Likert scale (ranging from strongly disagree to strongly agree)
M02.02	Using MEEGA+ I can evaluate the quality of the game without significantly interrupting the flow of the class.	
	MEEGA+ allows me to evaluate an educational game with minimal effort.	
	MEEGA+ allows me to quickly collect data with respect to the quality of the game.	
	MEEGA+ allows me to quickly analyse the data.	
M02.03	Learning how to use the MEEGA+ method was easy.	
M02.04	I think that the MEEGA+ method is easy to use.	
M02.05	I think that the MEEGA+ method is useful to evaluate the quality of educational games.	
	The evaluation results provide a useful feedback to improve and/or helps me to select the game(s) that I use in my classes.	
Quality characteristic: Authenticity		
M03.01	The MEEGA+ method adequately includes what is necessary to evaluate games.	Likert scale (ranging from strongly disagree to strongly agree)
	The MEEGA+ method provides more support than other methods/models/frameworks currently available for the evaluation of educational games.	
Quality characteristic: Correctness		

M04.01 **M04.02** **M04.03**	Did you find any error in the MEEGA+ model (the decomposition of the factors, questionnaires, analysis spreadsheets and/or scale)?	Yes/No (Please indicate any error in the following way: Item (questionnaire, spreadsheet, etc): description)
M10.01 **M10.02** **M10.03**	Did you find any error in the MEEGA+ process (phases, activities and work products)?	Yes/No (Please indicate any error in the following way: Activity code: description)
Quality characteristic: Completeness		
M05.01 **M05.02** **M05.03**	Is there anything missing in the MEEGA+ model?	Yes/No (Please indicate what is missing in the model/process in the following way: Item: description)
M11.01 **M11.02** **M11.03**	Did you notice the absence of any phase, activity and/or work products that you consider important in this process?	
Quality characteristic: Consistency		
M06.01 **M06.02** **M06.03**	Did you find any inconsistency in the MEEGA+ model (decomposition of the factors, questionnaires, analysis spreadsheets and/or scale)?	Yes/No (Please indicate any inconsistency in the following way: Item (questionnaire, spreadsheet, etc): description)
M12.01 **M12.02** **M12.03**	Did you find any inconsistency in the MEEGA+ process (phases, activities and work products)?	Yes/No (Please indicate any inconsistency in the following way: Activity code: description)
Quality characteristic: Understandability		
M07.01 **M07.02** **M07.03**	Did you find something incomprehensible in the MEEGA+ model (factor, metric, questionnaire item, scale quality level, etc.)?	Yes/No (Please indicate any incomprehensible item in the following way: Item (questionnaire, spreadsheet, etc): description)
M13.01 **M13.02** **M13.03**	Did you find something incomprehensible in the MEEGA+ process (phases, activities and work products)?	Yes/No (Please indicate something incomprehensible in the following way: Activity code: description)

Quality characteristic: Unambiguousness		
M08.01 **M08.02** **M08.03**	Did you find something ambiguous in the MEEGA+ model (factor, metric, questionnaire item, scale quality level, etc.)?	Yes/No (Please indicate any ambiguous item in the following way: Item (questionnaire, spreadsheet, etc): description)
M14.01 **M14.02** **M14.03**	Did you find something ambiguous in the MEEGA+ process (phases, activities and work products)?	Yes/No (Please indicate something ambiguous in the following way: Activity code: description)
Quality characteristic: Flexibility		
M09.01	The MEEGA+ model may be easily adapted to evaluate a game with different learning objectives.	Likert scale (ranging from strongly disagree to strongly agree)
M15.01	The MEEGA+ process may be easily adapted to evaluate a game with different learning objectives.	

Source: developed by the author.

Additional questions were also included in the questionnaire in order to collect demographic information about the experts, collect information about the previous use of the MEEGA+ method or the initial version of the MEEGA model, and to collect the perceptions of the main strengths and weaknesses of the MEEGA+ method.

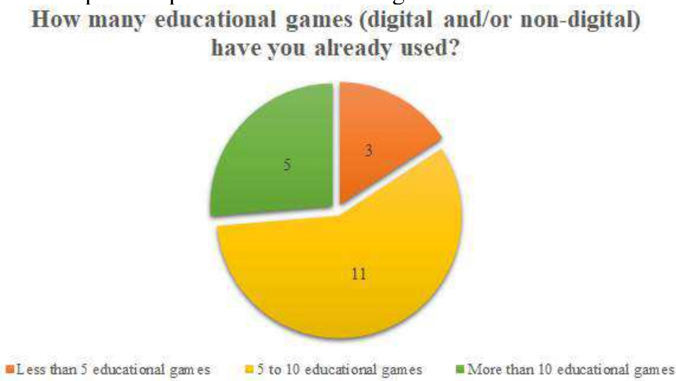
6.2.2 Execution

The expert panel was conducted in August and September 2018. We targeted experts in educational games from different backgrounds such as computing, education, statistics, etc. as recommended by Kitchenham et al. (2002). We define an expert, in the context of this study, as a person who has a PhD or is a PhD student in any knowledge area with scientific publications in recognised Brazilian and/or international journals and/or conferences in the field of educational games. We invited 34 experts to participate in the evaluation of the MEEGA+ method. All experts were personally invited by sending a private e-mail containing the technical report describing the MEEGA+ method (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2018c) and the questionnaire online as a Google Forms. In total, 19 experts accepted and answered the questionnaire completely (representing a response rate of 56%).

Analysing the demographic information, 14 experts are PhD and 5 are PhD students. Most of the participants have an academic background in Computer Science (13 experts), 5 participants have a background in Informatics in Education and 1 participant has a background in Linguistics.

The participants were asked about their practical experience in the use, development/customization, and evaluation of games. Regarding the use of games (Figure 61), most of the experts have used between 5 and 10 educational games, 3 experts have used less than 5 games, and 5 experts have used more than 10 educational games. This demonstrate that the majority of the experts has a practical experience in the use of educational games, including digital and non-digital ones.

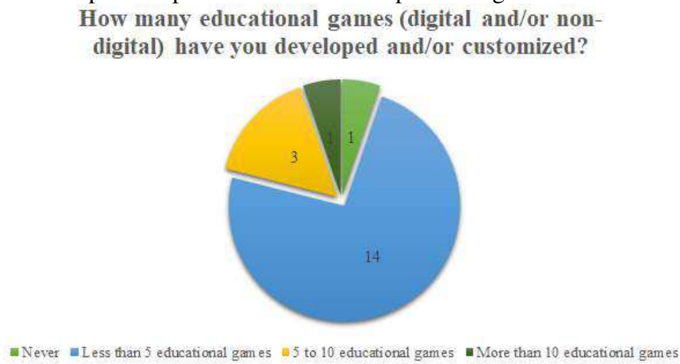
Figure 61 - Experts' experience in the use of games



Source: developed by the author.

Regarding the development and/or customization of educational games (Figure 62), most of the participants developed and/or customized less than 5 games, 4 experts developed/customized more than 5 educational games, and only one participant never developed and/or customized an educational game. Thus, based on these data, we can observe that a few numbers of games have been developed and/or customized by the experts. This result may be related to the degree of difficulty and/or effort necessary for the development of educational games, typically developed by researchers themselves, under restrictions of time and cost (BATTISTELLA; GRESSE VONWANGENHEIM, 2016; PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017).

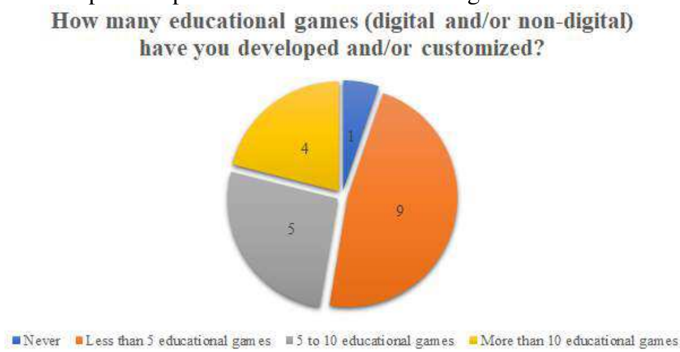
Figure 62 - Experts' experience in the development of games



Source: developed by the author.

Most of the participants evaluated less than 5 educational games (Figure 63), 5 participants evaluated between 5 and 10 games, 4 experts evaluated more than 10 games, and one participant never evaluated an educational game.

Figure 63 - Experts' experience in the evaluation of games



Source: developed by the author.

Based on these data, we can observe that, consistent with the few number of experts that have developed and/or customized games, is the few number of experts that have conducted an evaluation. This result may also be related to the effort necessary to conduct an evaluation (ALL; CASTELLAR; LOOY, 2016). Thus, also justifying the need of an evaluation method that provide a comprehensive support in order to guide researchers and minimize the effort in the conduct of such evaluations.

Based on the responses of the experts (n=19), we also analysed the reliability of the standardized items (items using a Likert scale) of the questionnaire used in the expert panel. The results of the analysis indicate an excellent reliability (Cronbach's alpha α=.907), indicating that the items of the questionnaire used in the expert panel are consistent and precise with respect to the quality evaluation of the MEEGA+ method.

6.2.3 Analysis and Discussion

In order to characterize the use of the MEEGA+ model, we asked the participants if they had already used the MEEGA+ model or the initial version of the MEEGA model to evaluate games. In this regard, most of the participants (13) used the MEEGA/MEEGA+ model one or more times to evaluate games, and 6 participants never used the model. Most of the experts have a previous practical knowledge in the MEEGA/MEEGA+ evaluation model. In addition, 12 participants also collected data using the MEEGA/MEEGA+ model and analysed the data collected using the spreadsheet provided by the MEEGA/MEEGA+ model. Thus, this previous knowledge in the MEEGA/MEEGA+ model may be useful in the evaluation of the MEEGA+ method. Results of the evaluation of the MEEGA+ method are presented for each defined analysis question.

Overall Evaluation of the MEEGA+ method

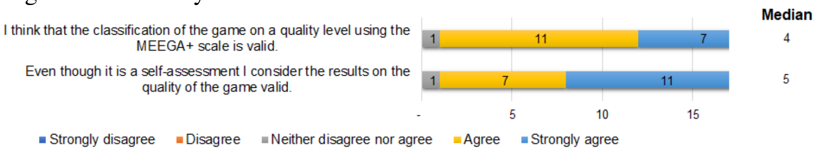
The response format of the items in the overall evaluation of the MEEGA+ method is a Likert scale ranging from strongly disagree (1) to strongly agree (5). Thus, the results of these items are interpreted based on the median and the response frequency for each item, graphically presented trough frequency diagrams.

AQ01: Does the MEEGA+ method provide valid feedback on the game's quality?

Analysing the validity of the MEEGA+ method based on the experts' perspective (Figure 64), the majority of the participants indicates that even the MEEGA+ method using a self-assessment they consider that the evaluation results are valid. This result seems to be consistent with the results of the reliability and validity analysis of the MEEGA+ measurement instrument (section 6.1), indicating that the MEEGA+ method provides a valid feedback on the game's quality through a reliable and valid measurement instrument.

The experts also indicated that the classification of the game on a quality level using the MEEGA+ scale is valid. Thus, the scale provided by the MEEGA+ method may be used as an effective instrument to classify the games' quality and, thus, contributing to the selection of games to use as an instructional strategy.

Figure 64 - Validity results of the MEEGA+ method



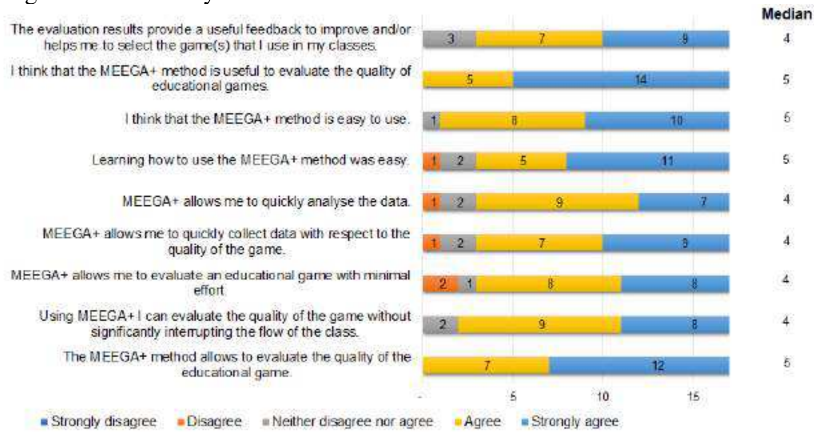
Source: developed by the author.

AQ02: Does the MEEGA+ method have usability?

Regarding the usability of the MEEGA+ method, in general, the majority of the experts indicated that the MEEGA+ method has a good usability, mainly in terms of usefulness, learnability and effectiveness.

Based on the results presented in Figure 65, we can identify that the experts consider that the MEEGA+ method besides to provide a useful feedback to improve and/or help researchers in the selection of games for their classes, it is considered easy to use, easy to learn to use, minimizes the interrupting of the flow of the class when conducting an evaluation, and provides efficient instruments to quickly collect and analyse the data collected.

Figure 65 - Usability results of the MEEGA+ method



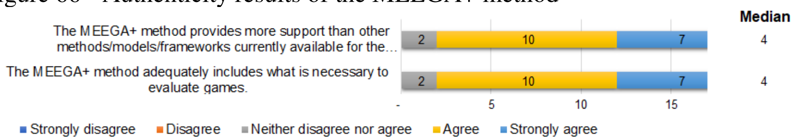
Source: developed by the author.

On the other hand, two experts disagree that the MEEGA+ method allows evaluate games with minimal effort. This result may be related to the several activities that need to be performed to conduct an evaluation, although the MEEGA+ method provides a comprehensive support, providing instruments to collect and analyse data and describing a step by step to guide researchers in the evaluation.

AQ03: Is the MEEGA+ method authentic?

Analysing the authenticity of the MEEGA+ method (Figure 66), we can observe that the majority of the experts (17) agree or strongly agree that the MEEGA+ method provides more support than other approaches currently available for evaluation of educational games, and adequately includes what is necessary to evaluate games. Thus, these results seem to be confirming the originality of the MEEGA+ method, based on the experts' perspective, providing a systematic and comprehensive support for game evaluations.

Figure 66 - Authenticity results of the MEEGA+ method



Source: developed by the author.

Detailed Evaluation of the MEEGA+ model

AQ04: Is the MEEGA+ model correct?

Regarding the correctness of the MEEGA+ model, 95% of the participants (18 experts) did not find any errors in the decomposition of the factors, questionnaire, analysis spreadsheet and scale. Thus, indicating that the MEEGA+ model is correct.

However, one participant indicates that the learning should not be evaluated only based on students' opinion. In this respect, although it is an important issue, in the current version of the MEEGA+ model we only evaluate the perceived learning due to the restrictions in performing game evaluations in class. As defined and justified in chapter 5, the MEEGA+ model has been developed to be used in non-experimental studies using case studies, in order to not interrupt the normal flow of a class and to not impair the participants involved in the study. However, the evaluation of the learning impact may be considered in a future version of the MEEGA+ model.

AQ05: Is the MEEGA+ model complete?

In terms of the completeness of the MEEGA+ model, 79% of the experts (15) did not find anything missing in the MEEGA+ model, thus, indicating that the MEEGA+ model is complete. One participant, although has indicated that there is not something missing in the MEEGA+ model, indicates that "there are new factors to consider, but it would mean a longer response time". In addition, 4 experts also indicate additional comments to this question, as presented in Table 50.

Table 50 - Experts' comments related to the completeness of the MEEGA+ model

Id	Experts' comments
C1	Objective measurements are missing. But I believe maybe it is not the goal of the instrument. That is why I have proposed some objective measures to complement your model. Krassmann, A. L., Falcade, A., Nunes, L. N., & Medina, R. D. Análise do modelo "Visão do Estudante" de avaliação de jogos sérios digitais. Informática na educação: teoria & prática, 20(3), 2017. Krassmann, A. L., Falcade, A., & Medina, R. D. Visão do Estudante: proposta de um modelo de avaliação de jogos sérios digitais. XXV Ciclo de Palestras sobre Novas Tecnologias na Educação, 2017.
C2	I think it would be important to consider (although I know that this has been removed due to its complexity) motivation (or even engagement). Perhaps consider the effort (which for some authors can be considered synonymous with motivation). Add questions "outside the curve" (security) to identify inconsistent responses (i.e.: This is a question to see if you are paying attention and reading questions. If you are paying attention, check option 2).
C3	I believe that the evaluation of student's learning could be clearer. Sometimes, when elaborating the questions related to bloom taxonomy I have several doubts.
C4	To better describe the population of the evaluation in terms of education: educational level, previous knowledge... Include some open questions.

Source: developed by the author.

Regarding comment C1, the expert indicated that objective measurements are missing and cites two research articles. The papers cited were not considered in our analysis of the state-of-the-art maybe due to the fact that they are written in Brazilian Portuguese and are not indexed in the analysed data sources. Analysing the mentioned articles, the authors consider as "objective measurements" the data collected from the interaction of the students with the game such as number of phases played, number of phases completed, total time spent, etc. As commented by the expert, in fact, the players' interactions with the game are out of the scope of this research. However, we intend to analyse the importance of these aspects and consider in a future version of the MEEGA+ method.

With respect to comment C2, as explained in section 5.1, as a result of our analysis of the state-of-the-art and the practice we identified a series of quality factors to evaluate games (including motivation and engagement), with similar conceptual definitions. Thus, we mapped these factors analysing their conceptual definitions and similarities in order to define the quality factors that compose the MEEGA+ model. With respect to the expert's suggestion to include questions "outside the curve", although it may be an interesting idea, we prefer not to include it in this version of the MEEGA+ model.

Regarding comment C3, as explained in section 5.1, the evaluation of the perceived learning is conducted based on standardized questionnaire items and based on items customized for each learning objective of the game, adopting the Bloom's taxonomy. In this respect, in order to assist the researchers in the definition of such items, we provided an example on how to define such statements based on the learning objectives of the game SCRUMIA.

With respect to comment C4, the expert suggests including educational aspects of the population involved in the evaluation such as learning level and previous knowledge. In this respect, as our context is higher computing education, we understand that all students that compose the evaluations are in the same learning level (higher education). However, the previous knowledge of the students is not considered in the current version of the MEEGA+ model. This aspect may be analysed and considered in a future version of the MEEGA+ model. Considering such differences in education stages we are also working on the customization of the MEEGA+KIDS model (GRESSE VON WANGENHEIM; PETRI; BORGATTO, 2018), an adaptation of MEEGA+ to be used in the context of K-12 education.

In summary, 2 of the 4 experts' suggestions/comments were already defined and explained in the MEEGA+ model but may not have been understood by the experts when evaluating the MEEGA+ method. Two other comments are not covered by the current version of the MEEGA+ model and may be analysed and considered in a future version.

AQ06: Is the MEEGA+ model consistent?

In general, the vast majority of the experts (18) indicates that the MEEGA+ model is consistent, representing 95% of the participants. However, one participant indicated an inconsistency related to the quality factors in the spreadsheet and the questionnaire (Table 51).

Table 51 - Expert's comment related to the consistency of the MEEGA+ model

Id	Experts' comments
C1	In the Figure 3, we can see that Player Experience and Usability are different quality factors. However, in the spreadsheet (Figure 4), the Usability is presented as part of the Player Experience. Besides that, I did not understand why Perceived Learning is not considered as a quality factor, once we also have statements to evaluate it in the MEEGA+ questionnaire.

Source: developed by the author.

In fact, the inconsistency reported in comment C1 was identified in the technical report provided to the experts. We corrected this inconsistency in the current version of the MEEGA+ model, as presented in Figure 26.

AQ07: Is the MEEGA+ model understandable?

Analysing the understandability of the MEEGA+ model, 89.5% of the experts (17) did not find anything incomprehensible in the decomposition of the factors, questionnaire, analysis spreadsheet and scale. Thus, indicating that the MEEGA+ model is understandable. However, two experts indicate a need for a more understandable explanation about the game quality scale and the definition of the quality factors (Table 52).

Table 52 - Expert's comment related to the understandability of the MEEGA+ model

Id	Experts' comments
C1	I think it is necessary a more comprehensible explanation about the Game Quality Scale. I understood how to calculate the metric, but I did not understand why this metric is suitable and how this metric measures the game quality. This point can be questionable regarding the construct validity of the case studies adopting MEEGA+.
C2	I missed a more detailed explanation/justification for the factors included in the model. Specifically, the second paragraph of page 10 details your sources but not the reasons for the selection of the quality factors and dimensions. Furthermore, the introduction of the items regarding the learning goals in page 13 appears a little disconnected from the quality model and can it be confusing for the reader how these factors relate to the quality model and the quality assessment.

Source: developed by the author.

With respect to comment C1, the technical report provided to the experts to evaluate the MEEGA+ method, presented only a description of the final version of the game quality scale, not presenting a detailed description of the development of the scale and how the scale determines the games' quality level based on anchor items. However, in section 5.1.1

of this work, we provide a detailed description of the game quality scale in order to clarify its understanding.

Similarly, with respect to comment C2, the factors included in the quality model are defined based on the results of the literature reviews and a mapping analysing the conceptual definition and similarities among the factors, as presented in section 5.1. This mapping was omitted of the technical report provided to the experts due to the fact that the focus of the report is to present the final version of the MEEGA+ method, not presenting a detailed description of the design of the method.

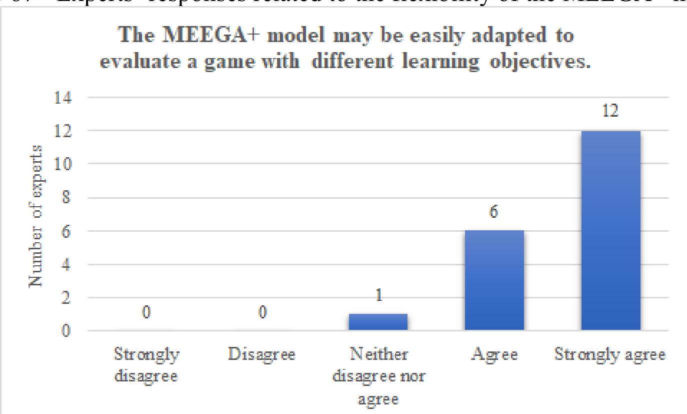
AQ08: Is the MEEGA+ model unambiguous?

95% of the participants (18 experts) did not find any ambiguity in the decomposition of the factors, questionnaire, analysis spreadsheet and scale. Thus, indicating that the MEEGA+ model is unambiguous. One expert indicated that s/he not evaluated this factor.

AQ09: Is the MEEGA+ model flexible?

We can identify that most of the experts agree or strongly agree (18 experts) with the statement that the MEEGA+ model is flexible (Figure 67). This result is consistent with the design of the model, as explained in the section 5.1 the MEEGA+ model, in fact, may be customized to evaluate games with different learning objectives (develop knowledge, skill and/or attitude) in different learning levels.

Figure 67 - Experts' responses related to the flexibility of the MEEGA+ model



Source: developed by the author.

In summary, the MEEGA+ model was positively evaluated based on the experts' opinion. A few numbers of inconsistencies and misunderstood items have been reported, being most of them corrected and/or explained in the current version of the MEEGA+ model. Thus, based on these results, we obtain a first indication that the MEEGA+ model is correct, complete, understandable, unambiguous, consistent, and flexible.

Detailed Evaluation of the MEEGA+ process

AQ10: Is the MEEGA+ process correct?

Analysing the MEEGA+ process in terms of correctness, all participants (19 experts) did not find any errors in the phases, activities and work products of the process. Thus, indicating that the MEEGA+ process is correct. As a suggestion, one participant indicated that the process could be synthesized in tables indicating the work products. In this respect, we provided a table for each phase of the MEEGA+ process (section 5.2), and the templates of the work products are available in Appendix E. In addition, another expert indicated some typos in the technical report, the suggestions were also corrected in this work.

AQ11: Is the MEEGA+ process complete?

Regarding the completeness of the MEEGA+ process, 95% of the participants (18 experts) did not find anything missing in the MEEGA+ process, thus, indicating that the MEEGA+ process is complete. However, two participants indicated a need to include a description of the roles typically involved in carrying out the process. The suggestions are presented in Table 53.

Table 53 - Experts' comments related to the completeness of the MEEGA+ process

Id	Experts' comments
C1	In the process description, it could have been included a description of the roles typically involved in carrying out the process and, consequently, an indication of the role responsible for carrying out each activity, in the case of more than one person carries out the process (researchers, teachers, students, fellows, junior researchers, etc.)
C2	It is interesting to identify the roles/responsible for the execution of each activity of the process. // Between activities 3.2 and 3.3 would there be an activity to prepare the participants? Whether in terms of training or control groups (in any specific situation)?

Source: developed by the author.

With respect to comments C1 and C2 related to the definition of roles for the execution of each activity, as explained in the section 5.2, in the MEEGA+ process we assumed that, typically, game evaluations are conducted by researchers. However, the MEEGA+ process is described in order to be applied by any student or instructor, not requiring advanced knowledge in education, measurement or statistics. Thus, in the MEEGA+ process, we define the generic role researcher representing any student/instructor/etc. responsible for the conduct of the activities of the process.

In addition, comment C2 also indicates that an activity to prepare students for the game could be included between activities 3.2 (Obtain participants' consent) and 3.3 (Execution of the game). However, the preparation of the students before the execution of the game is covered in the activity 3.3 (Execution of the game), as described in section 5.2.

AQ12: Is the MEEGA+ process consistent?

Analysing the MEEGA+ process in terms of consistency, 95% of the participants (18 experts) did not find any inconsistency in the in the phases, activities and work products of the MEEGA+ process. One participant indicated that the use of tables and the figure summarizing the process helped her/him in understanding the process.

However, one expert asked a question related to the premise of the activities (Table 54).

Table 54 - Expert’s comment related to the consistency of the MEEGA+ process

Id	Experts’ comments
C1	Why some activities have "premise" and others do not? There should be a pattern in the process description.

Source: developed by the author.

In this respect, we define a premise only in the first activity of the process (Activity 1.1. Select the object of study (educational game)). The premise indicates that the game selected for the evaluation must have already been developed, indicating that the process does not support the development of a new game. Based on the comment of the expert, we reclassify this premise as a premise of the whole process (and not only as a premise of activity 1.1). This correction has been performed in the current version of the MEEGA+ process.

AQ13: Is the MEEGA+ process understandable?

Regarding the understandability of the MEEGA+ process, 95% of the experts (18) did not find anything incomprehensible in the phases, activities and work products of the process. Thus, indicating that the MEEGA+ process is also understandable. On the other hand, only one participant asked a question related to the students’ consent in the evaluations (Table 55).

Table 55 - Expert’s comment related to the understandability of the MEEGA+ process

Id	Experts’ comments
C1	Why is the consent of the participants optional?

Source: developed by the author.

With respect to comment C1, we define activity 3.2 (Obtain participants' consent) as optional, due to the fact that evaluations conducted in educational institutions that do not have an Ethics Committee or that do not require the approval of the committee for this kind of research, obtaining consent from participants may be optional.

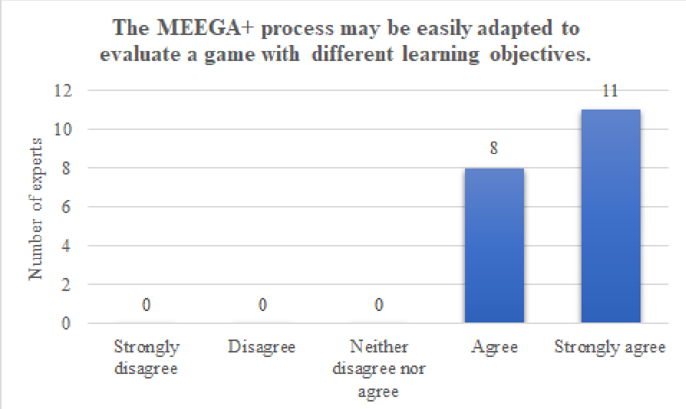
AQ14: Is the MEEGA+ process unambiguous?

Analysing the MEEGA+ process in terms of unambiguousness, all participants (19 experts) did not find any ambiguity in the phases, activities and work products of the process. Thus, indicating that the MEEGA+ process may be considered unambiguous.

AQ15: Is the MEEGA+ process flexible?

Regarding the flexibility of the MEEGA+ process to evaluate games with different learning objectives (Figure 68), we can identify that all the experts agree (8) or strongly agree (11) with the statement that the MEEGA+ process may be easily adapted to evaluate a game with different learning objectives.

Figure 68 - Experts' responses related to the flexibility of the MEEGA+ process



Source: developed by the author.

In general, the results of the detailed evaluation of the MEEGA+ process are similar to the results of the analysis of the MEEGA+ model, also indicating that the experts consider that the MEEGA+ process is correct, complete, understandable, unambiguous, consistent, and flexible.

Addition questions were also provided to the experts indicate the main strengths, weakness and additional comments identified during the evaluation of the MEEGA+ method. Table 56 presents the strengths, weakness and the additional comments reported by the experts.

Table 56 - Strengths, weakness and additional comments reported by the experts

Strengths of the MEEGA+ method reported by the experts
It is a powerful tool to evaluate digital games. It is a reference to our field, which I always recommend to my colleagues. Well theoretically founded, well defined, very complete, very useful and easy to use.
- MEEGA+ method is a reliable instrument to evaluate educational games. -The method enables us to aggregate results from different case studies about different educational games. - The method supports the planning, execution, analysis and reporting of educational game evaluations.
It defines a clear process to evaluate educational games, having establishing quality factors and its dimensions relevant and easy to understand.
Easy to use; Reduction of subjectivity. It has spreadsheet to assist in data collection. Theoretical background behind the questionnaire.
Allows to evaluate the quality of educational games. In addition, allows quickly both collect and analyse data with respect of the game. Provides a useful feedback to helps to select a game.
Technical and process quality of the model. Scientific and statistical rigor and accuracy. Full support to all involved activities.

Evaluate the quality of educational games in terms of usability and player experience from the students' perspective.
Ease of understanding, adaptation to any educational game.
Systematization
I believe that it is the ease of use and the delivery of the results ready for the discussion about the quality of the game. This makes it very practical to use. In addition, MEEGA + allows you to add new items to be queried to users.
It is a very broad model that comprises several aspects of educational games. In addition, it already has several evidences/experimental evidences of applicability.
Taking into consideration players experience - not only teachers and researchers' expectations. Also, allowing customization for different areas.
The main advantage of MEEGA+ is to be easy to use by any instructor. This helps a lot. In addition, its application process is simple and not complicated to be performed, it does not take so long to try to fit it into my teaching context.
Have spreadsheets and graphics attached as well as process that presents what to do.
It defines a clear process to evaluate educational games, having establishing quality factors and its dimensions relevant and easy to understand.
The support material (spreadsheet, activities, etc). After learning how to use it, having automated elements in the spreadsheet helps a lot. Support for quantitative analysis. The results of the evaluation convince me (at least to me). The site with material and the web application attracts users. The fact of consider digital and non-digital games.
The MEEGA+ method allows to perform in an easy way the evaluation of educational resources and it gives support to analyse the effectiveness of educational games.
A reliable, flexible and easy-to-use method to evaluate educational games. The process that supports the method application is complete and easy to understand and follow. The supporting tools for the analysis phase are very useful.
Weaknesses of the MEEGA+ method reported by the experts
As a thorough and rigorous method, it could be complex for inexperienced users. However, the technical document describing the method, as it is complete and very enlightening, can minimize this possible difficulty.
The lack of objective measurements, as I previously mentioned. As a future work, it could be developed a web system to insert the data and generate the results, instead of the spreadsheets. This system could include objective measures.
Time to complete
Regarding the main weaknesses, I believe there are other quality factors that can be considered, and the current factors may not be important in the future. However, I understand that the factors need to be constrained so that it does not become a very large assessment.
The main weakness I find is about customization of questions related to perceived learning. I still have many doubts about which ones to use, even though I have an example.
The description of the process should be simpler.
A potential problem are the self-report measures. Students may exaggerate their opinions, to make the game look worse or better than it actually is.
It would not be a weakness, but perhaps for teachers in areas outside the scope of technology they may have greater difficulties, especially as regards the part of data analysis and manipulation of tables.
Research Method: Figure 1 a little confusing and with colours that do not stand out / help in reading. Do the colours and rectangles have any specific meaning?

I felt lack of justification for some decisions. For example: why BPMN? Why defined values 15, 50, 0, 5, etc. Review English. I believe that on some points it can be improved. Ex: analyse whether it is "data collected" or "data collected". The reading itself, although detailed, tires the reader a little (maybe it's just the format / model of the text).
As weaknesses, perhaps if I can highlight the subjectivity of the evaluation since it depends on the perception of the players.
I can only say that I found a little gap between the quality model and the assessment of how well a game helps to achieve the educational goals.
Additional comments reported by the experts
Only congratulations and gratitude for this important work. I'm proud that it is Brazilian!
I liked the proposal and will be evaluating games that I have used in class.
MEEGA + is an excellent support tool both teachers and educational game developers. It has a relatively easy learning curve and clear instructions for application.
It was really innovative for me. Congrats for all the team!
I suggest creating a presentation video of the Model, explaining how it works. And also, a video tutorial on how to apply in a simple case of a game.
I can use the next semester in a Human-Computer Interface course, but I would need the spreadsheets.
Some threats to validity related to case studies with MEEGA+ could be explored by the authors in the technical report. As some students and instructors with no empirical software engineering knowledge can adopt MEEGA+, it could be useful to provide some guidelines in order to ensure the validity of the results.
Section 3.1 - I was curious because only two factors of quality (player experience and usability)? Was it the result of other work? Or was it a design decision? I think it's very useful for evaluating educational games. I have an interest in using it in my disciplines (where I already commented but did not apply). I think at first glance, using the method may scare you a little. I had already used previous versions (ref [1]) and had some difficulty. Specifically, in the report I missed some reading. Maybe by style / text format (very block, no indentation). Thus, the text was "heavy" to read. Maybe because the paragraphs are all together (without a space separating the lines). An index would also be very useful!
In the printed version of the document it is difficult/impossible to read some figures, especially the ones that are screenshots of the Excel file.
Maybe, it is desirable to try finding a way to perform the evaluations in a more objective way. But, I consider the MEEGA+ method is a well-defined method to evaluate educational games and get an impression of their effectiveness in education.

Source: developed by the author.

As shown in Table 56 several strengths, weaknesses and additional comments were reported by the experts. The reported weaknesses are typically related to the lack of objective metrics, evaluation of the learning impact, the lack of explanation/justification of some decisions, and the complexity of the method. Some of these issues were corrected in the current version of the method. However, other issues such as the inclusion of objective metrics and evaluation of the learning impact are out of the scope of this research and may be considered in a future version of the MEEGA+ method. On the other hand, the main strengths reported by the experts are related to the comprehensive and systematic support provided by the method, allowing a quick data collection and analysis through the questionnaire and the spreadsheet that compose the method.

Threats to validity. A main limitation in the conduction of an expert panel is related to the small sample size involved (BEECHAM et al., 2005). This is due to the fact that this type of research strategy is not intended to cover a relevant sample of the target audience, but rather to conduct an in-depth analysis of the object of study based on the perspective of a small group of experts appropriately qualified for such activity (KITCHENHAM et al., 2002). In this respect, although the sample involved in this study (19 experts) may be considered a small sample, it is composed of PhD and PhD students with relevant publications in recognized journals and/or conferences in the area of educational games.

Another limitation of the expert panel is the low response rate. However, from the 34 experts invited to participate of this research, 19 experts accepted to participate. Thus, representing a response rate of 56%. In this respect, a response rate of 55.6%, with a variation of +/- 19.7% is considered normal in this type of research (BARUCH, 1999; BEECHAM et al., 2005).

Another limitation refers to the reliability of the data collection instrument used in the expert panel (BEECHAM et al., 2005). In this respect, we also analysed the reliability of the questionnaire used in the expert panel. We obtained a satisfactory internal consistency (Cronbach's alpha α=.907), thus indicating the reliability of the questionnaire used in the expert panel.

6.3 CHAPTER CONCLUSIONS

This chapter presents the evaluation of the MEEGA+ method. The analysis of the reliability and validity of the measurement instrument of the MEEGA+ model is based on data collected from 62 case studies conducted involving a population of 1048 students. The results indicate that the MEEGA+ measurement instrument has a satisfactory reliability and construct validity. With respect to reliability, a Cronbach's alpha α=.927 indicates an excellent internal consistency, which means that the responses between the items are consistent and precise. The results of an exploratory factor analysis confirm that the quality of games for computing education is evaluated through two quality factors (player experience and usability). In addition, the results show evidence of convergent validity through a satisfactory degree of correlation found between the majority of the items for all quality factors. Furthermore, items of different quality factors present a small correlation, indicating evidence of discriminant validity. Thus, based on our results, we can conclude that the MEEGA+ measurement instrument has a high reliability and measures what it purports to measure: the quality of games used for computing education in terms of usability and player experience.

In order to evaluate the quality of the MEEGA+ method we conduct an expert panel. The results based on the opinion of 19 experts in the area of educational games, provide a first indication that the MEEGA+ method is valid, authentic, correct, complete, understandable, unambiguous, consistent, flexible, and has a good usability. However, some inconsistencies, misunderstood items and suggestions have been reported by experts. Some of the reported issues were corrected in the current version of the MEEGA+ method, and others, may be considered in a future version of the MEEGA+ method. Therefore, based on the results of the expert panel and considering the strengths reported by the experts, we can observe several indications that the MEEGA+ method achieves its objective, providing a systematic and comprehensive support for quality evaluation of games used as an instructional strategy for computing education.

7 RESULTS

In this chapter the results produced by this research are summarized and discussed. Section 7.1 presents the results produced in order to answer the research question and the objectives. In the section 7.2, the MEEGA+ method is compared with the related studies identified in the elicitation of the state-of-the-art. Section 7.3 describes the scientific publications produced by the author in the context of this research.

7.1 ANSWERING THE RESEARCH QUESTION

The research question addressed in this research is: "How to systematically conduct a quality evaluation of educational games used as an instructional strategy for computing/SE education?". Thus, in order to answer this question, we define as the main objective of this research: "to develop and evaluate a method for quality evaluation of games used as an instructional strategy for computing/SE education". In order to achieve this main objective, it was decomposed in six specific objectives.

To achieve the specific objective O1, we elicited the state of the art identifying existing approaches (methods, models, frameworks, and scales) for the systematic evaluation of educational games (PETRI; GRESSE VON WANGENHEIM, 2016). As a result, we identified that most of the existing approaches are frameworks rather than systematic evaluation methods/models, indicating a lack of support on how to conduct and operationalize such evaluations. In addition, in order to achieve the specific objective O2, we also analysed the state of practice, eliciting how games used for computing/SE education are currently evaluated (PETRI; GRESSE VON WANGENHEIM, 2016). As a result, we identified the MEEGA model (SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011) as a well-defined approach for the evaluation of games for computing/SE education that is being widely used in practice.

To achieve the specific objective O3, we conducted a large-scale study of the initial version of the MEEGA model analysing its validity and reliability. The analysis was conducted based on data collected in 60 case studies, involving a population of 1000 students (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2016; PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017). As a result, we identified that the initial version of the MEEGA model is acceptable in terms of reliability (α=.915). However, in terms of its validity, a conceptual overlap with respect to the factors motivation and user experience has been observed, indicating a need for the redesign of the MEEGA model.

Thus, based on the results of the literature reviews (PETRI; GRESSE VON WANGENHEIM, 2016; PETRI; GRESSE VON WANGENHEIM, 2017) and the large-scale analysis of the initial version of the MEEGA model (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2017), we designed a new evaluation method, the MEEGA+ method. The MEEGA+ method aims to provide a systematic support for the evaluation 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, guiding researchers on how to plan, execute and analyse the results of game evaluations. Thus, achieving our specific objectives O4 and O5, respectively.

In order to evaluate the MEEGA+ method and achieve our specific objective O6, we conduct a series of 62 case studies applying the MEEGA+ method evaluating 24 different educational games, involving 1048 students of eight educational institutions. Based on data collected in the case studies, we conduct a reliability and validity analysis of the MEEGA+ measurement instrument (PETRI; GRESSE VON WANGENHEIM; BORGATTO, 2018). As a result, we identified that the MEEGA+ measurement instrument has an excellent reliability through a Cronbach's alpha coefficient α=.927. In addition, results of validity analysis confirm the original structure of the MEEGA+ model, indicating that the quality of games for computing education is evaluated through two quality factors: usability and player experience. An additional evaluation of the MEEGA+ method, based on the opinion of 19 experts in educational games, was also conducted. As a result, we obtained a first indication that the MEEGA+ method is valid, authentic, correct, complete, understandable, unambiguous, consistent, flexible, and has a good usability

Therefore, based on these results, we can answer our research question, indicating that one way to systematically evaluate games for computing/SE education is adopting the MEEGA+ method. The results of its evaluation indicate that the MEEGA+ method is a reliable and valid method to evaluate games as well as was indicated by experts that it provides a systematic and comprehensive support for quality evaluation of games used as instructional strategy for computing/SE education.

7.2 COMPARING THE MEEGA+ METHOD WITH THE STATE OF THE ART

Comparing the MEEGA+ method with related approaches identified in the analysis of the state-of-the-art (Table 57), we can observe that the main originality of the MEEGA+ method is the comprehensive support for game evaluations through an evaluation model, defining quality factors to be evaluated through a standardized measurement instrument, a scale, classifying the quality level of the game, and a process, guiding researchers on how to plan, execute and analyse the results of game evaluations.

As shown in Table 57, most of the related approaches are frameworks, which are flexible and adaptable to specific contexts. However, the frameworks itself do not provide a comprehensive support on how to conduct the evaluation, data collection and analysis. In this regard, MEEGA+ defines a method, including an evaluation model, game quality scale, and evaluation process, providing a systematic and comprehensive support for conduction of game evaluations.

Most of the existing approaches also seem to be developed in a rather ad-hoc manner, not following a well-defined development guide/approach. In this respect, the MEEGA+ method has been developed adopting the GQM approach, the guide to development of scales and the questionnaire design, Item Response Theory as well as empirical study process and prescriptive modelling. In terms of data collection instruments used by the related approaches, the questionnaire is the most used one. However, typically, such questionnaires are developed in an ad-hoc way, not systematically decomposing measurement items from theoretical constructs, and no information of its validity and reliability is provided. In the MEEGA+ method a questionnaire is also used for data collection. However, it was systematically developed adopting the GQM approach, decomposing quality factors into measurement instrument items, and it was widely evaluated in terms of validity and reliability based on data collected from a series of case studies.

Regarding the evaluation of the related approaches, only two approaches (MEEGA and EGameFlow) provide information about the reliability and validity of their measurement instruments. In this regard, the MEEGA+ method besides to conduct a reliability and validity evaluation of its measurement instrument, also conducted a comprehensive evaluation, based on experts' perspective, of the complete method, indicating the correctness, authenticity, consistency, and unambiguousness of the MEEGA+ method.

Table 57 - Comparing the MEEGA+ method with related studies

Definition			Operationalization			Evaluation			
Name/ Reference	**Approach**	**Development Methodology**	**Study type**	**Data collection instrument**	**Data analysis methods**	**Evaluated factors**	**Applications/ Data points**	**Data analysis methods**	**Evaluation results**
MEEGA+ method	Method (model + scale + process)	GQM, Scale development guide, Guide for questionnaire design, Item response theory, Empirical study process, Prescriptive modelling	Non-experimental (case studies)	Questionnaire	Median, average and frequency of responses, frequency charts	Reliability, construct validity, correctness, completeness, consistency, authenticity, unambiguousness, usability, validity, understandability, and flexibility.	62 case studies 1048 data points 19 experts	Cronbach's alpha, exploratory factor analysis, Spearman's correlation, frequency of responses, frequency charts, expert panel	Cronbach's alpha = 0.927 Evidence of convergent and discriminant validity, confirming a structure of 2 factors
MEEGA (SAVI et al., 2011)	Model	GQM, Scale development guide	Non-experimental (case studies)	Questionnaire	Frequency charts, median	Validity Reliability Applicability	3 case studies 79 data points	Correlation of items, item-total correlation, variance, mean, Cronbach's alpha	Cronbach's alpha = 0.915 Only evidence of convergent validity
EGameFlow (FU et al., 2009)	Scale	Scale development guide	Ad-hoc evaluation	Questionnaire	Mean, SD, Qualitative analysis	Item analysis Validity Reliability	1 application (4 games) 166 data points	Mean, SD, extreme group comparison, test for homogeneity, T-Test, ANOVA, Pearson's correlation, Cronbach's alpha	Cronbach's alpha = 0.800
(CARVALHO, 2012)	Framework	Not informed	Ad-hoc evaluation	Questionnaire, interview, tests	Not informed	Not informed	Not informed	Not informed	Not informed
(AK, 2012)	Scale	Not informed	Not informed	Questionnaire	Not informed	Not informed	Not informed	Not informed	Not informed
(MAYER, 12)	Model	Not informed	Quasi-experimental	Not informed	Not informed	Not informed	2164 data points	Not informed	Not informed
(CONNOLLY et al, 2009)	Framework	Not informed	Not informed	Not informed	Not informed	Not informed	2 studies	Not informed	Not informed
(FREITAS; OLIVER, 2006)	Framework	Not informed	Not informed	Not informed	Not informed	Not informed	2 game sessions	Not informed	Not informed
(CHEW, 2017)	Framework	Not informed	Ad-hoc evaluation	Questionnaire	Mean	Not informed	3 game sessions 37 data points	Not informed	Not informed
(ABDELLATIF et al., 2018)	Framework	Not informed	Ad-hoc evaluation	Questionnaire	Mean	Not informed	1 application 15 data points	Not informed	Not informed
(GARCIA-MUNDO et al., 2015)	Model	Top-down approach	Not informed	Not informed	Not informed	Not informed	Not informed	Not informed	Not informed

7.3 SCIENTIFIC PUBLICATIONS

During the development of this research, partial results have been published as journal/conference papers, book chapters, technical reports, etc. Table 58 presents the scientific publications produced in the context of this research, in alignment with the research objectives.

Table 58 - Publications

Id	Reference	Result	Qualis Computer Science
Specific objective O1. Identify the state-of-the-art of approaches used to evaluate educational games			
1	**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.	Direct result: Analysis of the State-of-the-art (Section 3.1)	B1
Specific objective O2. Identify the state-of-the-practice on how games for computing education are evaluated			
2	**PETRI, G.**, & GRESSE VON WANGENHEIM, C. (2017). How games for computing education are evaluated? A systematic literature review, *Computers & Education*, 107, pp. 68-90.	Direct result: Analysis of the State-of-the-practice (Section 3.2)	A1
Specific objective O3. Conduct a reanalysis in terms of reliability and validity of the initial version of the MEEGA model			
3	**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 (ICSE-SEET)* (pp. 180-189). Buenos Aires/Argentina.	Direct result: Analysis of the reliability and validity of the initial version of the MEEGA model (Chapter 4)	A1
4	**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 (ICSE-SEET)* (pp. 150-159). Buenos Aires/Argentina.	Indirect result: Analysis of the games evaluated using the initial version of the MEEGA model	A1
5	**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°*	Indirect result: Evaluation results of a game using the initial version of the MEEGA model	B1

	Simpósio Brasileiro de Informática na Educação (SBIE) (pp. 320-329). Uberlândia/MG.		
6	**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. *Regae – Revista de Gestão e Avaliação Educacional*, 7 (14), pp. 9-29.	Indirect result: Analysis of the games evaluated using the initial version of the MEEGA model	-
7	**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.	Indirect result: Analysis of digital games evaluated using the initial version of the MEEGA model	-
8	**PETRI, G.**, GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. (2016). *A Large-scale Evaluation of a Model for the Evaluation of Educational Games*. Technical Report. INCoD/GQS.04.2016.E (July/2016).	Direct result: Analysis of the reliability and validity of the initial version of the MEEGA model (Chapter 4)	-
Specific objective O4. Evolve the MEEGA model based on the results of the state-of-the-art and practice and the analysis of its initial version **Specific objective O5.** Develop a process that provides a systematic support for the evaluation of games for computing education			
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.	Direct result: Initial results of the MEEGA+ model (Section 5.1)	B3
10	**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.	Direct result: Design and initial evaluation of the MEEGA+ model (Section 5.1)	-
11	**PETRI, G.**, GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. (2016). *MEEGA+: An Evolution of a Model for the Evaluation of Educational Games*. Technical Report. INCoD/GQS.03.2016.E (July/2016).	Direct result: Initial design of the MEEGA+ model (Section 5.1)	-
12	PETRI, G., GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. (2018). *MEEGA+: A Method for the Evaluation of Educational Games for Computing Education*. Technical Report. INCoD/GQS.05.2018.E (July/2018).	Direct result: Design of the MEEGA+ Method (Chapter 5)	-
13	**PETRI, G.**, GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. (2018). *Evolution of a Model for the Evaluation of Games for Software Engineering Education*. In: Cooper, K. (Ed.)	Direct result: Initial results of the MEEGA+ model (Section 5.1 and 6.1)	- (accepted for publication)

	Software Engineering Perspectives in Computer Game Development.		
Specific objective O6. Apply and evaluate the evaluation method in different computing courses and educational institutions			
14	**PETRI, G.**, GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. Design and Evaluation of a Model for the Evaluation of Games for Computing Education. *Informatics in Education*.	Direct result: Design and initial evaluation of the MEEGA+ model (Section 5.1 and 6.1)	B1 (in review)
15	**PETRI, G.**, CALDERÓN, A, GRESSE VON WANGENHEIM, C., BORGATTO, A. F., & RUIZ, M. Games for Teaching Software Project Management: An Analysis of the Benefits of Digital and Non-Digital Games. *Journal of Universal Computer Science*.	Indirect result: Analysis of the games evaluated using the MEEGA+ model	B1 (accepted for publication)
16	**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.	Indirect result: Analysis of the games evaluated using the MEEGA+ model	B3
17	**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.	Indirect result: Analysis of a game evaluated using the MEEGA+ model	B3
18	CALDERÓN, A., **PETRI, G.**, RUIZ, M., GRESSE VON WANGENHEIM, C. (2018). Desarrollando Competencias Personales y Habilidades Sociales en Ingeniería Informática Mediante el Uso de Juegos Serios. *In Proc. of the XXIV Jornadas sobre la Enseñanza Universitaria de la Informática (JENUI)* (pp. 127-134). Barcelona, Espanha.	Indirect result: Analysis of the games evaluated using the MEEGA+ model	-
Publications as co-author			
19	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.	Evaluation results of a game using the initial version of the MEEGA model	B2
20	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.	Design and evaluation of the AssistantMEEGA+	B1
Other publications indirectly related to the research topic of this thesis			

21	GRESSE VON WANGENHEIM, C., **PETRI, G.,** ZIBETTI, A. W., BORGATTO, A. F., HAUCK, J. C. R., PACHECO, F. S., & MISSFELDT FILHO, R. (2017). dETECT: A Model for the Evaluation of Instructional Units for Teaching Computing in Middle School. *Informatics in Education*, 16(2), pp. 301-318.	B1
22	GONÇALVES, R. Q., GRESSE VON WANGENHEIM, C., HAUCK, J. C. R., & **PETRI, G**. (2017). An Instructional Feedback Technique for Teaching Project Management Tools Aligned with PMBOK. *Informatics in Education*, 16(2), pp. 197-224.	B1
23	GRESSE VON WANGENHEIM, C., **PETRI, G.**, ZIBETTI, A. W., HAUCK, J. C. R., PACHECO, F.S., & MISSFELDT FILHO, R. (2017). dETECT: Um Modelo para a Avaliação de Unidades Instrucionais para o Ensino de Computação na Educação Básica. Technical Report. INCoD/GQS.02.2017.P (May/2017).	-
24	MIOTO, F., GRESSE VON WANGENHEIM, C., PACHECO, L. H. M., & **PETRI, G.** Como avaliar habilidades do século XXI no contexto do ensino de computação no ensino básico? Um mapeamento sistemático da literatura. *Revista Brasileira de Informática na Educação*.	B3 (in review)
25	MIOTO, F., **PETRI, G.**, GRESSE VON WANGENHEIM, C., & BORGATTO, A. F. (2018). bASES21 - um Modelo de Autoavaliação de Habilidades do Século XXI no Contexto do Ensino de Computação na Educação Básica, *Revista Brasileira de Informática na Educação*.	B3 (accepted for publication)
26	GRESSE VON WANGENHEIM, C., **PETRI, G.**, BORGATTO, A. F. (2017). MEEGA+KIDS: A Model for the Evaluation of Educational Games for Computing Education in Secondary School (Draft Version). Technical Report. INCoD/GQS.06.2018.E (August/2018).	-

Source: developed by the author.

8 CONCLUSIONS AND FUTURE WORK

As a result of this research, we develop and evaluate an evaluation method, the MEEGA+, providing a comprehensive support for quality evaluations of games used as an instructional strategy for computing education. The MEEGA+ method 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, guiding researchers on how to plan, execute and analyse the results of game evaluations. Based on a large-scale evaluation, the MEEGA+ measurement instrument has shown an excellent reliability through a Cronbach's alpha coefficient α=.927. In addition, results of validity analysis, through a factor analysis, confirm the original structure of the MEEGA+ model, indicating that the quality of games for computing education is evaluated through two quality factors: usability and player experience. The quality of the MEEGA+ method has also been confirmed by a panel of 19 experts.

Therefore, the main contribution of this research is the MEEGA+ method, being a well-defined and evaluated method, providing a systematic and comprehensive support for quality evaluations 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.

We intend to continue conducting case studies evaluating games (digital and non-digital ones) for computing/SE education using the MEEGA+ method in order to widely confirm its validity and reliability. In addition, we plan to contribute to the evolution of the AssistantMEEGA+, including all features available in the MEEGA+ method, in order to automatize the all activities needed to conduct a game evaluation. We also intend to study the analysis of additional factors to be considered in the evaluation, such as measures of the players' interactions with the game and evaluation of the learning impact. Moreover, we are currently analysing the correlation between measures of perceived learning and the learning impact when adopting games for computing education.

Although the emphasis of the MEEGA+ method is on the evaluation of games used for computing/SE education, we assume that the MEEGA+ method can be used and adapted for the evaluation of games to teach others knowledge areas, as performed by Gomes (2016), Herpich et al. (2017), Pereira et al. (2017), Silva et al. (2017a), Silva et al. (2017b), and Moosa, Al-Maadeed, and AlJa'am (2018). However, when transferring the method to other knowledge areas, further empirical studies are necessary to evaluate and confirm the reliability and validity of the MEEGA+ method also in these areas.

REFERENCES

ABDELLATIF, A. J.; MCCOLLUM, B.; MCMULLAN, P. Serious Games: Quality Characteristics Evaluation Framework and Case Study. In: INTEGRATED STEM EDUCATION CONFERENCE, 2018, Princeton. **Proceedings…**. Princeton, USA, 2018. p. 112-119.

ABT, C. C. **Serious Games**. Lanhan: University Press of America, 2002.

ACM; AIS; IEEE. **Computing Curricula 2005:** The Overview Report, 2005. Available at: <https://www.acm.org/binaries/content/assets/education/curricula-recommendations/cc2005-march06final.pdf> Acess: 13 sep 2018.

ACM; IEEE-CS. **Computer Science Curricula 2013**: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science, 2013. Available at: < http://www.acm.org/education/CS2013-final-report.pdf> Acess: 10 mar 2015.

ACUÑA, S. T.; ANTONIO, A.; FERRÉ, X.; LÓPEZ, M.; MATÉ, L. The Software process: modeling, evaluation and improvement. **Handbook of Software Engineering and Knowledge Engineering**. Singapura: World Scientific Publishing Company, 2000.

ACUÑA, S. T.; FERRÉ, X. Software Process Modelling. In: WORLD MULTICONFERENCE ON SYSTEMICS, CYBERNETICS AND INFORMATICS, 2001, Orlando. **Proceedings…**. Orlando, EUA, 2001. p. 1-6.

ADAMS, E; ROLLINGS, A. **Fundamentals of game design**. Saddle River: Prentice Hall, 2006.

AK, O. A Game Scale to Evaluate Educational Computer Games. **Procedia - Social and Behavioral Sciences**, v. 46, p. 2477-2481, 2012.

ALL, A.; CASTELLAR, E. P. N.; LOOY, J. V. Assessing the effectiveness of digital game-based learning: Best practices. **Computers & Education**, v. 92–93, p. 90-103, 2016.

ANDERSON, L. W.; KRATHWOHL, D. R.; BLOOM, B. S. **A taxonomy for learning, teaching, and assessing: a revision of**

Bloom's taxonomy of educational objectives. London: Longman, 2001.

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria de Resposta ao Item: conceitos e aplicações.** ABE — Associação Brasileira de Estatística, 4º SINAPE, 2000.

ANDRADE, H.; VALTCHEVA, A. Promoting learning and achievement through self-assessment. **Theory into Practice**, v. 48, p. 12-19, 2009.

BACKLUND, P.; HENDRIX, M. Educational games - Are they worth the effort? A literature survey of the effectiveness of serious games. In: INTERNATIONAL CONFERENCE ON GAMES AND VIRTUAL WORLDS FOR SERIOUS APPLICATIONS, 5., 2013, Poole. **Proceedings…**. Poole: IEEE, 2013. p. 1-8.

BAKER, A.; NAVARRO, E. O.; VAN DER HOEK, A. Problems and Programmers: An Educational Software Engineering Card Game. In: INTERNATINAL CONFERENCE ON SOFTWARE ENGINEERING, 25., 2003, Portland. **Proceedings…**. Portland, USA: IEEE/ACM, 2003, p. 614-619.

BARUCH, Y. Response Rate in Academic Studies – A Comparative Analysis. **Human Relations**, v. 52, n. 4, 1999.

BASILI, V. R.; CALDIERA, G.; ROMBACH, H. D. **Goal, Question Metric Paradigm**. In: MARCINIAK, J. J. (Ed.). Encyclopedia of Software Engineering. John Wiley & Sons, 1994. p. 528–532.

BATTISTELLA, P. E.; GRESSE VON WANGENHEIM, C. **Systematic Literature Review (SLR) of Game-based Learning for Computing**. Technical Report. INCoD – Nº 002/2014 – E – GQS, August, 2014.

BATTISTELLA, P. E.; GRESSE VON WANGENHEIM, C. Games for Teaching Computing in Higher Education – A Systematic Review. **IEEE Technology and Engineering Education (ITEE) Journal**, v. 9, n. 1, p. 8-30, 2016.

BATTISTELLA, P. E.; GRESSE VON WANGENHEIM, C.; VON WANGENHEIM, A.; MARTINA, J. E. Design and Large-scale Evaluation of Educational Games for Teaching Sorting Algorithms. **Informatics in Education**, v. 16, n. 2, p. 141-164, 2017.

BAVOTA, G.; LUCIA, A.; FASANO, F.; OLIVETO, R.; ZOTTOLI, C. Teaching software engineering and software project management: an integrated and practical approach. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING, 34., 2012, Piscataway. **Proceedings…**. Piscataway, USA: IEEE Press, 2012. p. 1155-1164.

BEDNARIK, R.; GERDT, P.; MIRAFTABI, R.; TUKIAINEN, M. Development of the TUP Model – Evaluating Educational Software. In: IEEE INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 4., 2004, Joensuu. **Proceedings…**. Joensuu, Finland: IEEE, 2004. p. 699–701.

BEECHAM, S.; HALL, T.; BRITTON, C.; COTTEE, M.; RAINER, A. **Using an Expert Panel to Validate a Requirements Process Improvement Model.** The Journal of Systems and Software, v. 76, 2005.

BELL, J.; SHETH, S.; KAISER, G. Secret ninja testing with HALO software engineering. In: INTERNATIONAL WORKSHOP ON SOCIAL SOFTWARE ENGINEERING, 4. 2011, Szeged. **Proceedings…**. Szeged, Hungary: ACM, 2011. p. 43-47.

BENALI K.; DERNIAME J.C. Software processes modeling: What, who, and when. In: Derniame JC. (Eds) **Software Process Technology.** Lecture Notes in Computer Science, vol 635. Springer, Berlin, Heidelberg, 1992.

BLOOM, B. S. **Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain**. New York: Toronto: Longmans, Green, 1956.

BOURQUE, P.; FAIRLEY, R. E. **Guide to the software engineering body of knowledge** (SWEBOK (R)): Version 3.0. IEEE Computer Society Press, 2014.

BOWMAN, D. D. Declining Talent in Computer Related Careers. **Journal of Academic Administration in Higher Education**, v. 14, n. 1, p. 1-4, 2018.

BOYLE, E. A.; HAINEY, T.; CONNOLLY, T. M.; GRAY, G.; EARP, J.; OTT, M.; LIM, T.; NINAUS, M.; RIBEIRO, C.; PEREIRA, J. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. **Computers & Education**, v. 94, p. 178-192, 2016.

BRANCH, R. **Instructional design: The ADDIE approach**. 2nd ed. Athens: Springer, 2009.

BRANDÃO, H. P.; BORGES-ANDRADE, J. E. Causas e efeitos da expressão de competências no trabalho: para entender melhor a noção de competência. **Revista de Administração Mackenzie**, v. 8, n. 3, p. 32–49, 2007.

BRANDÃO, H. P.; GUIMARÃES, T. A. Gestão de competências e gestão de desempenho: tecnologias distintas ou instrumentos de um mesmo constructo? **Revista de Administração de Empresas (FGV)**, v. 41, n. 1, p. 08–15, 2001.

BRANSFORD, J.; BROWN, A.; COCKING, R. (2000). **How people learn: Brain, mind, experience, and school**. 2nd ed. Washington: National Academies Press, 2000.

BROOKE, J. SUS: A "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (Eds.), **Usability evaluation in industry.** Taylor and Francis, London, 1996.

BROWN, G.; ANDRADE, H.; CHEN, F. Accuracy in student self-assessment: Directions and cautions for research. **Assessment in Education Principles Policy and Practice**, v. 22, n. 4, p. 1-26, 2015.

BROWN, T. A. **Confirmatory factor analysis for applied research**. New York: The Guilford Press, 2006.

BRUNER, J. S. **The process of education.** Cambridge: Harvard University Press, 1977.

BRUNO-FARIA, M. F.; BRANDÃO, H. P. Competências Relevantes a Profissionais da área de T&D de uma Organização Pública do Distrito Federal. **Revista de Administração Contemporânea**, v. 7, p. 35-56, 2003.

BUDGEN, D.; TURNER, M.; BRERETON, P.; KITCHENHAM, B. Using mapping studies in Software Engineering. In: ANNUAL WORKSHOP OF PSYCHOLOGY OF PROGRAMMING INTEREST GROUP, 20., 2008, Lancaster. **Proceedings…**. Lancaster, USA, 2008. p. 195-204.

CALDERÓN A.; RUIZ M. A Systematic Literature Review on Serious Games Evaluation: an Application to Software Project Management. **Computers & Education**, v. 87, p. 396-422, 2015.

CALDERÓN, A.; RUIZ M.; O'CONNOR, R. A multivocal literature review on serious games for software process standards education. **Computer Standards & Interfaces**, v. 57, p. 36-48, 2018.

CALDERÓN, A.; RUIZ, M. ProDec: a serious game for software project management training. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING ADVANCES, 8., 2013, Venice. **Proceedings…**. Venice, Italy, 2013. p. 565-570.

CARMINES, E. G.; ZELLER, R. A. **Reliability and validity assessment.** 5th ed. Beverly Hills: Sage Publications Inc., 1982.

CARVALHO, C. V. Is game-based learning suitable for engineering education? In: GLOBAL ENGINEERING EDUCATION CONFERENCE, 2012. Marrakech. **Proceedings…**. Marrakech, Morocco: IEEE, 2012. p.1-8.

CATTELL, R. B. **The Scientific Use of Factor Analysis in Behavioral and Life Sciences**. Nova York: Plenum, 1978.

CATTELL. R. B. The Scree Test for the Number of Factors. **Multivariate Behavioral Research**, v. 1, n. 2, p. 245-276, 1966.

CAULFIELD, C.; XIA, J.; VEAL, D.; MAJ, S. P. A systematic survey of games used for software engineering education. **Modern Applied Science**, v. 5, n. 6, p. 28-43, 2011.

CHEN, P. Y.; POPOVICH, P. M. **Correlation: Parametric and nonparametric measures**. Thousand Oaks: Sage Publications, 2002.

CHEW, B. An efficient framework for game-based learning activity. In: IEEE INTERNATIONAL CONFERENCE ON TEACHING, ASSESSMENT, AND LEARNING FOR ENGINEERING, 6., 2017, Hong Kong. **Proceedings...**. Hong Kong, China: IEEE, 2017. p. 147-150.

COELHO, M. P. C.; FUERTH, L. R. A Influência da Gestão por Competência no Desenvolvimento Profissional. **Revista Cadernos de Administração**, v. 1, n. 3, 2009.

COHEN, J. **Statistical Power Analysis for the Behavioral Sciences**. New York: Routledge Academic, 1998.

CONNOLLY, T. M.; BOYLE, E. A.; MACARTHUR, E.; HAINEY, T.; BOYLE, J. M. A systematic literature review of empirical evidence on computer games and serious games. **Computers & Education**, v. 59, n. 2, p. 661-686, 2012.

CONNOLLY, T. M.; STANSFIELD, M. H.; HAINEY, T. Development a General Framework for Evaluating Games-based learning. In: EUROPEAN CONFERENCE ON GAMES-BASED LEARNING, 2., 2008. Barcelona. **Proceedings...**. Barcelona, Spain, 2008. p. 105-114.

CONNOLLY, T. M.; STANSFIELD, M. H.; HAINEY, T. **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, Idea-Group Publishing: Hershey, 2009.

CONNOLLY, T. M.; STANSFIELD, M.; HAINEY, T. An application of games-based learning within software engineering. **British Journal of Educational Technology**, v. 38, p. 416-428, 2007.

COSENZA, R. N.; GUERRA, L. B. **Neurociência e Educaçao:** como o cérebro aprende. Porto Alegre: Artmed, 2011.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. **Psychometrika**, v. 16, n. 3, p. 297–334, 1951.

CUESTA, M. **Unidimensionalidade**. In J. Muñiz (Ed.), Psicometría. Madrid: Editorial Universitas, 1996.

DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. **MIS Quarterly**, v. 13, n. 3, p. 319-340, 1989.

DAVIS, F. D.; BAGOZZI, R. P.; WARSHAW, P. R. User acceptance of computer technology: A comparison of two theoretical model. **Manage. Sci.**, v. 35, n. 8, p.982-1003, 1989.

DAWES, J. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. **International Journal of Market Research**, v. 50, n. 1, p. 61-77, 2008.

DEVELLIS, R. F. **Scale development: theory and applications.** 4th ed. Thousand Oaks: SAGE Publications, 2016.

DEVELLIS, R. F. **Scale development: theory and applications.** Thousand Oaks: SAGE Publications, 2003.

DICK, W., & CAREY, L. **The systematic design of instruction**. New York: Haper Collins College Publishers, 2006.

DJAOUTI, D.; ALVAREZ J.; JESSEL J. P.; RAMPNOUX O. **Origins of Serious Games**. In: Ma M., Oikonomou A., Jain L. (Eds) Serious Games and Edutainment Applications. London: Springer, 2011.

DOLAN, E. L.; COLLINS, J. P. We must teach more effectively: here are four ways to get started. **Molecular Biology of the Cell**, v. 26, n. 12, p. 2151-2155, 2015.

DRISCOLL, M. P. **Psychology of Learning for Instruction**. Boston: Allyn & Bacon, 1994.

DURAND, T. L'alchimie de la Compétence. **Revue Française de Gestion**, v. 32, n. 160, p. 2-30, 1999.

DZIUBAN, C. D.; SHIRKEY, E. C. When is a correlation matrix appropriate for factor analysis? Some decision rules. **Psychological Bulletin**, v. 81, p. 358-361, 1974.

EAGLE, M.; BARNES, T. Experimental evaluation of an educational game for improved learning in introductory computing. In: ACM TECHNICAL SYMPOSIUM ON COMPUTER SCIENCE EDUCATION, 40., 2009, Chattanooga. **Proceedings…**. Chattanooga, USA: ACM, 2009. p. 321-325.

FENTON, N.E., PFLEEGER, S. L. **Software Metrics: A Rigorous and Practical Approach**. 2nd ed. Boston: PWS Pub. Co., 1998.

FERNÁNDEZ, F. A. **Didática y optimización del processo de ensenãnza-aprendizaje.** Instituto Pedagógico Latinoamericano y Caribeño. La Havana, Cuba, 1998.

FILATRO, A. **Design Instrucional na Prática.** São Paulo: Pearson Education do Brasil, 2008.

FINKELSTEIN, A.; KRAMER, J.; NUSEIBEH, B. **Software Process Modelling and Technology**. Research Studies Press, 1994.

FRANCH, X.; CARVALLO, J. P. Using quality models in software package selection. **IEEE Software**, v. 20, n. 1, p. 34-41, 2003.

FREEDMAN, D.; PISANI, R.; PURVES, R. **Statistics**. 4th ed. New York: W. W. Norton & Company, 2007. 720p.

FREEMAN, S.; EDDY, S. L.; MCDONOUGH, M.; SMITH, M. K.; OKOROAFOR, N.; JORDT, H.; WENDEROTH, M. P. Active learning increases student performance in science, engineering, and mathematics. **National Academy of Sciences of the United States of America**, v. 111, n. 23, p. 8410-8415, 2014.

FREITAS, I. A.; BRANDÃO, H. P. Trilhas de aprendizagem como estratégia para desenvolvimento de competências. In: ENCONTRO ANUAL DA ASSOCIAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO – ENANPAD, 2005. Brasília. **Proceedings....** Brasília, 2005, p. 1-16.

FREITAS, S. D.; OLIVER, M. How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? **Computers & Education**, v. 46, n. 3, p. 249-264, 2006.

FU, F.; SU, R.; YU, S. EGameFlow: A scale to measure learners' enjoyment of e-learning games. **Computers & Education**, v. 52, n. 1, p. 101-112, 2009.

GAGNE, R. M.; BRIGGS, L. J.; WAGER, W. W. **Principles of instructional design**. 4th ed. Fort Worth: Harcourt Brace Jovanovich, 1992.

GÁMEZ, E. H. C. **On the Core Elements of the Experience of Playing Video Games**. PhD Thesis. Interaction Centre Department of Computer Science, University College London, GB, 2009.

GARCIA-MUNDO, L.; GENERO, M.; PIATTINI, M. Towards a Construction and Validation of a Serious Game Product Quality Model. In: INTERNATIONAL CONFERENCE ON GAMES AND VIRTUAL WORLDS FOR SERIOUS APPLICATIONS, 7., 2015, Skovde. **Proceedings…**. Skovde, Sweden, 2015. p. 1-8.

GARSON, D. **Factor Analysis**. Raleigh: Statistical Associates Publishing, 2013.

GHEZZI, C.; MANDRIOLI, D. The challenges of software engineering education. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING EDUCATION IN THE MODERN AGE, 27., St. Louis. **Proceedings…**. St. Louis, USA: Springer-Verlar, 2005. p. 115-127.

GIBSON, B.; BELL, T. Evaluation of games for teaching computer science. In: WORKSHOP IN PRIMARY AND SECONDARY COMPUTING EDUCATION, 8., 2013. New York. **Proceedings…**. New York: ACM, 2013. p. 51-60.

GLOGER, B. **Ball Point Game**. Available at: <https://borisgloger.com/wp-cotent/uploads/2016/08/Ball_Point_Game.pdf> Access: 27 November 2017.

GLORFELD, L. W. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. **Educational and Psychological Measurement**, v. 55, n. 3, p. 377-393, 1995.

GOMES, M. N. **Desenvolvimento de uma Atividade Gamificada Voltada ao Turismo de Caxias Do Sul**. Undergraduate Thesis (Computer Science Program), Universidade de Caxias do Sul, 2016.

GOUWS, L. A.; BRADSHAW, K.; WENTWORTH, P. Computational thinking in educational activities: an evaluation of the educational game light-bot. In: ACM CONFERENCE ON INNOVATION AND TECHNOLOGY IN COMPUTER SCIENCE EDUCATION, 18., 2013, New York. **Proceedings…**. New York, USA: ACM, 2013. p. 10-15.

GRAF, S.; VIOLA, S. R.; LEO, T.; KINSHUK. In-depth analysis of the Felder-Silverman learning style dimensions. **Journal of Research on Technology in Education**, v. 40, n. 1, p79-93, 2007.
GRESSE VON WANGENHEIM, C. **PM Master**. Available at: <http://www.gqs.ufsc.br/pm-master/> Access: 06 Feb 2018.

GRESSE VON WANGENHEIM, C.; CARVALHO, O. P.; BATTISTELLA, P. E. Ensinar a Gerência de Equipes em Disciplinas de Gerência de Projetos de Software. **Revista Brasileira de Informática na Educação**, v. 21, n. 1, p. 15-22, 2013.

GRESSE VON WANGENHEIM, C.; KOCHANSKI, D.; SAVI, R. Revisão Sistemática sobre Avaliação de Jogos Voltados para a Aprendizagem de Engenharia de Software no Brasil. In: FÓRUM DE EDUCAÇÃO EM ENGENHARIA DE SOFTWARE, 2009. Fortaleza. **Proceedings...**. Fortaleza, Brazil, 2009. p. 1-8.

GRESSE VON WANGENHEIM, C.; PETRI, G.; BORGATTO, A. F. **MEEGA+KIDS: A Model for the Evaluation of Educational Games for Computing Education in Secondary School (Draft Version)**. Technical Report. INCoD/GQS.06.2018.E. August, 2018.

GRESSE VON WANGENHEIM, C.; RAUSIS, B.; SOARES, G.; SAVI, R.; BORGATTO, A. F. Project Detective a Game for Teaching Earned Value Management. **International Journal of Teaching and Case Studies**, v. 5, n. 3/4, p. 216-234, 2014.

GRESSE VON WANGENHEIM, C.; SAVI, R.; BORGATTO, A. F. DELIVER! – An educational game for teaching Earned Value Management in computing courses. **Information and Software Technology**, v. 54, n. 3, p. 286-298, 2012.

GRESSE VON WANGENHEIM, C.; SAVI, R.; BORGATTO, A. F. SCRUMIA - An educational game for teaching SCRUM in computing courses. **Journal of Systems and Software**, v. 86, n. 10, p. 2675-2687, 2013.

GRESSE VON WANGENHEIM, C.; SHULL, F. To Game or Not to Game? **IEEE Software**, v. 26, n. 2, p. 92-94, 2009.

GRESSE VON WANGENHEIM, C.; THIRY, M.; KOCHANSKI, D. Empirical evaluation of an educational game on software measurement. **Empirical Software Engineering**, v. 14, p. 418–452, 2009.

GRESSE VON WANGENHEIM, C.; VON WANGENHEIM, A. **Ensinando Computação com Jogos.** Florianópolis: Bookes, 2012.

GSI TEACHING & RESOURCE CENTER. **Learning: Theory and Research**. 2015. Available at: <http://gsi.berkeley.edu/gsi-guide-contents/learning-theory-research/>. Acess: 30 mar 2015.

HAINEY, T.; CONNOLLY, T. M.; BOYLE, E. A. A Refined Evaluation Framework for Games-based Learning. In: EUROPEAN CONFERENCE ON GAMES-BASED LEARNING, 4., 2010. Copenhagen. **Proceedings…**. Copenhagen, Denmark, 2010. p 1-11.

HAINEY, T.; CONNOLLY, T. M.; STANSFIELD, M.; BOYLE, E. A. Evaluation of a game to teach requirements collection and analysis in software engineering at tertiary education level. **Computers & Education**, v. 56. n. 1, p. 21-35, 2011.

HAKULINEN, L. Using serious games in computer science education. In: INTERNATIONAL CONFERENCE ON COMPUTING EDUCATION RESEARCH, 11., 2011. Koli. **Proceedings…**. New York: ACM, 2011. p. 83-88.

HAMBLETON, R. K., SWAMINATHAN, H. AND ROGERS, H. J. **Fundamentals of Item Response Theory**. Newbury Park: Sage Publications, 1991.

HAMEY, L. G. C. Using the Security Protocol Game to teach computer network security. In: IMPROVING LEARNING OUTCOMES THROUGH FLEXIBLE SCIENCE TEACHING, 2003, Sydney. **Proceedings…**. Sydney, Australia, 2003. p. 96-101.

HAYS, R.T. **The Effectiveness of Instructional Games: A Literature Review and Discussion**. Naval Air Warfare Center Training System Division, Orlando, FL, USA, 2005.

HERCULANO-HOUZEL, S. **Fique de bem com seu cérebro**. Sextante, 2007.

HERPICH, F.; NUNES, F. B.; VOSS, G. B.; SINDEAUX, P.; TAROUCO, L. M. R.; LIMA, J. V. Realidade Aumentada em Geografia: uma atividade de orientação no ensino fundamental. **Revista Novas Tecnologias na Educação**, v. 15, n. 2, p. 1–10, 2017.

HERZ, J. C. **Joystick nation: how videogames ate our quarters, won our hearts, and rewired our minds.** Boston: Little, Brown and Company, 1997.

HEVNER, A.; CHATTERJEE, S. **Design Research in Information Systems: Theory and Practice**. Integrated Series in Information Systems, Springer, 2010.

HILGARD, E. R. **Teorias da Aprendizagem**. 4. ed. São Paulo: Editora Pedagógica e Universitária., 1975.

HUANG, W.; HUANG, W.; DIEFES-DUX, H.; IMBRIE, P. K. A Preliminary Validation of Attention, Relevance, Confidence and Satisfaction Model-Based Instructional Material Motivational Survey in a Computer-Based Tutorial Setting. **British Journal of Educational Technology**, vol. 37, no. 2, p. 243-259, 2006.

IBRAHIM, R.; YUSOFF, R. C. M.; OMAR, H. M.; JAAFAR, A. Students perceptions of using educational games to learn introductory programming. **Computer and Information Science**, v. 4, n. 1, p. 205-216, 2011.

IEEE. **Standard Glossary of Software Engineering Terminology**. Std 610.12-1990(R2002), 2002.

IEEE. **Systems and Software Engineering – Vocabulary.** ISO/IEC/IEEE 24765:2010(E), pp.1-418, 2010.

INTERNATIONAL STANDARD ORGANIZATION (ISO). **ISO/IEC 25010: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models**, Tech. Rep., 2011.

IZQUIERDO, M. C.; PEDRERO, E. F. Estimating the reliability coefficient of tests in presence of missing values. **Psicothema**, v. 26, n. 4, p. 516-523, 2014.

JEDLITSCHKA A.; CIOLKOWSKI M.; PFAHL D. **Reporting Experiments in Software Engineering**. In: Shull F., Singer J., Sjøberg D.I.K. (eds) Guide to Advanced Empirical Software Engineering. London: Springer, 2008.

JOHNSON, B.; CHRISTENSEN, L. **Educational Research: Quantitative, Qualitative, and Mixed Approaches.** 6th ed. Thousand Oaks: SAGE Publications, Inc., 2016.

JOHNSON, M. A pilot study examining the motivational effect of instructional materials on EFL learning motivation. **Journal of Language and Culture of Hokkaido**, vol. 10, p. 39-47, 2012.

JORDAN, C.; KNAPP, M.; MITCHELL, D.; CLAYPOOL, M.; FISLER, K. **CounterMeasures: a game for teaching computer security**. In: ANNUAL WORKSHOP ON NETWORK AND SYSTEMS SUPPORT FOR GAMES, 10., 2011. Piscataway. **Proceedings...**. Piscataway: IEEE, 2011. Article 7, 6 pages.

KARNER, F. W.; HÄRTEL, G. **Theory and Taxonomies of Serious Games.** ENTRExplorer Project, 2011. Available at: http://www.entrexplorer.com/pdf/Theory_and_Taxonomies_of_Serious_Games.pdf. Acess: 15 may 2015.

KASUNIC, M. **Designing an Effective Survey.** Hadbook CMU/SEI 2005-HB-004, Software Engineering Institute/Carnegie Mellon, 2005.

KAZIMOGLU, C.; KIERNAN, M.; BACON, L.; MACKINNON, L. A Serious Game for Developing Computational Thinking and Learning Introductory Computer Programming. **Procedia - Social and Behavioral Sciences**, v. 47, p. 1991-1999, 2012.

KELLER, J. Development and Use of the ARCS Model of motivational Design. **Journal of Instructional Development**, v. 10, n. 3, p. 2-10, 1987.

KELLER, J. **Motivational Design for Learning and Performance: The ARCS Model Approach.** Springer, 2009.

KIMBERLIN, C. L.; WINTERSTEIN, A. G. Validity and reliability of measurement instruments used in research. **American Journal of Health-System Pharmacy**, v. 65, n. 23, p. 2276-84, 2008.

KIRKPATRICK, D. L.; KIRKPATRICK, J. D. **Evaluating training programs: the four levels.** 3. ed. USA: Berrett-Koehler Publishers, USA, 2006.

KISH, L. Multipopulation Survey Designs: Five Types with Seven Shared Aspects. **International Statistical Review**, v. 62, n. 2, p. 167-186, 1994.

KITCHENHAM, B. et al. Preliminary guidelines for empirical research in software engineering. **IEEE Transactions on Software Engineering**, v. 28, n. 8, p. 721–734, 2002.

KITCHENHAM, B. Systematic literature reviews in software engineering – A tertiary study**. Information and Software Technology**, v. 52, n. 1, p. 792-805, 2010.

KITCHENHAM, B.; PFLEEGER, S. L.; FENTON, N. Towards a Framework for Software Measurement Validation. **IEEE Transactions on Software Engineering**, v. 21, n. 12, p. 929-944, 1995.

KOLB, D. A.; FRY, R. **Toward an applied theory of experiential learning**. London: John Wiley, 1975.

KORDAKI, M., GOUSIOU, A. Computer Card Games in Computer Science Education: A 10-Year Review. **Journal of Educational Technology & Society**, v. 19, n. 4, p. 11-21, 2016.

KORDAKI, M., GOUSIOU, A. Digital card games in education: A ten year systematic review. **Computers & Education**, v. 109, p. 122-161, 2017.

KOSA, M.; YILMAZ, M.; O'CONNOR, R.; CLARKE, P. Software engineering education and games: a systematic literature review. **Journal of Universal Computer Science**, v. 22, n. 12, p. 1558-1574, 2016.

KRATHWOHL, D. R.; BLOOM, B. S.; MASIA, B. B. **Taxonomy of Educational Objectives, the Classification of Educational Goals**. Handbook II: Affective Domain. New York, 1973.

MALHOTRA, N. K.; BIRKS, D. F. **Marketing Research: An Applied Approach.** 3rd ed. Philadelphia: Trans-Atlantic Publications, 2008.

MARCONI, M. A.; LAKATOS, E. M**. Fundamentos de metodologia científica.** 7. ed. São Paulo: Atlas, 2010.

MATOOK, S.; INDULSKA, M. Improving the Quality of Process Reference Models: A Quality Function Deployment-Based Approach. **Decision Support Systems**, v. 47, 2009.

MAYER, I. Towards a Comprehensive Methodology for the Research and Evaluation of Serious Games, **Procedia Computer Science**, v. 15, p. 233-247, 2012.

MAYER, I.; BEKEBREDE, G.; HARTEVELD, C.; WARMELINK, H.; ZHOU, Q.; RUIJVEN, T.; LO, J.; KORTMANN, R.; WENZLER, I. The research and evaluation of serious games: Toward a comprehensive methodology, **British Journal of Educational Technology**, v. 45, n. 3, p. 502-527, 2014.

MAYER, I.; BEKEBREDE, G.; WARMELINK, H.; ZHOU, Q. **A Brief Methodology for Researching and Evaluating Serious Games and Game-Based Learning**. In Coonolly et al., (Eds) Psychology, Pedagogy, and Assessment in Serious Games (pp.357-393). IGI Global, 2013.

MELERO, J.; HERNÁNDEZ-LEO, D.; BLAT, J. Considerations for the design of mini-games integrating hints for puzzle solving ICT-related concepts. In: IEEE INERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 12., 2012, Rome. **Proceedings…**. Rome, Italy: IEEE, 2012. p. 154-158.

MERRILL, M. D. et al. **Reclaiming instructional design.** Educational Technology, v. 36, n. 5, p. 5–7, 1996.

MITAMURA, T.; SUZUKI, Y.; OOHORI, T.. Serious Games for Learning Programming Languages. In: INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CYBERNETICS, 2012, Seoul. **Proceedings…**. Seoul, Korea, 2012. p. 1812-1817.

MOHAMED, H.; JAAFAR, A. Development and Potential Analysis of Heuristic Evaluation for Educational Computer Game (PHEG). In: INTERNATIONAL CONFERENCE ON COMPUTER SCIENCES AND CONVERGENCE INFORMATION TECHNOLOGY, 5., 2010, Seoul. **Proceedings…**. Seoul, South Korea: IEEE, 2010. p. 222-227.

MOOSA, A. M.; AL-MAADEED, N.; ALJA'AM, J. M. A Simple Health-Based Game for Children. In: INTERNATIONAL CONFERENCE ON COMPUTER AND APPLICATIONS. 2018, Beirut. **Proceedings…**. Beirut, Lebanon, 2018. p. 309-312.

MORRISON, G., ROSS, S., KEMP, J., & KALMAN, H. **Designing Effective Instruction** (6th ed.). USA: John Wiley & Sons, 2010.

MUMMS. **Measuring the usability of multi-media systems.** Available at: http://www.ucc.ie/hfrg/questionnaires/mumms/index.html. Access: 31 Aug 2015.

MÜNCH, J.; ARMBRUST, O.; KOWALCZYK, M.; SOTO, M. **Software Process Definition and Management**. New York: Springer-Verlag Berlin Heidelberg, 2012.

NAVARRO, E. O.; VAN DER HOEK, A. Design and Evaluation of an Educational Software Process Simulation Environment and Associated Model. In: INTERNATIONAL CONFERENCE ON SOFTWARE

ENGINEERING EDUCATION AND TRAINING, 18. 2005, Ottawa. **Proceedings…**. Ottawa, Canada, 2005. p. 25-32.

NAVARRO, E.; VAN DER HOEK, A. **Comprehensive Evaluation of an Educational Software Engineering Simulation Environment**. In: CONFERENCE ON SOFTWARE ENGINEERING EDUCATION & TRAINING, 20., 2007. Dublin. **Proceedings…**. Dublin: IEEE, 2007. p.195-202.

NAVARRO, E.; VAN DER HOEK, A. **Multi-site evaluation of SimSE**. In: TECHNICAL SYMPOSIUM ON COMPUTER SCIENCE EDUCATION, 40., 2009. Chattanooga. **Proceedings…**. New York: ACM, 2009. p. 326-330.

O'BRIEN, H. L.; TOMS, E. G. The Development and Evaluation of a Survey to Measure User Engagement. **Journal of the American Society for Information Science and Technology**, 61(1), 50–69, 2010.

OLIVEIRA, M. K. DE. **Vygotsky - Aprendizado e desenvolvimento:** um processo sócio-histórico. São Paulo: Scipione, 1993.

PARSONS, P. Preparing computer science graduates for the 21st Century. **Teaching Innovation Projects**, v. 1, n. 1, article 8, 2011.

PASQUALI, L.; PRIMI, R. Fundamentos da Teoria da Resposta ao Item –TRI. **Avaliação Psicológica**, v. 2, n. 2, p. 99-110, 2003.

PEREIRA, C. X.; HOFFMANN, Z.; CASTRO, C. P.; SANTOS, G. S.; AIRES, T. A.; FRANCISCO, R. E. CHEMIS3: A game for learning chemical concepts through elements of nature. In: IBERIAN CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGIES, 12., 2017, Lisbon. **Proceedings…**. Lisbon, Portugal, 2017. p. 1-5.

PETERSEN, K. et al. Systematic Mapping Studies in Software Engineering. In: INTERNATIONAL CONFERENCE ON EVALUATION AND ASSESSMENT IN SOFTWARE ENGINEERING, 12., 2008. **Proceedings....** Italy: BCS Learning & Development, 2008.

PETRI, G.; BATTISTELLA, P.; CASSETTARI, F.; GRESSE VON WANGENHEIM, C.; HAUCK, J. Um Quiz Game para a revisão de conhecimento em Gerenciamento de Projetos. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 27., 2016. Uberlândia. **Proceedings...**. Uberlândia, Brazil, 2016. p. 320-329.

PETRI, G.; CHIAVEGATTI, N. C. Um Role Playing Game para o Ensino de Elicitação e Análise de Requisitos. **Revista Novas Tecnologias na Educação**, v. 13, n. 1, p. 1-10, 2015.

PETRI, G.; GRESSE VON WANGENHEIM, C. How games for computing education are evaluated? A systematic literature review. **Computers & Education**, vol. 107, 2017, pp. 68-90.

PETRI, G.; GRESSE VON WANGENHEIM, C. How to evaluate educational games: a systematic literature review. **Journal of Universal Computers Science**, vol. 22, no. 7, 2016, pp. 992-1021.

PETRI, G.; GRESSE VON WANGENHEIM, C.; BORGATTO, A. F. A Large-scale Evaluation of a Model for the Evaluation of Games for Teaching Software Engineering. In: IEEE/ACM INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING: SOFTWARE ENGINEERING EDUCATION AND TRAINING TRACK, 39., 2017. Buenos Aires, Argentina. **Proceedings…**. Buenos Aires: IEEE/ACM, 2017.

PETRI, G.; GRESSE VON WANGENHEIM, C.; BORGATTO, A. F. Design and Evaluation of a Model for the Evaluation of Games for Computing Education. **Journal of Informatics in Education**, 2018b. (under review).

PETRI, G.; GRESSE VON WANGENHEIM, C.; BORGATTO, A. F. **MEEGA+, Systematic Model to Evaluate Educational Games.** In Newton Lee (Eds) Encyclopedia of Computer Graphics and Games, (pp. 1-7). Cham: Springer, 2018a.

PETRI, G.; GRESSE VON WANGENHEIM, C.; BORGATTO, A. F. **MEEGA+: A Method for the Evaluation of Educational Games for Computing Education**. Technical Report. INCoD/GQS.05.2018.E. July, 2018c.

PETRI, G.; GRESSE VON WANGENHEIM, C.; BORGATTO, A. F. **A Large-scale Evaluation of a Model for the Evaluation of Educational Games.** Technical Report. INCoD/GQS.04.2016.E, July, 2016.

PFAHL, D.; RUHE, G.; KOVAL, N. **An Experiment for Evaluating the Effectiveness of Using a System Dynamics Simulation Model in Software Project Management Education**. In: INTERNATIONAL SYMPOSIUM ON SOFTWARE METRICS, 7., 2001. London. **Proceedings...**. London: IEEE, 2001. p. 97-109.

PMDOME. (2017). **PMDome Workshop**. Available at: <https://ricardo-vargas.com/pt/workshops/pmdome/> Access: 27 November 2017.

POELS, K.; KORT, Y. D.; IJSSELSTEIJN, W. **It is always a lot of fun!: exploring imensions of digital game experience using focus group methodology**. In: CONFERENCE ON FUTURE PLAY, 7., 2007. Toronto. **Proceedings…**. New York: ACM, 2007. p. 83-89.

POZO, J. I. **Aprendizes e mestres:** a nova cultura da aprendizagem. Porto Alegre: Penso, 2002.

PPGCC. **Programa de Pós-Graduação em Ciência da Computação. Linhas de Pesquisa**. Disponível em: < http://ppgcc.posgrad.ufsc.br/linhas-de-pesquisa-2/>. Acesso em: 19 Abr. 2017.

PRENSKY, M. **Digital Game-Based Learning**. New York: Paragon House, 2007.

RAICHE, G.; WALL, T. A.; MAGIS, D.; RIOPEL, M.; BLAIS, J.-G. Nongraphical solutions for the Cattell's scree test. **Methodology**, v. 9, n. 1, p. 1-12, 2013.

RAMAN, R., LAL, A.; ACHUTHAN, K. Serious games based approach to cyber security concept learning: Indian context. In: INTERNATIONAL CONFERENCE ON GREEN COMPUTING COMMUNICATION AND ELECTRICAL ENGINEERING, 2014, Coimbatore. **Proceedings…**. Coimbatore, Indian, 2014. p. 1-5.

RATIER, R. **Como aprendem nossos alunos**. Revista Nova Escola. 2010. Available at: <https://novaescola.org.br/conteudo/1940/teorias-da-aprendizagem> Acess: 15 aug. 2018.

RITTERFELD, U.; CODY, M.; VORDERER, P. (Eds.). **Serious Games**. New York: Routledge, 2010.

RITTGEN, P. Quality And Perceived Usefulness of Process Models. In: the ACM SYMPOSIUM ON APPLIED COMPUTING. Nova York, EUA. **Proceedings…**. Nova York: ACM, 2010, p. 65-72.

ROSS, J. A. The reliability, validity, and utility of self-assessment. **Practical Assessment, Research & Evaluation**, v. 11, n. 10, p. 1-13, 2006.

ROSSIOU, E.; PAPADAKIS, S. Educational Games in Higher Education: a case study in teaching recursive algorithms. In: INTERNATIONAL CONFERENCE EDUCATION IN A CHANGING ENVIRONMENT, 2007, Salford. **Proceedings…**. Salford, UK, 2007. p. 149-157.

RUSU, A., RUSSELL, R., BURNS, E. & FABIAN, A. Employing software maintenance techniques via a tower-defense serious computer game. In: INTERNATIONAL CONFERECE ON E-LEARNING AND GAMES, EDUTAINMENT TECHNOLOGIES, 6., 2011, Taipei. **Proceedings…**. Taipei, Taiwan: Springer-Verlag, Berlin, Heidelberg, 2011. p. 176-184.

SALLA, F. **Neurociência:** como ela ajuda a entender a aprendizagem. Revista Nova Escola. 2012. Available at: <https://novaescola.org.br/conteudo/217/neurociencia-aprendizagem> Acess: 15 aug. 2018.

SAMEJIMA, F. **Estimation of latent ability using a response pattern of graded scores**. Psychometric Monograph. Educational Testing Service. New Jersey: Princeton, 1969.

SANTOS, A. L.; SOUZA, M. R. A.; FIGUEIREDO, E.; DAYRELL, M. Game Elements for Learning Programming: A Mapping Study. In: INTERNATIONAL CONFERECENCE ON COMPUTER

SUPPORTED EDUCATION, 10., 2018. Funchal. **Proceedings…**. Funchal, Portugal, 2018. p. 89-101.

SASKATCHEWAN EDUCATION. **Instructional Approaches:** A Framework for Professional Practice**.** Canada: Saskatchewan Education, 1991.

SAUNDERS, M. N. K.; LEWIS P.; THORNHILL, A. **Research Methods for Business Students**. (5th ed.). Harlow: Prentice Hall, 2009.

SAVI, R. Avaliações de jogos voltados para a disseminação de conhecimentos. PhD Thesis (Doutorado em Engenharia e Gestão do Conhecimento). Universidade Federal de Santa Catarina, Florianópolis, Brazil, 2011.

SAVI, R.; GRESSE VON WANGENHEIM, C.; BORGATTO, A. F. A Model for the Evaluation of Educational Games for Teaching Software Engineering. In: SIMPÓSIO BRASILEIRO DE ENGENHARIA DE SOFTWARE, 25., 2011. São Paulo. **Proceedings...** São Paulo: [S.I.], 2011. p. 194–203.

SAVI, R.; GRESSE VON WANGENHEIM, C.; ULBRICHT, V. R.; VANZIN, T. Proposal of a model for evaluation of educational games. **Revista Novas Tecnologias na Educação**, v. 8, n. 3, p. 1-10, 2010.

SBC. **Currículo de Referência da SBC para Cursos de Graduação em Bacharelado em Ciência da Computação e Engenharia de Computação.** Sociedade Brasileira de Computação. 2005. Disponível em: <http://www.sbc.org.br/documentos-da-sbc/summary/131-curriculos-de-referencia/760-curriculo-de-referencia-cc-ec-versao2005>. Acesso em: 11 abril 2017.

SCAGNOLATO, L. A. S. **A construção da afetividade para Jean Piaget.** 2015. Available at: <http://www.webartigos.com/articles/16153/1/a-construcao-da-afetividade-par-jean-piaget/pagina1.html>. Acess in: 30 mar. 2015.

SCHANZENBACH, D. W. Limitations of Experiments in Education Research. **Education Finance and Policy**, v. 7, n. 2, p. 219-232, 2012.

SCHIFFLER, A. **A heuristic taxonomy of computer games.** Ferzkopp's Philosophy Work, 1–13, 2006.

SCHILDGEN, J.; DEßLOCH, S. SQL-Grundlagen spielend lernen mit dem Text-Adventure SQL Island. In: GI-FACHTAGUNG DATENBANKSYSTEME FÜR BUSINESS, TECHNOLOGIE UND WEB, 16., 2015, Hamburg. **Proceedings…**. Hamburg, Germany, 2015. p. 1-4.

SCHNEIDER, M. F.: SCRUM'ed: **Um jogo de RPG para ensinar Scrum.** Undergraduate thesis. Federal University of Santa Catarina, Florianópolis, SC, Brazil, 2015.

SHADISH, W. R.; COOK, T. D.; CAMPBELL, D. T. **Experimental and quasi-experimental designs for generalized causal inference.** New York: Houghton Mifflin Company, 2002.

SHARMA, R.; JAIN, A.; GUPTA, N.; GARG, S.; BATTA, M.; DHIR, S. K. Impact of self-assessment by students on their learning. **International Journal of Applied and Basic Medical Research**, v. 6, n. 3, p. 226–229, 2016.

SHENG, S.; MAGNIEN, B.; KUMARAGURU, P.; ACQUISTI, A.; CRANOR, L.; HONG, J.; NUNGE, E. Anti-Phishing Phil: The Design and Evaluation of a Game that Teaches People Not to Fall for Phish. In: SYMPOSIUM ON USABLE PRIVACY AND SECURITY, 2007, Pittsburgh. **Proceedings…**. Pittsburgh, USA, 2007. p. 88-99.

SIGMAN, M. et al. Neuroscience and education: prime time to build the bridge. **Nature neuroscience**, v. 17, n. 4, p. 497–502, 2014.

SILVA, J. P.; CARVALHO, F. C.; SILVESTRE, A.; LUIZ, M. F.; ALENCAR, R. L.; SILVEIRA, I. F. Forca dos Sinônimos: Um Jogo da Forca Multiplayer para o ensino de Tesauros em Língua Portuguesa como Ferramenta de busca na Web. In: Brazilian Symposium on Games & Digital Entertainment, 16., 2017, Curitiba. **Proceedings…**. Curitiba, Brazil, 2017a. p. 1128-1131.

SILVA, J. P.; JORGE, A. A.; GAIA, C. C.; COSTA, G. C. Quais as chances? Um jogo de dados e cartas para o ensino do cálculo de probabilidades. In: CONGRESSO DE INOVAÇÃO, CIÊNCIA E

TECNOLOGIA DO IFSP, 8. 2017, Cubatão. **Proceedings...**. Cubatão/SP, Brasil, 2017b. p. 1-3.

SIMPSON, E. J. **The classification of educational objectives, psychomotor domain.** Washington: Gryphon House, 1972.

SINDRE, G.; MOODY, D. **Evaluating the Effectiveness of Learning Interventions: an Information Systems Case Study**. In: EUROPEAN CONFERENCE ON INFORMATION SYSTEMS, 11., 2003. Naples. **Proceedings…**. Naples: AISeL, 2003.

SINDRE, G.; NATVIG, L.; JAHRE, M. Experimental validation of the learning effect for a pedagogical game on computer fundamentals. **IEEE Transactions on Education**, v. 52, n. 1, p. 10–18, 2009.

SINGH, J.; DORAIRAJ, S. K.; WOODS, P. **Learning Computer Programming Using a Board Game - Case Study On C-Jump**. Masters thesis, Multimedia University, Cyberjaya, Malaysia, 2007.

SITZMANN, T.; ELY, K.; BROWN, K. G.; BAUER, K. N. Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure? **Academy of Management Learning & Education**, v. 9, n. 2, p. 169-191, 2010.

SOARES, R; PETRI, G.; GRESSE VON WANGENHEIM, C.; CONTE, T.; MARQUES, A. B. Assistant MEEGA+: Uma ferramenta de apoio para avaliação de jogos educacionais usando modelo MEEGA+. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 29., 2018. Fortaleza. **Proceedings...**. Fortaleza, Brazil, 2018.

SPRINTHAL, N. A.; SPRINTHAL, R. **Psicologia Educacional.** Rio de Janeiro: McGraw Hill, 1993.

SWEETSER, P.; WYETH, P. GameFlow: a model for evaluating player enjoyment in games. **Computers in Entertainment**, v. 3, n. 3, p. 1-24, 2005.

TABACHNICK, B. G.; FIDELL, L. S. **Using Multivariate Statistics**. 5th ed. Boston: Allyn and Bacon, 2007.

TAHIR, R.; WANGMAR, A. I. State of the art in Game Based Learning: Dimensions for Evaluating Educational Games. In: EUROPEAN CONFERENCE ON GAMES BASED LEARNING, 2017, Graz. **Proceedings…**. Graz, Astria, 2017. p. 641-650.

TAKATALO, J.; HÄKKINEN, J.; KAISTINEN, J.; NYMAN, G. **Presence, Involvement, and Flow in Digital Games**. In: Bernhaupt, R. (Ed.). Evaluating User Experience in Games: Concepts and Methods, Springer, 2010. p. 23-46.

TEZZA, R.; BORNIA, A. C.; ANDRADE, D. F. D. Measuring web usability using item response theory: principles, features and opportunities. **Interacting with Computers**, v. 23, n. 2, p. 167-175, 2011.

THIRY, M.; ZOUCAS, A.; SILVA, A. C. Empirical study upon software testing learning with support from educational game. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING & KNOWLEDGE ENGINEERING, 23., 2011, Miami Beach. **Proceedings…**. Miami Beach, USA: IEEE, 2011. p. 482-484.

THOMAS, G.; MARTIN, D.; PLEASANTS, K. Using self- and peer-assessment to enhance students' future-learning in higher education. **Journal of University Teaching & Learning Practice**, v. 8, n. 1, p. 1-17, 2011.

TONUSSI, L. P.; HAUCK, J. C. R. Um Módulo de Jogo de Perguntas e Respostas para apoio ao Ensino de Gerência de Projetos Integrado ao Moodle. In: COMPUTER ON THE BEACH, 2017, Florianópolis. **Proceedings….** Florianópilis, Brazil, 2017. p. 100-109.

TROCHIM, W. M.; DONNELLY, J. P. **Research methods knowledge base.** 3. ed. Mason, OH: Atomic Dog Publishing, 2008.

TULLIS, T.; ALBERT, W. **Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics**, 2008. Morgan Kaufmann. University, Pittsburgh, 2005.

WAGNER, R.W. Edgar Dale: Professional Theory into Practice**. Taylor & Francis, Ltd.**, v. 9, n. 2, p.89-95, 1970.

WESKE, M. **Business Process Management: Concepts, Languages, Architectures.** 2nd ed. Hidelberg: Springer, 2012.

WIEBE, E. N.; LAMB, A.; HARDY, M.; SHAREK, D. Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. **Computers in Human Behavior**, 32, 123-132, 2014.

WOHLIN, C.; RUNESON, P.; HOST, M.; OHLSSON, M. C.; REGNELL, B.; WESSLEN, A. **Experimentation in Software Engineering**. Hidelberg: Springer, 2012.

WOLF, J. **The medium of the video game.** Austin: University of Texas Press, p. 113-134, 2001.

YIN, R. K. **Case study research: design and methods**. (6th ed.) Thousand Oaks: SAGE Publications, 2017.

ZAIBON, S. B. User Testing on Game Usability, Mobility, Playability, and Learning Content of Mobile Game-Based Learning. **Journal Teknologi**, v. 77, n. 29, p. 131-139, 2015.

ZAIBON, S. B.; SHIRATUDDIN, N. Heuristics Evaluation Strategy for Mobile Game-Based Learning. In: IEEE INTERNATIONAL CONFERENCE ON WIRELESS, MOBILE AND UBIQUITOUS TECHNOLOGIES IN EDUCATION, 6., 2010, Kaohsiung. **Proceedings…**. Kaohsiung, Taiwan: IEEE, 2010. p. 127-131

APPENDIX A – Detailed description of the State of the Art

Data extracted to answer the questions AQS1 and AQS2

AQS1: Which models, methods, scales, or frameworks (approaches) exist to systematically evaluate educational games?				**AQSQ2:** Which quality and/or sub-quality factors are evaluated?		
Id	**Reference**	**Name**	**Instructional Strategy**	**Quality factors**	**Quality subfactors**	**Theoretical basis**
1	(SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011)	MEEGA (Model for the Evaluation of Educational Games)	Educational games	Motivation User Experience Learning	Motivation: attention, relevance, confidence, satisfaction. User experience: fun, competence, challenge, social interaction, immersion. Learning	ARCS Model [Keller, 87] [Sweetser, 05; Poels07; Gámez, 09; Takatalo, 10] [Bloom, 56; Sindre, 03]
2	(FU; SU; YU, 2009)	EGameFlow	E-learning games	Concentration Goal clarity Feedback Challenge Control Immersion Social Interaction Knowledge improvement	Not Informed	[Sweetser, 05]
3	(CARVALHO, 2012)	Not defined	Game-based learning	Beta testing: - Game play - Game story - Mechanisms/Usability	Not Informed	Not Informed
		Not defined	Game-based learning	Gamma testing: - Knowledge	Knowledge	Not Informed

				- Motivation - Satisfaction	Motivation: competence, interest, motivation for Computer Games Satisfaction: interest/enjoyment, perceived competence, user experience.	
4	(AK, 2012)	Not defined	Educational computer games	Enjoyment Learning	Enjoyment: challenge, curiosity & mystery, clear goals, social interaction, diversion (fun), fantasy, arousal, flow. Learning	[Sherry, 06; Fu, 09; Garris, 02; Berlyne, 60; Kiili, 05; Csikszentmihalyi, 91; Freitas, 06; Squires, 99]
5	(MAYER, 2012; MAYER et al., 2013; MAYER et al., 2014)	Not defined	Serious games	Game performance Game play Game experience Player satisfaction Learning	Game performance: time, avoidable mistakes Game play: dominance, influence, power Game experience: flow, immersion, presence Post-game: Game experience: engagement, fun Player satisfaction: clarity, relevance, attractiveness, ease of use, interaction with others students (social interaction), student's efforts, motivation, role identification, facilitator Learning: player learning satisfaction, self-reported, self-perceived learning, measured changes in knowledge, attitudes, skills, behaviors, asking clients, participants. Measured changes in team: safety, commitment, performance.	[Tallir, 07; Oslin, 98; Baba, 93; Trepte, 11; Blumberg, 00; Csikszentmihalyi, 91; Admiraal, 11; Martin, 08; Mayes, 11; Boyle, 12; Schuurink, 08; Olsen, 11; Reichlin, 11]

6	(CONNOLLY; STANSFIELD; HAINEY, 2008; CONNOLLY; STANSFIELD; HAINEY, 2009; HAINEY; CONNOLLY; BOYLE, 2010)	Evaluation Framework for Effective Games-based Learning	Game-based learning	Learner performance/Learning Learner/academic motivation Learner/academic perceptions Learner/academic preferences GBL environment Collaboration between players where appropriate	Learner performance Learner/instructor motivation Learner/instructor perceptions Learner/instructor preferences Learner/instructor attitudes GBL environment: virtual environment, scaffolding, usability, level of social presence, deployment. Collaboration	Not clearly informed
7	(FREITAS; OLIVER, 2006)	Four-Dimensional Framework	Games- and Simulation-based learning	Pedagogic considerations Learner specification Context Mode of representation	Pedagogic considerations: learning models used, approaches taken. Learner specification: learner profile, pathways, learning background, group profile. Context: classroom-based, outdoors, access to equipment, technical support Mode of representation: level of fidelity, interactivity, immersion.	Not Informed
8	(CHEW, 2017)	Not defined	Game-based learning	Behavior Cognitive Emotional Agentic Immersion Challenge	Not Informed	J. Reeve and C. Tseng, ”Agency as a fourth aspect of students engagement during learning activities”, Contemporary Educational Psychology, vol. 36, no. 4, pp. 257-267, 2011

9	(ABDELLATIF; MCCOLLUM; MCMULLAN, 2018)	Not defined	Serious games	Usability Understandability Motivation Engagement User experience	Usability: learnability, usefulness, errors, ease of use. Understandability: clarity, need for supervision. Motivation: challenge, enjoyment, curiosity. Engagement: purpose, interest, control. User experience: competence, social interaction, fun.	Not Informed
10	(GARCIA-MUNDO; GENERO; PIATTINI, 2015)	QSGame-Model	Serious games	Functional suitability Performance efficiency Compatibility Usability Reliability Security Maintainability Portability		Product Quality Model of ISO/IEC 25010

Data extracted to answer the questions AQS3, AQS4 and AQS5

AQS3: How data collection and analysis is operationalized?					**AQS4:** How these approaches have been developed?	**AQS5:** How these approaches have been evaluated?			
Id	**Study type**	**Data collection instrument(s)**	**Response format**	**Data analysis method(s)**	**Development methodology**	**Evaluated factors**	**Number of applications**	**Data points**	**Data analysis method(s)**
1	Non-experimental with case study: one-	Questionnaire	Likert scale Ordinal scale	Descriptive statistics: histogram, frequency	GQM (BASILI et al., 1994)	Validity Reliability Applicability	Applied in two courses with 3 games in	79	Intercorrelation of scale items Item-total correlation Variance Mean

	shot post-test only			diagram, median.	Scale Development (DEVELLIS, 2003)		each application		Cronbach's alpha coefficient
2	Ad-hoc evaluation: pre-test/post-test	Questionnaire	Likert scale	Descriptive statistics: mean, SD, Pearson correlation coefficient Hypothesis testing: ANOVA, T-test Qualitative analysis	Scale Development (DEVELLIS, 2003)	Item analysis Validity Reliability	One application in one course with 4 e-learning games	166	Mean Standard deviation Extreme group comparison Test for homogeneity T-Test ANOVA Pearson's correlation Cronbach's alpha coefficient
3	Ad-hoc evaluation: one-shot post-test only	Questionnaire Semi-structured interview	Likert scale	Not informed	Not informed	Not informed	Not informed	Not informed	Not informed
	Ad-hoc evaluation: pre-test/post-test	Questionnaire Tests	Likert scale Ordinal scale	Not informed	Not informed	Not informed	Applied in two classes with only one game	Not informed	Not informed
4	Not Informed	Questionnaire	Not Informed	Not informed	Not informed	Not informed	Not Informed	Not Informed	Not informed
5	Quasi-experimental: pre-test/post-test design	Not Informed	Not Informed	Not Informed	Not informed	Not informed	Several hundreds of sessions	2164	Not informed

6	Not Informed	Not Informed	Not Informed	Not Informed	Not informed	Not informed	Evaluated through 2 studies	Not informed	Not informed
7	Not Informed	Not Informed	Not Informed	Not Informed	Not informed	Not informed	Applied to evaluate two games	Not informed	Not informed
8	Ad-hoc evaluation: one-shot post-test only	Questionnaire	Likert scale	Mean	Not Informed	Not Informed	Applied to evaluate 3 games	37	Not informed
9	Ad-hoc evaluation: one-shot post-test only	Questionnaire	Ordinal scale	Mean	Not Informed	Not Informed	One application	15	Not informed
10	Not Informed	Not Informed	Not Informed	Not Informed	Top-down approach methodology (FRACH; CARVALLO, 2003)	Not Informed	Not informed	Not informed	Not informed

APPENDIX B – Detailed description of the State of the Practice

Overview on the evaluated games

ID	Game	Game Description	Computing Knowledge Area	References
1	Wu's Castle	Wu's Castle is a game where students program changes in loops and arrays in an interactive, visual way. Players interact in two ways: by manipulating arrays by changing loop parameters, and by physically walking the game character through loop execution.	Software Development Fundamentals	(Eagle & Barnes, 2009)
2	X-MED	The objective of this game is to reinforce the remembering and understanding of basic concepts and to train measurement application. In this game, the learner takes the role of a measurement consultant in a software organization. During the game session, the player passes sequentially through all steps of a GQM-based measurement program creating measurement artifacts based on a series of constrained selections with pre-defined alternatives. For each decision, the game gives a feedback and score. At the end a total score and feedback report is provided.	Software Engineering	(Gresse von Wangenheim, Thiry & Kochanski, 2009) (Gresse von Wangenheim, Thiry, Kochanski, Steil & Lino, 2009)
3	SimSE	SimSE is a single-player game in which the player takes on the role of project manager of a team of developers who must successfully complete a "virtual" software engineering project.	Software Engineering	(Navarro & van der Hoek, 2009) (Navarro, 2006) (Navarro & van der Hoek, 2005) (Navarro & van der Hoek, 2007) (Navarro & van der Hoek, 2008)

4	EleMental: The recurrence	A game that provides computer science students the opportunity to write code and perform interactive visualizations to learn about recursion through depth-first search of a binary tree.	Software Development Fundamentals	(Chaffin, Doran, Hicks, & Barnes, 2009).
5	DELIVER!	DELIVER! is a board game to teach Earned Value Management in monitoring and controlling the execution of a software project. The purpose of the game is to reinforce EVM concepts and to exercise the application of EVM.	Software Engineering	(Gresse von Wangenheim, Savi & Borgatto, 2012)
6	Saving Sera	"Saving Sera" is a 2D exploratory game, where the player learns of the kidnapping of the princess and determines to rescue her. The user must perform various tasks involving programming concepts: correctly reordering a while loop statement of a confused old fisherman's mind; correcting nested for loop placing eggs in crates; and visually piecing together a quicksort algorithm. When the player makes a mistake, the character must fight a script bug, which asks the users various computer science questions in order to fight the bug.	Software Development Fundamentals	(Barnes, Powell, Chaffin, & Lipford, 2008)
7	The Catacombs	"The Catacombs" is a 3D game where the user is an apprentice wizard who must perform three progressively more complicated tasks to save children stuck in the catacombs. The first involves unlocking the door to the catacombs involving two if statements; the second, building a bridge brick by brick with nested for loops; and the third, solving a cryptogram using more nested for loops. In the second and third quests, incorrect answers resulted in decreasing player health. The game uses dialogue with a sarcastic spell book named Grimore and multiple-choice questions (a dialogue tree) to create the code and complete the tasks.	Software Development Fundamentals	(Barnes, Powell, Chaffin, & Lipford, 2008)

8	The Risk Management Game	In the game, each player assumes the role of a project manager and the players compete against one another. The game's objective from a player's perspective is to develop a product, sell it in the market and win by having more money at the end of the game than all the other players. The game has 5 stages: planning, requirements, architecture & design, implementation and testing; yet unlike traditional board games, players are free to roam around, choose where to go and what to do, without the need for throwing a die to advance. Though players can follow a natural waterfall like cycle between the phases, they are not required to do so.	Software Engineering	(Taran, 2007)
9	Program your robot	Program your robot is a serious game designed to enable students to practice working with introductory programming constructs, within an environment that explicitly supports the acquisition of CT skills (such as algorithm building, debugging and simulation). The goal of the game is to assist a robot and help him to escape from a series of platforms by constructing an escape plan called a solution algorithm. Players construct their solution algorithm by giving various commands to the robot to perform.	Software Development Fundamentals	(Kazimoglu, Kiernan, Bacon, & Mackinnon, 2012a) (Kazimoglu, Kiernan, Bacon, & Mackinnon, 2012b)
10	SortingGame	SortingGame is a card games about sorting algorithms and related concepts. The basic concept is that the players have to match an algorithm and criteria that are valid for the algorithm. A criterion can be, for example, stable, which means that all the algorithms that are stable can be matched to that criterion. In SortingGame, players first place down some criteria and then they have to place down an algorithm that matches the criteria. The player who placed down an algorithm with the best time complexity, wins the round.	Algorithms and Complexity	(Hakulinen, 2011)

11	SortingCasino	SortingCasino is a card games about sorting algorithms and related concepts. The basic concept is that the players have to match an algorithm and criteria that are valid for the algorithm. A criterion can be, for example, stable, which means that all the algorithms that are stable can be matched to that criterion. In SortingCasino, players collect cards from the table by matching corresponding algorithm with a valid criterion.	Algorithms and Complexity	(Hakulinen, 2011)
12	Secret rule	Secret rule covers all the topics of the course. The deck consists of cards that have all the data structures and algorithms taught in the course. The main idea is that the dealer makes up a secret rule that the other players try to figure out. A secret rule can be for example: "A linear data structure". On each turn, a player puts one card on the table trying to match the secret rule, and the dealer acts as a judge. The advantage of this game is that the difficulty level is purely defined by the secret rule, which is made by the students.	Algorithms and Complexity	(Hakulinen, 2011)
13	Draw and guess	In Draw and guess, students are divided into two groups. The deck of cards consists of course concepts. On each turn, a student has to draw the concept to the blackboard. If his or her own team guesses the correct concept, they receive a point. The opposite team can also guess the concept and steal the point.	Algorithms and Complexity	(Hakulinen, 2011)
14	SCRUMIA	SCRUMIA is a manual paper & pencil game to teach SCRUM for managing software projects. The purpose of the game is to strengthen the understanding of SCRUM concepts and to exercise the application of the SCRUM process. After playing the game, students are expected to be more able to remember the names of concepts related to SCRUM (remembering) as well as to be more able to explain and distinguish SCRUM roles, meetings and	Software Engineering	(Gresse von Wangenheim, Savi & Borgatto, 2013)

		artefacts (understanding). Yet, the principal objective is that students should be able to execute the SCRUM process creating the respective artefacts to manage software projects in concrete situations (applying).		
15	Software engineering education game in Second Life	Based on the SimSE, in the game, six software engineer roles are provided. The players can choose any role they want but can only bein charge of the chosen role. All the players joining the game will form a software development team. During the game, the player can interact with each other through various communication means provided in SL. A team score will be given at the end of game if the team can deliver the product before the deadline. To get the good score, all the players need to not only work on their own part, but also collaborate with each other.	Software Engineering	(Wang & Zhu, 2009)
16	BOTS	An online, multiplayer game that allows players to engage in simple programming and problem-solving puzzles in an engaging visual environment. This environment, sharing elements of successful social games and websites like Farmville, Fantastic Contraption, and the online Scratch community, will encourage consistent social interaction and sharing of knowledge.	Software Development Fundamentals	(Hicks, 2010)
17	SpITKom Project	SpITKom aims at utilizing the pedagogical potential computer games provide by offering a learning scenario based on a Browser Game. The game guides the learner through building- and construction projects. Its main intention is to bring the target group (learners difficult to reach) “in touch” with the integrated IT-knowledge. A more elaborate engagement, i.e. the actual learning, takes place within the IT-Café.	IT Knowledge	(Schmitz, Czauderna, Klemke, & Specht, 2011)

18	Prog&Play	Prog&Play is based on an open source real-time strategy (RTS) game called Kernel Panic [Ker]. Kernel Panic uses computer science metaphors, such as bits and pointers, as units (i.e. graphical objects which are controlled by the player). It is a simplified RTS with the following features: there is no resource management except for time and space; all units are free to create; it has a small technology improvement tree with fewer than 10 units; and it uses low-end vectorial graphics which match the universe. Owing to these characteristics, differences between two players are about strategies and tactics used (and not about knowledge of units features and relative advantages). Thus, the game is action-oriented while always remaining user friendly. Starting with Kernel Panic, we designed an applicative programming interface that enables students to interact with the game through programming.	Software Development Fundamentals	(Muratet, Torguet, Viallet, & Jessel, 2011) (Muratet, Delozanne, Torguet, & Viallet, 2012).
19	Light-Bot	Light-Bot is an educational Flash game developed by Armor Games. The objective of the game is to program a small robot to light up all the blue blocks on a board. This objective is achieved by giving the robot a series of instructions from a limited set of commands, with a note instruction space.	Software Development Fundamentals	(Gouws, Bradshaw, & Wentworth, 2013)
20	Software Development Manager (SDM)	In the SDM game, the player has a team of employees who are used to develop software that are required by customers. The gameplay and game mechanics are modeled presenting possibilities to the player to decide strategies for development and define the roles for each staff member. The software required by customers may have requirements that must be respected during development.	Software Engineering	(Kohwalter, Clua, & Murta, 2011)

21	PM Master	A board game with questions and answers about project management, where players alternately answer questions from a category (eg, scope, time, cost, human resources, etc.). Wins the player who correctly answer a question from each category. The educational goal of the game is to strengthen and fix concepts of process groups and project management knowledge areas.	Software Engineering	(Savi, Gresse von Wangenheim & Borgatto, 2011)
22	Requirements Collection and Analysis Game (RCAG)	The basic idea of the game is for the team (comprising one or more players) to manage and deliver a number of software development projects. Each player has a specific role, such as project manager, systems analyst, systems designer or team leader. Each scenario has an underlying business model. Additional resources can be brought in for a project although this will have a cost and timescale (delay) associated with it. The project manager has overall responsibility for the delivery of each project on budget and on time and is given a short brief for each project. Communication is one of the key aspects of the game and the project manager must communicate relevant details of the project to the other players.	Software Engineering	(Hainey, Connolly, & Boyle, 2009) (Hainey, Connolly, Stansfield, & Boyle, 2011)
23	Innov8	The game employs a first-person role-playing approach where the player assumes the role of a consultant within a company that is experiencing challenges. The player is given a goal which he/she must achieve in order to successfully complete the game. The goal is to re-engineer a call center process for the company in order to make it more efficient and effective. In order to achieve this goal, the player has to complete certain tasks which evolve as the game progresses.	Information Systems	(Roodt & Joubert, 2009)
24	Kernel Panic	This game is designed to allow the user to easily program and integrate his/her AIs. It is aimed at users who already know how to program. In this game, there are three races:	Software Development Fundamentals	(Muratet, Torguet, Jessel, & Viallet, 2009)

		The System, the Network and the Hacker, all having their own advantages and disadvantages. The game takes place inside of a computer, leading to intense, fast paced gameplay. There are no resources in Kernel Panic, so the player can build units until the map is full. The game's textures and sounds resemble old games, and maps are like computer boards and chips.		
25	Immersive Security Education Environment (I-SEE)	The learning module simulates a scenario in which two teams are competing against each other: one acting as a group of attackers (a.k.a. red team) and the other acting as a group of defenders (a.k.a. blue team). Each team has a base consisting of a packet assembly board, a building containing a money counter and a console, and a pipeline representing a network connection from the base to the gateway router.	Information Assurance and Security	(Ryoo, Techatassanasoontorn, Lee, & Lothian, 2011)
26	hACMEgame	The game is organized as a series of levels where the player must overcome a set of challenges in order to unlock access to the next level. The player begins at the lowest level (level 1) and proceeds through increasingly numbered levels with gradually increasing difficulty. Each level contains a set of challenges, which the player must complete in order to advance to the next level.	Information Assurance and Security	(Nerbraten & Rostad, 2009)
27	Evaluators	Evaluators is an educational system that involves both instructors and students. The system provides instructors with different tools: an authoring tool for creating exercises concerning processing tasks formalized as attribute grammars, a customizer tool for generating the serious games from the exercises, and an analytic tool to evaluate student performance on these games.	Programming Languages	(Rodriguez-Cerezo, Sarasa-Cabezuelo, Gomez-Albarran, & Sierra, 2014) (Rodriguez-Cerezo, Sarasa-Cabezuelo, Gomez-Albarran & Sierra, 2012)
28	Simsoft	Simsoft players are formed into teams of 2 or more and they are given a scenario that describes the requirements	Software Engineering	(Bavota, Lucia, Fasano, Oliveto, & Zottoli, 2012)

		for a small software development project. Taking the role of project manager, the team must manage the project from start-up to final delivery. The players gather around a printed game board to discuss the current state of the project and to decide on their next move. The board shows the flow of the game while plastic counters are used to represent the staff of the project.		(Caulfield, Veal, & Maj, 2011)
29	The Requirements Island	The game is contextualized in a playful environment, which the goal is to entertain the players, through its story and its characters. During the game, the player will need to carry out a bunch of challenges in order to evaluate the player knowledge about the Requirements Engineering.	Software Engineering	(Thiry & Gonçalvez, 2010)
30	Motherboard-assembly pairing game	The motherboard-assembly pairing game is designed both to lead students to remember computer components by looking at realistic photographs of the equipment and to simulate them to assemble a motherboard in an interactive venue	Architecture and Organization	(Fu, Su, & Yu, 2009)
31	Describing computer parts	The game describing computer parts was developed with the purpose of guiding the student to understand how to purchase computer accessories based on his or her personal needs	Architecture and Organization	(Fu, Su, & Yu, 2009)
32	Hands-on OS game	The purpose of the hands-on OS game is to acquaint the student with common problems associated with a computer's operating system	Operating Systems	(Fu, Su, & Yu, 2009)
33	Bear-cub's computer game	The bear-cub's computer game introduces the player to a wide range of computer software and imparts basic knowledge regarding the software.	Architecture and Organization	(Fu, Su, & Yu, 2009)
34	Maze	The setting of the level is a technologically advanced three-dimensional maze. The maze has three floors; higher floors are reached with three consecutive correct answers. Each floor has a different texture and look. To	Software Development Fundamentals	(Adamo-Villani, Haley-Hermiz, & Cutler, 2013)

		win the game the player must get to the third floor and successfully answer three problems in a row. All problems require the application of various operator precedence rules.		
35	Coloring Map	The goal of the game is to help students to understand CSP (Constraint Satisfaction Problem) solving. The map-coloring game is a one-player game played coloring the map, using a set of available colors so that the adjacent regions the map do not have the same color.	Intelligent Systems	(Hatzilygeroudis, Grivokostopoulou, & Perikos, 2012)
36	Scrum LEGO Challenge	Players execute sprints building LEGO houses and vehicles from user stories following the SCRUM process and realizing SCRUM ceremonies.	Software Engineering	(Paasivaara, Heikkilä, Lassenius, & Toivola, 2014)
37	Star Chef	The game included the following five learning units: stack game, queue game, bubble sort game, tree traversal game, and binary search game. For example, unit 1 is Fry meat balls game, where learners must follow the rule of last in, first out (LIFO) to deliver fried meat balls (data list) to each customer.	Software Development Fundamentals	(Liu, Chu, & Tan, 2012)
38	Object Access Tool (OAT)	In this game, we are teaching basic principles, such as class, objects and relationship between them in the beginning of object-oriented programming (OOP) course and postpone learning of specific objective language to later time	Software Development Fundamentals	(Livovský & Porubän, 2014)
39	DesigMPS	DesigMPS is a computer game designed to support the teaching of SPM (Software Process Modeling) by reinforcing relevant concepts and providing SPM exercises.	Software Engineering	(Chaves, Gresse von Wangenheim, Furtado, Oliveira, Santos & Favero, 2015)
40	Data Defender	The goals were to design and build a serious game prototype for use in helping to develop secure coding abilities to novice programmers and to develop high quality introductory computing laboratory exercises that	Information Assurance and Security	(Oania, 2011)

		incorporate game activities as part of the laboratory assignments.		
41	Ztech	Ztech de Object-Oriented is a stand-alone game in which each player will play and learn in their own application. The game will guide users in an appropriate way so that users can truly understand the concept of Object-Oriented. The purpose of having the gaming part is just to increase and foster the interest of users to learn the knowledge. Through the game, the application will provide users with all the basics of Object-Oriented concept like encapsulation, inheritance and polymorphism.	Software Development Fundamentals	(Seng & Yatim, 2014)
42	InspectorX	The game aims at training in the software inspection process activities and the learning of concepts and taxonomies of software inspection, including the inspection process itself, its management, the necessary roles, and how to inspect artifacts and classify them accordingly.	Software Engineering	(Potter, Schots, Duboc, & Werneck, 2014)
43	SPARSE	SPARSE is a game to practical approach for teaching and learning Software Engineering, based on games and simulation.	Software Engineering	(Souza, Resende, Prado, Fonseca, Carvalho, & Rodrigues, 2010)
44	SORTIA	SORTIA is a game of Heapsort sorting algorithms where the student must simulate manually and thus actively experiencing the process of sorting a set of integers.	Software Development Fundamentals	(Batistella, Wangenheim & Gresse von Wangenheim, 2012)
45	InspSoft	The InspSoft is an educational game designed for software inspection in a requirements specification document that provides a playful environment enabling the choice of an avatar for the player, one character for each learning objective and specific sounds for successes and failures in the game. The InspSoft simulates a software inspection situation in a software development company. The objective is to learn about the roles of each	Software Engineering	(Lopes, Marques, & Conte, 2012)

		participant in the inspection process and the types of defects found in a requirements document		
46	SimMS	SimMS is a digital educational game single-player whose goal is to support the teaching of Software Maintenance, with emphasis on the process described by the standard IEEE 1219 (IEEE 1998). The game simulates the environment of a software company, which implements maintenance requests requested by a customer in a specific software. The goal of the game is to meet the previously registered maintenance requests, according to the norm.	Software Engineering	(Simão, Correa, & Parreira Júnior, 2014)
47	UbiRE	The game aims to promote the understanding of ER in the context of a ubiquitous system in a smart home environment	Software Engineering	(Lima, Campos, Santos, & Werner, 2012)
48	CyberCIEGE	The CyberCIEGE game consists of a 3-dimensional office environment consist of some characters who access the assets to achieve goals. This game gives a virtual money concept based on this. The building blocks of CyberCIEGE consist of several elements: a unique simulation engine, a domain-specific scenario definition language, a scenario development tool, and a video enhanced encyclopedia.	Information Assurance and Security	(Raman, Lal, & Achuthan, 2014)
49	ProGames	The main objective of ProGames is facilitating learning and foster motivation into the programming area. To this purpose, we propose a comprehensive set of programming games that are, in fact, sets of exercises arranged into categories that every student selects and completes according to his particular tastes.	Software Development Fundamentals	(Hijon-Neira, Velazquez-Iturbide, Pizarro-Romero, & Carrico, 2014)
50	IT Billionare Game Board	IT Billionaire is a (turn based) board game intended for 2 through 4 players in which players attempt to become a billionaire by running a globally operating software	Software Engineering	(Solingen, Dullemond, & Gameren, 2011)

		engineering company. This game has been developed by the researchers themselves with the purpose to investigate if such games provide an effective learning tool to teach GSE dynamics to computer science students. The game takes place on a world map and every round of play represents a day.		
51	CAPTAIN3	CAPTAIN3 is a puzzle- based programming training system with motivational learning methods, in which learners create programs similar to solving a puzzle game.	Software Development Fundamentals	(Nunohiro, Matsushita, Mackin, & Ohshiro, 2013)
52	The Last Java Code	The Lost Java Code was developed by two undergraduate computer science students with GameMaker 8.1 during the summer of 2011. It aims to help novice programmers review the concepts on decision structures such as if, if-else, if-else-if and switch statements and practice it in a fun way.	Software Development Fundamentals	(Zhang, Smith, Caldwell, & Perkins, 2014)
53	Tower-Defense	The goal is to try to stop enemies from crossing the map by building towers which shoot at them as they pass. The enemies and towers usually have varied abilities and costs. When an enemy is defeated, the player earns money, which is used to buy more towers. In our game, the student is faced with ants, which are a metaphor for software bugs and lacking features for an existing software system.	Software Engineering	(Rusu, Russell, Burns, & Fabian, 2011)
54	The aMAZEing Labyrinth	The game consists of a 7x7 tiled area, where every tile is an I-shaped, T-shaped, or L-shaped corridor. Every other row and column can be shifted. There is an extra tile that is used to do so, thereby changing the paths in the labyrinth. Each player has a wizard with a distinct home position on the board, plus a randomly generated list of treasures to collect before returning home. Treasures are depicted on some of the tiles and the treasure is collected	Software Development Fundamentals	(Bezakova, Heliotis, & Strout, 2013)

		whenever a player's wizard lands on the corresponding tile.		
55	Quoridor	This game consists of a 9x9 board and works for 2 or 4 players. Each player places a single piece in the middle of one edge of the board. Its goal is to arrive somewhere along the edge at the opposite side of the board. Additionally, to their piece, each player has ten walls of length 2. In each move, a player either moves their piece to an adjacent location or places a wall. A wall must be aligned with the grid lines and it cannot cross an already placed wall or extend out of the board. Walls are used to obstruct and therefore detour opponents' movement. However, the wall placement cannot completely block the opponent from reaching their destination.	Software Development Fundamentals	(Bezakova, Heliotis, & Strout, 2013)
56	San Francisco Cable Cars	This game for 2-6 players is played on an 8x8 initially empty board, with 32 cable car stations distributed along the sides of the board. The rules specify which stations belong to each player. Each player gets a randomly drawn tile that species four connections, each between two of the eight end-points on the tile's sides. In every move, a player rotates their tile and places it on the board adjacently to an already placed tile or to a side of the board. By doing so, the players build cable car routes (theirs or their opponents'- or both). The game finishes after all tiles have been placed and the player with the overall longest cable car routes wins.	Software Development Fundamentals	(Bezakova, Heliotis, & Strout, 2013)
57	NeverWinter Nights 2	The game was developed to achieve the above learning outcomes in a 3D Role-Playing-Game (RPG) created through the Aurora toolset from NeverWinter Nights 2. In this game, learners control an avatar that they can personalize at the beginning of the game. The avatar can	Information Management	(Soflano, Connolly, & Hainey, 2015a) (Soflano, Connolly, & Hainey, 2015b)

		be controlled by point-and-click using a mouse and also by using keyboard shortcuts. The same method also controls the game's 3D camera. The game is designed to require the learner to retrieve the required data and evidence from the fictional database by using appropriate SELECT statements in order to get the warrant. The game structures learning by increasing the level of complexity of the SQL queries that the learner must construct as the game progresses. As the process of real-life crime investigation parallels to some extent the process of investigation reflected in the game story, this element may help the learner to understand how the learning content is applied in real life.		
58	Problems and Programmers	A card game that simulates the software process from requirements specification to product delivery based on the waterfall life cycle. Players take the role of the project leader in the same project and compete to be the first to complete the project. They pass through the phases of the software process and draw cards and take actions to continue the development as well as to react to problems. The winner of the game is the player who first achieves a sufficient number of integrated code cards without bugs.	Software Engineering	(Baker, Navarro, & van der Hoek, 2005) (Baker, Navarro, & van der Hoek, 2003) (Baker, Navarro, & van der Hoek, 2004)
59	SESAM	SESAM (Software Engineering Simulation by Animated Models) is a model of a software project. Users run the model loaded with its initial project state and then tweak it to simulate different scenarios before running it again. Players take the role of a project manager and must plan and control a simulated project. Rather than a graphical user interface, players control the game by typing commands in a modelling language. Players analyze their performance through an after-game analysis tool.	Software Engineering	(Drappa & Ludewig, 1999) (Drappa & Ludewig, 2000) (Mandl-Striegnitz, 2001)

60	Open Software Solution (OSS)	Case studies are presented through a simulated office environment and then completed outside of the game environment.	Software Engineering	(Sharp & Hall, 2000)
61	SimjavaSP	Players act as a project manager to deliver a product within time and budget constraints. SimjavaSP uses discrete-event simulation as the game engine.	Software Engineering	(Shaw & Dermoudy, 2005)
62	Incredible Manager	The Incredible Manager, the player sets project parameters such as staffing and work hours and executes the project for a period of time. The simulation can be stopped so the parameters can be tweaked.	Software Engineering	(Dantas, Barros, & Werner, 2004a) (Barros, Dantas, Veronese, & Werner, 2006) (Dantas, Barros, & Werner, 2004b)
63	Requirements Game	Requirements game is a one-and-half-hour to two-hour simulation of the software development process. Participants are in charge of the development of a software application, with special requirements previously established by the customer. This game reproduces some of the real situations of method-oriented software development projects.	Software Engineering	(Zapata & Awad-Aubad, 2007)
64	Lecture Quiz	The lecturer acts as a game show host and students answer multiple-choice questions about a particular software design issue through their laptop or mobile phone. Players have to answer correctly to get to the next round. The winner is the last person standing.	Software Engineering	(Wang, Fsdahl, & Morch-Storstein, 2008)
65	PlayScrum	The card game PlayScrum, based on its predecessor Problems and Developers, is a competition game in which each student plays the role of a Scrum Master in a software development that follows the practices of Scrum. PlayScrum can be played by between 2 and 5 players. The game is divided into sprints that differ from project to	Software Engineering	(Fernandes & Sousa, 2010)

		project and during which each player must develop a number of tasks defined at the start of the game.		
66	Management in Software Companies: a Classroom Game	The game is about simulating three main functions of management in software companies: estimation of cost and time, clear definition of roles inside a software company, and general knowledge about management issues. The game employs different methods for every task to be simulated. Firstly, estimation of cost and time is simulated by a bidding process, in which players in the role of managers must "guess" the cost and time of software projects. Secondly, the clear definition of roles inside a software company is provided by the role definition in the game. Finally, the general knowledge about management issues is simulated by a questionnaire in which we gather several issues on management techniques (planning, estimation, software project management, etc.).	Software Engineering	(Zapata, 2010)
67	Age of Computers	The game has a simple plot. Initially, the player is caught in a time-storm and transported to the past, where he or she meets von Neumann in his study. The objective is to earn enough points to travel back to the present by walking through various rooms and solving problems.	Architecture and Organization	(Sindre, Natvig, & Jahre, 2009) (Natvig & Line, 2004) (Natvig, Line, & Djupdal, 2004)
68	Software project simulator	The simulator offers various exercises in which learner simulates software projects comparing life cycle models, risk management, software inspections, etc. The student's tasks are related to software project management activities, such as, planning and monitoring & control. The learner can provide input, monitor and adjust project variables via a graphical control panel.	Software Engineering	(Collofello, 2000)
69	System Dynamics (SD) simulation	Scenario-driven interactive single-player web-based environment in which the learner has to plan and control a	Software Engineering	(Pfahl, Ruhe, & Koval, 2001)

		software project in the role of a project manager. For example, s/he can take corrective actions to complete the project considering the given resources and constraints, each action associated with project management principles and linked to model parameters. The system uses a System Dynamics (SD) simulation model, which represents three phases in a simplified, generic waterfall-model: design, implementation and test. The system presents simulation results as well as the possibility to analyze and interpret them.		(Pfahl, Laitenberger, Dorsch, & Ruhe, 2003) (Pfahl, Laitenberger, Ruhe, Dorsch, & Krivobokova, 2004) (Rodriguez, Sicilia, Cuadrado-Gallego, & Pfahl, 2006)
70	Groupthink exercise in Second Life	Multi-player Second Life version of the Groupthink exercise game focusing on requirements engineering. The objective of the game is to test the ability of a group of learners to reach consensus on software specifications. After discussing the specification within the group, players individually answer a set of questions on the specifications and the system evaluates the number of agreeing answers and presents performance statistics and the winner.	Software Engineering	(Ye, Liu, & Polack-Wahl, 2007)
71	MO-SEProcess	MO-SEProcess is a multiplayer online SE process game based on SimSE supporting the waterfall life cycle. Learners can choose one of six SE roles forming a development team together with other players. During the game, a player executes the selected role and can interact with other players. A team score is given at the end of game, if the team delivers the product before the deadline.	Software Engineering	(Ye, Liu, & Polack-Wahl, 2007)
72	SE•RPG	SE • RPG is a tool that simulates the software development environment through a game which is a fictional scenario development company various characters with whom the student must interact during the	Software Engineering	(Benitti & Molléri, 2008)

		progress of game, being supported by a system of rules by which the actions are validated.		
73	The MIS Game	The game proposes an understanding of traditional information systems development life-cycle concept, IS development project management and decision making.	Software Engineering	(Martin, 2000)
74	Information Systems Project Manager Game - ISPM Game	Aims to provide understanding on IS development project management and decision making.	Software Engineering	(Martin, 2000)
75	Paths	Paths string together a series of questions that students must complete in sequence so as to complete a path. They may either be multiple choice questions or more general coding questions where automated public and private test cases are run to verify the students' answers.	Software Development Fundamentals	(Leong, Koh, & Razeen, 2011)
76	Tower of Cubes	A list of cubes in the tower represents the data in a stack or queue depending on the selected mode on Stack or Queue. When the game starts, 5 cubes with random 2 colors are dropped into the tower. Subsequently, a new cube with a random color is dropped into the tower automatically either to the top/rear of the tower depending on the mode. When two consecutive cubes with the same color meet, both cubes will be disappeared. The player needs to move the cube in and out in order to clear the cubes in the tower and the score will be increased. When all cubes are cleared in the tower, the player wins.	Algorithms and Complexity	(Tan & Seng, 2010)
77	Software Development Game	Players must build origami boxes with one of the following four groups of letters, SO, FT, WA or RE. Every box represents a software module (a part of a software piece that can be exchangeable with others). One group of four modules forms one software piece (a complete word, SOFTWARE, made of four modules). Every module must accomplish a set of pre-defined	Software Engineering	(Zapata, 2009)

		requirements which can be discussed—for the sake clarity—with the director of the game.		
78	DesigMPS	The player assumes the role of a process engineer and given process descriptions (based on the Brazilian software process improvement model MPS.BR) has to model the process. The game offers 4 levels with increasing degree of difficulty in terms of given elements. The created process models are compared to pre-defined solutions and based on their degree of similarity a score is assigned. Goal of the game is to achieve maximum score.	Software Engineering	(Chaves, Miranda, Tavares, Oliveira, & Favero, 2011)
79	CounterMeasures	CounterMeasures is designed as a single-player game. The player is guided through several missions, each teaching a new aspect of security. Each mission has a title, a description, a score for completing the mission, a skill as the focus of the mission, objectives required to complete the mission, help given to guide the user, and a list of commands learned during the mission. The missions build upon each other, allowing the player to utilize previously learned skills in each new mission.	Information Assurance and Security	(Jordan, Knapp, Mitchell, Claypool, & Fisler, 2011)
80	Algorithms Recursive Game	The game has 44 questions of three (3) levels of difficulty covered various aspects of recursive algorithms. They aimed to revise "recursion" and develop critical reflection on what they have learned and what they have not consolidated sufficiently. The students played in pairs. During the game, each student rolls the dice; whoever gets the highest number begins the match and each one takes a 'turn'. Players put their tokens at the start. Token moves according to the number indicated on the dice. If the number is six, the player rolls the dice again. If a player arrives at the bottom of a ladder or at the head of a snake, a question is displayed. If player answers correctly, his or	Algorithms and Complexity	(Rossiou & Papadakis, 2007)

		her token moves up, at the top of the ladder or stays at the head of the snake, accordingly. Otherwise, the token stays at the bottom of the ladder or moves down to the tail of the snake. The first player who arrives at the square 100 wins the game		
81	Detective Game – what killed the project?	The game takes place in the context of a hypothetical software company that completed a software project for the development of a pizza ordering web site. The project failed and now the company is contracting the players as consultants in order to identify what went wrong. Therefore, the players receive a set of project documents. The players have to analyze the project documentation and track project progress by revising the weekly status reports applying EVM. For each correct calculation and correctly identified time or cost overrun, the groups receive a point. The winner is the group of players that obtains the largest number of points.	Software Engineering	(Soares & Rausis, 2011)
82	Computer Architecture Mini-Game	The player has to solve several problems on different levels of difficulty on logic gates and elements of computer architecture. The game start with an initial score (1000 points) and each time a player makes a wrong decision the punctuation of the score is decreased. Textual hints are provided.	Architecture and Organization	(Melero, Hernández-Leo, & Blat, 2012)
83	Computing Networks Mini-Game	The game encourages players to solve three different situations (one per level): how to connect different elements to have ADSL in home, how the information travels through the network, and how information is sent from a personal computer to another. Routers, IPs, Computer ports, browsers or frames are examples of puzzle pieces in the game. The game starts with no score and increments the punctuation when players make correct actions and decrements when making mistakes.	Networking and Communications	(Melero, Hernández-Leo, & Blat, 2012)

84	Programming Fundamentals Mini-Game	The player has to solve several problems on different levels of difficulty. The game provides the player with a skeleton of the code where s/he then has to place the different code pieces. The game start with an initial score (1000 points) and each time a player makes a wrong decision we decrease the punctuation of the score. Visual color-coded hints are provided.	Software Development Fundamentals	(Melero, Hernández-Leo, & Blat, 2012)
85	PDConsole	The basis of the game is a barebone, but fully functional integrated website that combines features of social networking and video sharing. The site displays (mock) advertisements when different pages are viewed. The game generates artificial traffic against the site, and the overall system metric of health is the number of advertisements served and resulting revenue. The game is a sort of firefighting exercise that begins when the game administrator breaks parts of the system in some way. Players notice that the system's performance has degraded because ads and revenue drop. Then they have to identify the problem and fix it.	Parallel and Distributed Computing	(Wein, Kourtchikov, Cheng, Gutierez, Khmelichek, Topol, & Sherman, 2009)
86	Crossword Puzzles	The Crossword Puzzles are designed more for knowledge reinforcement than for discovery of new concepts. Students in a senior-level Fundamentals of Computer Theory course were offered Crossword Puzzles as a means of reinforcing definitions of terms.	Algorithms and Complexity	(Hill, Ray, Blair, & Carver Jr., 2003)
87	Jeopardy!® Game	In many courses there is a large amount of information that falls into the “know” or “be familiar with” or “know how to” categories. Definitions of terms are one example, as are simple problems, calculations, or algorithms. These are ideal candidates for inclusion in a Jeopardy!®-style game.	Algorithms and Complexity	(Hill, Ray, Blair, & Carver Jr., 2003)

88	BattleThreads Game	The class is broken down into one controller and two teams. The players are responsible for the placement of one ship each, and for firing a shot from that ship each turn until their own ship is destroyed (or their side wins by sinking all of their opponents' ships). The controller gets the enemy team's ship layout and announces the effect (hit or miss) of each shot. At the end of a turn, the controller compares battle damage and reports the results. No other means of communications are allowed.	Operating Systems	(Hill, Ray, Blair, & Carver Jr., 2003)
89	Process State Transition Game	Each group of players is given a game board representing the seven-state process transition model. One of the students is selected to be the operating system (OS), one the timekeeper (TK), and the others become programs, each keeping track of some number of processes as they are managed by the operating system. When a process moves into a state in which it must be present in memory (Ready, Running, etc.), the player who owns that process places its memory markers on a grid representing available memory. When the process is suspended, the memory markers are lifted, indicating that the process has been moved out of main memory.	Operating Systems	(Hill, Ray, Blair, & Carver Jr., 2003)
90	ERPsim	Using a continuous-time simulation, students are put in a situation in which they have to run their business with a real-life ERP system. Students, thus operate a company; be it a bottled water distributor, a make-to-stock cereal manufacturer, or a Dairy company, and must interact with suppliers and customers by sending and receiving orders, delivering their products and completing the whole cash-to-cash cycle.	Information Systems	(Leger, Charland, Feldstein, Robert, Babin, & Lyle, 2011)

91	Security Protocol Game	Students play the game in groups of 4-5 players. Within each group, one student is selected to play each of Alice and Bob, the two communicating parties. Another student is selected to play Gavin. The same student may also take the role of Colin. The remaining student or students take the role of Trudy the intruder. The game commences with the students seated around a table: Alice and Bob at opposite ends, Trudy on one side and Gavin opposite her. The students select a game scenario to play, and a protocol to use in the scenario. In a typical scenario, Alice wishes to purchase computer software from Bob over the Internet using her credit card for payment. The students may choose to simulate the Transport Layer Security protocol (TLS; formerly called SSL and used to secure transactions on the world wide web) for this scenario, or other protocols, some of which are vulnerable to various attacks. The protocols involve messages being passed between Alice, Bob and Gavin. All messages are actually passed via Trudy, who may attempt to attack the protocol by monitoring or modifying the messages. The students find this a stimulating group activity as they help each other run the protocol correctly and try to think up ways to subvert it.	Information Assurance and Security	(Hamey, 2003)
92	U-Test	U-Test is a simulation game to support software testing with focus on unit tests and black box techniques, approaching theoretical and practical questions. The game is based on a case where the player is seen as a candidate to a position in a software company. After an interview the player must solve challenges to prepare unit test cases.	Software Engineering	(Thiry, Zoucas, & Silva, 2011)
93	Operating systems role plays	In the game, students assume the roles of processes (instructors representing processors). The game itself involves running sample sets of program code in a step-	Operating Systems	(Leverington, Yuksel, & Robinson, 2009)

		wise fashion so that students can see the actions and consequences of each segment of code.		
94	Dealing with difficult people	During this activity, groups of learners realize a project kick-off meeting. One of the group members represents the project manager, who has to coordinate the meeting and to make sure that in the end of the meeting all members give their commitment. The problem is that each of the other group members is assigned a role of a difficult person being instructed to “act” as e.g., a whiner, no-person etc. Once the project manager reacts in an adequate way, the person turns into a cooperative member and gives her commitment to the project.	Software Engineering	(Battistella, Wangenheim & Carvalho, 2013)
95	Starter MMO	The game combines routing and forwarding. Students start out in an area, marked start on a map representing a maze. In this area there is a Non-Player Character (NPC) that gives them quests to perform. These quests all consist of delivering a package (representing a 'data packet') to another NPC (host) somewhere. The students are neither told where the destination is located nor about the layout of the maze. The aim is to run to the NPCs location (simulating transmission of the data packet) and deliver the package to them.	Networking and Communications	(Graven, Hansen, & MacKinnon, 2009)
96	AMEISE	AMEISE is based on SESAM. In the game, students assume the role of a technical project manager. They can hire and fire personnel, structure the project, and allocate tasks. Students are challenged to manage a project according to a particular model of the problem structure, selected by the instructor. It is up to the instructor to select the number of trials (simulation runs) to solve given tasks within specified constraints. Students can learn from previous simulation runs, change their strategies and	Software Engineering	(Mittermeir, Hochmüller, Bollin, Jäger, & Nusser, 2003)

		measure their own success using the AMEISE self-assessment feature.		
97	C-Jump	The goal of the game is to find the most efficient way to "ski" down a mountain. Players are to imagine themselves as either skiers or snowboarders, racing with each other to reach the finishing line. The catch is that the player must make decisions based on common computer programming syntax, such as if(X==1) you can go down a certain path. The first player to move all his or her pieces of pawns past the finishing line would be the winner.	Software Development Fundamentals	(Singh, Dorairaj, & Woods, 2007)
98	BINX	The game takes place in the scenario of a computer being attacked by a virus. The main character is Chip, an integrated circuit designed with the purpose to wipe all traces of the malicious virus from the computer. Inside the computer, information is flowing from the motherboard to all output devices attached to the computer except the monitor. The adventure takes place inside the CPU. The player must navigate a path through the bus to find the viral infection plaguing the graphics processor and resolving several missions on different levels.	Architecture and Organization	(Morsi & Richards, 2012)
99	Anti-Phishing Phil	The main character of the game is Phil, a young fish living in the Interweb Bay. Phil wants to eat worms, so he can grow up to be a big fish, but has to be careful of phishers that try to trick him with fake worms (representing phishing attacks). Each worm is associated with a URL, and Phil's job is to eat all the real worms (which have URLs of legitimate web sites) and reject all the bait (which have phishing URLs) before running out of time. The other character is Phil's father, who is an experienced fish in the sea. He occasionally helps Phil out by giving Phil some tips on how to identify bad worms (and hence, phishing web sites). The game is split into	Information Assurance and Security	(Sheng, Magnien, Kumaraguru, Acquisti, Cranor, Hong, & Nunge, 2007)

		four rounds, each one more difficult than the previous and focusing on a different type of deceptive URLs. Players have to correctly recognize at least six out of eight URLs within two minutes to move on to the next round. If a player loses all three lives the game is over.		
100	Lego Factory	The simulation shows a physical production process in the form of a miniature factory for producing "bubble gums" using Lego bricks and Mindstorms NXT. As the factory production manager, the player's task is to find the best configuration (production flow and buffer sizes) possible for the production process in order to maximize profit. The player wins the game, if the optimal configuration of the factory is found. In order to find optimal solutions, players have to develop AI algorithms.	Intelligent Systems	(Syberfeldt & Syberfeldt, 2010)
101	Control-Alt-Hack	The player acts as white hat hackers in a security consulting company. Each player is given a Hacker card. Gameplay is centered around missions—a variety of audit jobs and pro bono work that require the selective application of hacker skills: Hardware Hacking, Network Ninja, Cryptanalysis, Forensics, etc. The character's skill levels and player's dice rolls determine whether the player succeeds or fails at a mission. Players can increase their skill levels by purchasing useful items (Bag of Tricks); opponents can hinder player's efforts to complete a mission by playing Lightning Strikes on them. Mission successes and failures lead to the gain and loss of Hacker Cred. Players win the game by accruing enough Hacker Cred and becoming the CEO of their own consulting company.	Networking and Communications	(Denning, Kohno, & Shostack, 2012)
102	3DAR Lego Game	Players have to re-construct a 3D model using its decomposed pieces. The game provides 3 game levels with increasing difficulty. At level 1, all pieces are a	Human Computer Interaction	(Do & Lee, 2009)

		precise component of the model, at level 2 each piece is a primitive geometry (cube, cone, etc.) that has to be customized and at level 3 no pieces are provided, and the players have to create the model from primitive geometries.		
103	Database concurrency control card game	In the game, the students play cards and build a schedule for a given set of simultaneous transactions. They use their knowledge about transactions and concurrency control protocols to simulate the work of a transaction processing system. The sequence of activities is simulated through a control card or Kanban that is passed from student to student. The card exchange is similar to a procedure call. Each student has to decide what to do at his turn, based on the concepts s/he learned, and the concurrency control protocol being used.	Information Management	(Kern, Stotz, & Bento, 2008)
104	Digital System Game	The game starts with the player in one corner of an imaginary 3D world similar to those found in first person-shooter games. The player's goal is to reach the exit, which can be accomplished by unlocking several doors and obtaining two skill upgrades. At each locked door, the player is presented with a sum-of-products combinational circuit problem. The game switches to a 2D environment for the digital circuit design problems. The game updates the external outputs automatically to indicate the values of the outputs of the current circuit for the specified input values.	Architecture and Organization	(Srinivasan, Butler-Purry, & Pedersen, 2008)
105	Mystery of Traffic Lights	Designed from a first-person perspective, the game starts at the major intersection of a small town, where an engineer character, Jack, is standing there frustrated by the busy and messy traffic due to malfunctioning traffic lights. He then invites students to help him redesign the controller with the right logic for the current traffic flow.	Architecture and Organization	(Tang, Shetty, & Chen, 2012)

106	Project Execution Game (PEG)	In this game, the players work in groups in order to manage a project. The players receive a detailed project plan and the success evaluation criterion, which is to complete the project with the minimum possible budget. Project overruns, caused by the players' decisions, cause penalties and overhead cost, which negatively affect the team's success.	Software Engineering	(Zwikael & Gonen, 2007)

Analysis factors used for the evaluation of games for computing education

ID	Game	Quality factors
1	Wu's Castle	Learning, Enjoyability, Confidence, Fun
2	X-MED	Competence, Learning, Relevance, Correctness, Sufficiency, Challenge, Sequence, Adequacy, Satisfaction, Fun, Interest
3	SimSE	Learning, Enjoyability, Confidence
		Learning, Enjoyability, Confidence, Fun, Attention, Relevance, Satisfaction
		Learning, Enjoyability, Confidence, Fun
		Learning, Attention, Confidence, Relevance, Satisfaction
4	EleMental: The recurrence	Learning, Enjoyability, Quality
5	DELIVER!	Learning, Attention, Relevance, Confidence, Satisfaction, Immersion, Challenge, Competence, Fun, Social Interaction
6	Saving Sera	Learning, Enjoyability, Quality
7	The Catacombs	Learning, Enjoyability, Quality
8	The Risk Management Game	Realism, Enjoyability, Ease of use, Learning
9	Program your robot	Quality

10	SortingGame	Usefulness, Confidence, Relevance, Fun, Learning
11	SortingCasino	Usefulness, Confidence, Relevance, Fun, Learning
12	Secret rule	Usefulness, Confidence, Relevance, Fun, Learning
13	Draw and guess	Usefulness, Confidence, Relevance, Fun, Learning
14	SCRUMIA	Learning, Attention, Relevance, Confidence, Satisfaction, Immersion, Challenge, Competence, Fun, Social Interaction
15	Software engineering education game in Second Life	Learning
16	BOTS	Learning
17	SpITKom Project	Fun, Interaction, Challenge
18	Prog&Play	Learning, Fun, Clear goal, Helpfulness, Ease of use, Intuitive, Visually pleasing, Feedback
19	Light-Bot	Learning
20	Software Development Manager (SDM)	Enjoyability, Confidence, Learning, Fun
21	PM Master	Learning, Attention, Relevance, Confidence, Satisfaction, Immersion, Challenge, Competence, Fun, Social Interaction
22	Requirements Collection and Analysis Game (RCAG)	Learning, Engagement, Fun, Quality
23	Innov8	Learning, Attention, Relevance, Confidence, Satisfaction, Attention, Clear goal, Retrieval, Challenge, Helpfulness, Responding, Feedback, Realism
24	Kernel Panic	Learning
25	I-SEE	Learning
26	hACMEgame	Learning
27	Evaluators	Interaction, Adequacy, Learnability, Ease of use, Enjoyability, Recommendation
28	Simsoft	Learning, Satisfaction, Recommendation
29	The Requirements Island	Learning, Quality

30	Motherboard-assembly pairing game	Concentration, Clear Goal, Feedback, Challenge, Control, Immersion, Social Interaction, Learning
31	Describing computer parts	Concentration, Clear Goal, Feedback, Challenge, Control, Immersion, Social Interaction, Learning
32	Hands-on OS game	Concentration, Clear Goal, Feedback, Challenge, Control, Immersion, Social Interaction, Learning
33	Bear-cub's computer game	Concentration, Clear Goal, Feedback, Challenge, Control, Immersion, Social Interaction, Learning
34	Maze	Engagement, Visually pleasing, Usefulness
35	Coloring Map	Enjoyability, Challenge, Confidence, Learning, Recommendation, Helpfulness
36	Scrum LEGO Challenge	Learning
37	Star Chef	Usefulness, Ease of use , Recommendation
38	Object Access Tool (OAT)	Learning, Quality
39	DesigMPS	Learning
40	Data Defender	Quality
41	Ztech	Learning, Quality
42	InspectorX	Learning, Interest
43	SPARSE	Ease of use, Interest, Learning
44	SORTIA	Learning, Attention, Relevance, Confidence, Satisfaction, Immersion, Challenge, Competence. Fun. Social Interaction
45	InspSoft	Learning, Attention, Relevance, Confidence, Satisfaction, Immersion. Challenge, Competence, Fun, Social Interaction
46	SimMS	Ease of use, Usefulness
47	UbiRE	Feedback, Control, Intuitive, Efficiency, Visually pleasing, Helpfulness, Learnability, Interaction
48	CyberCIEGE	Learning

49	ProGames	Usefulness, Ease of use, Learnability, Satisfaction
50	IT Billionare Game Board	Learning
51	CAPTAIN3	Learning
52	The Last Java Code	Learning, Quality
53	Tower-Defense	Learning, Quality
54	The aMAZEing Labyrinth	Learning, Quality
55	Quoridor	Learning, Quality
56	San Francisco Cable Cars	Learning, Quality
57	NeverWinter Nights 2	Learning
58	Problems and Programmers	Learning, Enjoyability, Confidence, Fun
59	SESAM	Learning
		Learning
		Quality
60	Open Software Solution (OSS)	Attention, Control, Efficiency, Helpfulness, Learnability, Immersion
61	SimjavaSP	Learning, Enjoyability, Confidence, Fun
62	Incredible Manager	Learning, Fun
		Quality
		Learning, Fun
63	Requirements Game	Structure, Style, Systems, Strategy, Shared Values, Staff, Skills
64	Lecture Quiz	Usefulness, Ease of use, Learnability, Intuitive
65	PlayScrum	Enjoyability, Confidence, Learning, Fun, Quality

66	Management in Software Companies: a Classroom Game	Learning, Quality
67	Age of Computers	Learning
		Learning, Quality
68	Software project simulator	Quality
69	System Dynamics (SD) simulation	Learning
		Learning
		Learning
		Learning
70	Groupthink exercise in Second Life	Learning, Quality
71	MO-SEProcess	Learning, Quality
72	SE•RPG	Learning
73	The MIS Game	Quality
74	Information Systems Project Manager Game - ISPM Game	Quality
75	Paths	Learning, Social Interaction
76	Tower of Cubes	Learning, Fun, Satisfaction
77	Software Development Game	Quality
78	DesigMPS	Interest, Motivation, Recommendation
79	CounterMeasures	Learning, Interest
80	Algorithms Recursive Game	Ease of use, Learning, Quality
81	Detective Game – what killed the project?	Learning, Attention, Relevance, Confidence, Satisfaction, Immersion, Challenge, Competence, Fun, Social Interaction

82	Computer Architecture Mini-Game	Quality
83	Computing Networks Mini-Game	Quality
84	Programming Fundamentals Mini-Game	Quality
85	PDConsole	Quality
86	Crossword Puzzles	Learning
87	Jeopardy!® Game	Quality
88	BattleThreads Game	Learning
89	Process State Transition Game	Quality
90	ERPsim	Learning
91	Security Protocol Game	Enjoyability, Confidence, Learning, Quality
92	U-Test	Learning, Relevance, Correctness, Sufficiency, Confidence
93	Operating systems role plays	Quality
94	Dealing with difficult people	Learning, Attention, Relevance, Confidence, Satisfaction, Immersion, Challenge, Competence, Fun, Social Interaction
95	Starter MMO	Enjoyability, Learning, Confidence, Usefulness
96	AMEISE	Learning
97	C-Jump	Ease of use, Clear goal, Learning, Fun, Visually pleasing
98	BINX	Helpfulness, Confidence
99	Anti-Phishing Phil	Learning, Quality
100	Lego Factory	Quality
101	Control-Alt-Hack	Quality

102	3DAR Lego Game	Interest, Practical, Intuitive, Confidence, Learning
103	Database concurrency control card game	Learning
104	Digital System Game	Learning, Quality
105	Mystery of Traffic Lights	Interest, Motivation, Learning
106	Project Execution Game (PEG)	Learning, Challenge, Interest, Fun Confidence

Mapping and definition of the analysis factors

Analysis factors		Definition	Related Analysis factors	Definition
Learning		The evaluation of this factor often refers to the improvement of competence by comparing their competence level after game playing with their competence level beforehand, typically based on a post-test score or a self-assessment. Few studies evaluate the learning effect with regard to a systematic definition of learning levels, e.g., based on Bloom's taxonomy.		
Motivation	Confidence	Confidence means to enable students to make progress in the study of educational content through their effort and ability (e.g., through exercises with increasing level of difficulty) (Keller, 1987).		
	Relevance	Relevance refers to the students need to realize that the educational proposal is consistent with their goals and that they can link content with their professional or academic future (Keller, 1987).	Content relevancy (Gresse von Wangenheim, Thiry & Kochanski, 2009) Importance (Gresse von Wangenheim,	The relevance of the content addressed in the game (Gresse von Wangenheim, Thiry & Kochanski, 2009). Refers to how important the student considers learn the content covered in the game (Gresse von Wangenheim, Thiry & Kochanski, 2009)

			Thiry & Kochanski, 2009)	
	Satisfaction	Satisfaction means that the students must feel that the dedicated effort results in learning (Keller, 1987).		
	Attention	Attention refers to students' cognitive responses to instructional stimuli. It is desirable to obtain and maintain a satisfactory level of attention of students during a learning period (Keller, 1987).	Attractiveness (Sharp & Hall, 2000)	Affect how much the product captures the user's emotional responses (MUMMS, 2015).
	Concentration	Screen out distraction and make concentration possible (Fu, Su, & Yu, 2009)		
	Interest	Interest refers to how much the game is interesting and attractive for the students' learning (Gresse von Wangenheim, Thiry & Kochanski, 2009)		
User Experien ce	Fun	Fun refers to students' feeling of pleasure, happiness, relaxing and distraction (Poels, Kort, & Ijsselsteijn, 2007)	Entertainment (Muratet, Torguet, Viallet, & Jessel, 2011)	The game is fun and rewarding (Muratet, Torguet, Viallet, & Jessel, 2011)
	Social interaction	Social interaction refers to the creation of a feeling of shared environment and being connected with others in activities of cooperation or competition (Fu, Su, & Yu, 2009).	Interaction (Leong, Koh, & Razeen, 2011)	The game helped to improve their interactions with the teaching staff and with their peers (Leong, Koh, & Razeen, 2011).
	Challenge	Challenge means that a game needs to be sufficiently challenging with respect to the player's competency level. The increase of difficulty should occur at an appropriate pace accompanying the learning curve. New obstacles and situations should be presented throughout the game to minimize fatigue and to keep the students interested (Sweetser & Wyeth, 2005)	Selective perception (Roodt & Joubert, 2009) Difficulty (Gresse von Wangenheim, Thiry & Kochanski, 2009)	Presenting the stimulus. The challenge within the game is given by the tasks the learner has to complete in order to advance within the game (Roodt & Joubert, 2009). Corresponds to the difficulty degree of the game levels (Gresse von Wangenheim, Thiry & Kochanski, 2009)

Immersion	Immersion allows the player to have an experience of deep involvement within the game, creating a challenge with real-world focus, so that s/he forgets about the outside world during gameplay (Fu, Su, & Yu, 2009).	Excitement (Sharp & Hall, 2000)	This is the extent to which end users feel that they are 'drawn into' the world of the multi-media application, and it seems to capture some of the fascination which the best multi-media apps exercise over their users (MUMMS, 2015)
Control	The game should also allow the player to have a sense of control over the game interactions that should be easy to learn and allow them to explore the game freely and at their own pace (Poels, Kort, & Ijsselsteijn, 2007; Takatalo, Häkkinen, Kaistinen, & Nyman, 2010).	Autonomy (Fu, Su, & Yu, 2009)	Make it possible to adjust opportunities for action to our capabilities (Fu, Su, & Yu, 2009).
Competence	Players need to realize that their competencies are at a level where it is possible to overcome the challenges of the game. As the difficulty increase, challenges should require the player to develop their competencies to advance in the game and have fun (Sweetser & Wyeth, 2005; Poels et al., 2007; Takatalo et al., 2010).		
Recommendat ion	Recommendation refers to if the students suggest the using of the game to other students (Bavota, Lucia, Fasano, Oliveto & Zottoli, 2012).	Attitute toward use (Liu, Chu, & Tan, 2012)	The students liked the game and recommend its use (Liu, Chu, & Tan, 2012)
Engagement	The game was engaging, and the game sustained their engagement (Hainey, Connolly, & Boyle, 2009)		
Realism	How real is the game (Taran, 2007)	Generalization (Roodt & Joubert, 2009)	The game involves a “real life type scenario”. This realism allows the learners to generalize the knowledge and skills they have acquired to practical examples (Roodt & Joubert, 2009).
Enjoyability	Enjoyability refers to how much the game is enjoyable (Eagle & Barnes, 2009; Navarro & van der Hoek, 2005).		

Usefulness	Usefulness	Usefulness corresponds to if the game is helpful, effective, could enhance the learning and is a sensible teaching method (Hakulinen, 2011).	Suitability (Wang, Fsdahl, & Morch-Storstein, 2008)	
Usability	Ease of use	Ease of use refers to how much the game's information is clear, organized, if the students know where they are and how to get where they want, it is user friendly and simple to use.	Easiness (Rossiou & Papadakis, 2007; Liu, Chu, & Tan, 2012) Playability (Singh, Dorairaj, & Woods, 2007) Simplicity (Taran, 2007)	Refers to the difficulties in accessing the game, if the instructions of the game helped students to play (Rossiou & Papadakis, 2007). Easiness (the user interface is friendly, It's easy of use to assist my learning) (Liu, Chu, & Tan, 2012). The players immediately started to play the game or needs to read the instructions to learn how to play (Singh, Dorairaj, & Woods, 2007). How simple was this game to play? (Taran, 2007)
	Learnability	Refers to how much the game is ease of learn, if I learned to use it quickly and I easily remember how to use (Hijon-Neira, Velazquez-Iturbide, Pizarro-Romero, & Carrico, 2014; MUMMS, 2015)		
	Helpfulness	Helpfulness measures the degree to which the software is self-explanatory, as well as more specific things like the adequacy of help facilities and documentation (MUMMS, 2015)	Semantic encoding (Roodt & Joubert, 2009)	Providing learning guidance. The game provides continual guidance in what needs to be achieved and how to achieve it (Roodt & Joubert, 2009).
	Visually pleasing	Refers to the quality of graphics (Adamo-Villani, Haley-Hermiz, & Cutler, 2013)	Aesthetic appearance (Singh, Dorairaj, & Woods, 2007)	Is the game attractive and appealing (Singh, Dorairaj, & Woods, 2007).

	Efficiency	Efficiency measures the degree to which users feel that the software assists them in their work and is related to the concept of transparency (MUMMS, 2015)		
	Intuitiveness	Refers to the degree of intuition, understanding and comprehension of the activities (Do & Lee, 2009).		
	Interaction	It's easy to identify the state of problem solving (Rodriguez-Cerezo, Sarasa-Cabezuelo, Gomez-Albarran, & Sierra, 2014)		
Instructional aspects	Clear goal	Have clearly defined goals with manageable rules (Fu, Su, & Yu, 2009)	Expectancy (Roodt & Joubert, 2009) Rules (Singh, Dorairaj, & Woods, 2007)	Informing learners of the objective. At the beginning of the game the learner is informed of the actions required to finish the game as well as how they are expected to perform these actions (Roodt & Joubert, 2009). The rules are clear enough for the players to understand how the game should be played (Singh, Dorairaj, & Woods, 2007).
	Sequence	The logical sequence of steps in the game (Gresse von Wangenheim, Thiry & Kochanski, 2009)		
	Adequacy	Adequacy corresponds to if the game suitably complements the lectures (Rodriguez-Cerezo, Sarasa-Cabezuelo, Gomez-Albarran, & Sierra, 2014).	Teaching method (Gresse von Wangenheim, Thiry & Kochanski, 2009) Pedagogical utility (Rodriguez-Cerezo, Sarasa-Cabezuelo, Gomez-Albarran, & Sierra, 2014)	The teaching method used in the evaluation, e.g an educational game (Gresse von Wangenheim, Thiry & Kochanski, 2009). The game helps me understand the content (Rodriguez-Cerezo, Sarasa-Cabezuelo, Gomez-Albarran, & Sierra, 2014)

	Recall	Stimulating recall of prior learning. In order for the learner to successfully complete certain tasks they are required to recall information given to them during the course of the game. (Roodt & Joubert, 2009)		
	Responding	Eliciting performance. The very nature of the game, i.e. a first person role-playing approach, allows the learner to actively participate in the learning process. (Roodt & Joubert, 2009).		
	Practical	Refers to if the players can practice skills (Do & Lee, 2009)		
	Feedback	Provide clear information on how the participants are doing (Fu, Su & Yu, 2009)	Reinforcement (Roodt & Joubert, 2009)	Providing feedback (Roodt & Joubert, 2009).
Correctness	Correctness	The correctness of the game in accordance with the content addressed in the game (Gresse von Wangenheim, Thiry & Kochanski, 2009)		
Completeness	Sufficiency/Completeness	Corresponds to the sufficiency degree of the game to teach a specific content (Gresse von Wangenheim, Thiry & Kochanski, 2009)		
Quality	Quality	Generic term used to describe evaluations that used open-ended questions to obtain the general perceptions of students after playing the game.		
7S-Model	Structure	Every team must establish organizational procedures. How many people were required in every organizational function? Was task specialization convenient for executing tasks? Was an organizational hierarchy needed? How the information must have flowed? Were functions clearly defined? (Zapata & Awad-Aubad, 2007)		
	Style	It reflects the behavior patterns inside the organization. What kind of organization was used? What was the		

		preferred behavior of the leaders (leadership or authority)? How was the organizational atmosphere? (Zapata & Awad-Aubad, 2007)
	Systems	Formal processes and procedures for organizational management. What was the kind of planning? How was the organization of task execution? Were processes established? (Zapata & Awad-Aubad, 2007)
	Strategy	What was the goal of the team? How was the plan for achieving this goal? What factor had higher importance: accuracy in deliveries, quality of the final product, or financial impact of decision making process? (Zapata & Awad-Aubad, 2007)
	Shared values	What values were identified in the organization? Did the player behavior reflect their own ethics? (Zapata & Awad-Aubad, 2007)
	Staff	If you could select people for your organization, what would the criterion be? Did the companies make good use of people abilities? Were synergies possible? (Zapata & Awad-Aubad, 2007)
	Skills	Did the companies identify skills, competence and expertise of their employees? (Zapata & Awad-Aubad, 2007)

Study types, evaluation models/methods and data collection instruments used in the studies

ID	Game	Study Type	Evaluation model/metho d	Instrument
1	Wu's Castle	Experimental: randomized pre-test/post-test control group (R O X1 O O R O X2 O O X1-Wu's castle X2-Programming assignment)	NI	Tests and Questionnaire
2	X-MED	Experimental: randomized pre-test/pos-test control group (R O X1 O R O O X1- X-MED)	NI	Questionnaire and Exercises

3	SimSE	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
		Experimental: randomized pre-test/post-test control group (R O X1 O R O O X1-SimSE)	NI	Questionnaire
		Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
		Ad-hoc evaluation: one-shot post-test only	NI	Observations, interview and questionnaire
4	EleMental: The recurrence	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Challenges and Questionnaire
5	DELIVER!	Non-experimental with a case study: one-shot post test only X O	MEEGA	Questionnaire
6	Saving Sera	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Questionnaire, tests and interviews
7	The Catacombs	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Questionnaire, tests and interviews
8	The Risk Management Game	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
9	Program your robot	Ad-hoc evaluation: collected feedback after play the game	NI	Questionnaire
10	SortingGame	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
11	SortingCasino	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
12	Secret rule	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
13	Draw and guess	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
14	SCRUMIA	Non-experimental with case study: one-shot post test only X O	MEEGA	Questionnaire
15	Software engineering education game in Second Life	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
16	BOTS	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Tests

17	SpITKom Project	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire, interviews and observations
18	Prog&Play	Experimental: pre-test/post-test control group O X1 O O O X1- Prog&Play	Questions derived from literature	Questionnaires
19	Light-Bot	Ad-hoc evaluation by researchers to obtain a computational thinking score for Light-Bot.	Computational Thinking Framework (CTF)	Questionnaire
20	Software Development Manager (SDM)	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
21	PM Master	Non-experimental with case study: one-shot post-test only X O	MEEGA	Questionnaire
22	Requirements Collection and Analysis Game (RCAG)	Experimental: randomized pre-test/post-test control group R O X1 O R O X2 O R O X3 O X1- RCAG, X2-Role-Play case study, X3-Paper-Based case study	Evaluation Framework for Effective Games-based Learning.	Questionnaire and tests
23	Innov8	Ad-hoc evaluation: one-shot post-test only	Bloom, ARCS, Gagne	Questionnaire and observations
24	Kernel Panic	Ad-hoc evaluation: one-shot post-test only	NI	Challenges
25	I-SEE	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Scores
26	hACMEgame	Ad-hoc evaluation with challenges	NI	Challenges
27	Evaluators	Ad-hoc evaluation: one-shot post-test only and analysis of student's grades in the last five years.	TUP (Technology, Usability and Pedagogy) Model	Questionnaire
28	Simsoft	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Questionnaire

29	The Requirements Island	Ad-hoc evaluation	NI	Questionnaire
30	Motherboard-assembly pairing game	Ad-hoc evaluation: one-shot pre-test/post-test	EGameFlow	Questionnaire
31	Describing computer parts	Ad-hoc evaluation: one-shot pre-test/post-test	EGameFlow	Questionnaire
32	Hands-on OS game	Ad-hoc evaluation: one-shot pre-test/post-test	EGameFlow	Questionnaire
33	Bear-cub's computer game	Ad-hoc evaluation: one-shot pre-test/post-test	EGameFlow	Questionnaire
34	Maze	Quasi-experimental: one-shot post-test only	NI	Questionnaire
35	Coloring Map	Ad-hoc evaluation: static group comparison group post-test (X1 O, X2 O, X1=game, X2= non game)	NI	Questionnaire
36	Scrum LEGO Challenge	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire, learning diary and interview
37	Star Chef	Quasi-experimental: static group comparison group post-test : X1 O, X2 O (X1 game, X2 learning tool)	TAM	Questionnaire, observations and interviews
38	Object Access Tool (OAT)	Ad-hoc evaluation: one-shot post-test only	NI	Observation and questionnaire
39	DesigMPS	Experimental: randomized pre-test/pos-test control group R O X1 O R O O X1- DesigMPS X2- Project-based learning method	NI	Tests
40	Data Defender	Quasi-experimental: randomized pre-test/post-test R O X O	NI	Questionnaire
41	Ztech	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
42	InspectorX	Non-experimental: one-shot pre test /post test O X O	Metrics derived using GQM	Questionnaire and Tests
43	SPARSE	Non-experimental: one-shot pre test /post test O X O	NI	Questionnaires
44	SORTIA	Non-experimental with case study: one-shot post test only X O	MEEGA	Questionnaire

45	InspSoft	Non-experimental: one-shot post test only X O	MEEGA	Questionnaire
46	SimMS	Non-experimental: one-shot post test only X O	TAM	Questionnaire
47	UbiRE	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Questionnaire
48	CyberCIEGE	Quasi-experimental: static group comparison group post-test (X O, O)	NI	Test
49	ProGames	Experimental: pre-questionnaire/pos-questionnaire experimental and control group O X1 O O O X1- ProGames	NI	Questionnaire
50	IT Billionare Game Board	Experimental: randomized pre-test/pos-test control group R O X1 O R O O	NI	Test
51	CAPTAIN3	Quasi-experimental: time series	NI	Test
52	The Last Java Code	Quasi-experimental: 2 static groups comparison group post-test (X1 O, X2 O, X1=game, X2=non game)	NI	Test, Questionnaire and interview
53	Tower-Defense	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire and Test
54	The aMAZEing Labyrinth	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
55	Quoridor	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
56	San Francisco Cable Cars	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
57	NeverWinter Nights 2	Experimental: randomized pre-questionnaire/post-test two group2 R O X1 O R O X2 O X1- non-adaptive game X2- adaptive game	Felder-Silverman learning style questionnaire and Tests	Questionnaire and test
58	Problems and Programmers	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
59	SESAM	Ad-hoc evaluation: one-shot post-test only	NI	Challenges
		Experimental: pre-test/post-test experimental and control group O X1 O O O X1- SESAM	NI	Questionnaire and Project Plan

		Ad-hoc evaluation where the participants play two sessions of SESAM and their tutor analyzed their performance and provided feedback in between.	NI	Interviews
60	Open Software Solution (OSS)	Ad-hoc evaluation: one-shot post-test only	MUMMS usability	Questionnaire
61	SimjavaSP	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
62	Incredible Manager	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
		Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
		Experimental with 2 Case studies: one-shot pre-test/post- test (training O X debriefing X debriefing O)	NI	Questionnaire
63	Requirements Game	Ad-hoc evaluation: one-shot post-test only	7-S model	Questionnaire
64	Lecture Quiz	Ad-hoc evaluation: one-shot post-test only	SUS	Questionnaire
65	PlayScrum	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
66	Management in Software Companies: a Classroom Game	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
67	Age of Computers	Experimental: pretest/post-test control-group design, O X1 O, O X2 O, O X3 O, X1=Game, X2=paper exercises, X3=read paper	NI	Tests
		Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
68	Software project simulator	Non-experimental with a case study: one-shot post test only X O	NI	Questionnaire
69	System Dynamics (SD) simulation	Experimental: Randomized pre-test/post-test control group R O X1 O R O X2 O X1: SD simulation model with roleplay scenario X2: COCOMO without role-play scenario	NI	Questionnaire and Test
		Experimental: Randomized pre-test/post-test control group R O X1 O R O X2 O X1: SD simulation model with roleplay scenario X2: COCOMO without role-play scenario	NI	Questionnaire and Test
		Experimental: Randomized pre-test/post-test control group R O X1 O R O X2 O X1: SD simulation model with roleplay scenario X2: COCOMO without role-play scenario	NI	Questionnaire and Test

		Experimental: Randomized pre-test/post-test control group R O X1 O R O X2 O X1: SD simulation model X2: COCOMO	NI	Questionnaire and Test
70	Groupthink exercise in Second Life	Non-experimental with a case study: one-shot post test only X O	NI	Questionnaire
71	MO-SEProcess	Non-experimental with a case study: one-shot post test only X O	NI	Questionnaire
72	SE•RPG	Non Experimental: one-shot pre-test/post- test O X O	NI	Questionnaire
73	The MIS Game	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
74	Information Systems Project Manager Game - ISPM Game	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
75	Paths	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
76	Tower of Cubes	Non-experimental with a case study: one-shot post test only X O	NI	Questionnaire
77	Software Development Game	Non-experimental with a case study: one-shot post test only X O	NI	Questionnaire
78	DesigMPS	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
79	CounterMeasures	Experimental: Randomized pre-test/post-test control group R O X1 O R O X2 O X1: game X2: read	NI	Performance statistics and questionnaires
80	Algorithms Recursive Game	Non-experimental with a case study: one-shot post test only X O	NI	Questionnaire and interviews
81	Detective Game – what killed the project?	Non-experimental with a case study: one-shot post test only X O	MEEGA	Questionnaire
82	Computer Architecture Mini-Game	Ad-hoc evaluation: static group comparison group post-test (X1 O, X2 O X1-play mini-game with hints and X2-play mini-game without assistance)	NI	Questionnaire
83	Computing Networks Mini-Game	Ad-hoc evaluation: static group comparison group post-test (X1 O, X2 O X1-play mini-game with hints and X2-play mini-game without assistance)	NI	Questionnaire

84	Programming Fundamentals Mini-Game	Ad-hoc evaluation: static group comparison group post-test (X1 O, X2 O X1-play mini-game with hints and X2-play mini-game without assistance)	NI	Questionnaire
85	PDConsole	Ad-hoc evaluation with game sessions and observations	NI	Observations
86	Crossword Puzzles	Ad-hoc evaluation: one-shot post-test only	NI	Interviews
87	Jeopardy!® Game	Ad-hoc evaluation: one-shot post-test only	NI	Interviews
88	BattleThreads Game	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
89	Process State Transition Game	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
90	ERPsim	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
91	Security Protocol Game	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
92	U-Test	Experimental: Randomized pre-test/post-test control group R O X1 O R O X2 O X1: U-Test X2: placebo	NI	Questionnaire
93	Operating systems role plays	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
94	Dealing with difficult people	Non-experimental with case study: one-shot post test only X O	MEEGA	Questionnaire
95	Starter MMO	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
96	AMEISE	Ad-hoc evaluation: one-shot post-test only	NI	Informal
97	C-Jump	Ad-hoc evaluation: one-shot post-test only	NI	Interviews, observations, recording verbal communications and questionnaire
98	BINX	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Questionnaire
99	Anti-Phishing Phil	Experimental: Randomized pre-test/post-test control group R O X1 O R O X2 O X1: game group X2: tutorial group	NI	Questionnaire

10 0	Lego Factory	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
10 1	Control-Alt-Hack	Ad-hoc evaluation: one-shot post-test only	NI	Informal
10 2	3DAR Lego Game	Ad-hoc evaluation: one-shot post-test only	NI	Questionnaire
10 3	Database concurrency control card game	Ad-hoc evaluation during the semester	NI	Tests, questions and students average
10 4	Digital System Game	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Test and Questionnaire
10 5	Mystery of Traffic Lights	Ad-hoc evaluation: static group comparison group pre/post-test	NI	Tests and Questionnaire
10 6	Project Execution Game (PEG)	Ad-hoc evaluation: one-shot pre-test/post-test	NI	Questionnaire

Sample size and replication of the evaluations

ID	Game	Sample size categories	Replication
1	Wu's Castle	41 - 60 participants	One replication where the experimental and control groups switch roles in same context
2	X-MED	1 - 20 participants	Three experiments in parallel without any modifications
3	SimSE	> 120 participants	Multi-site study in three different universities.
		1 - 20 participants	NI
		21 - 40 participants	NI
		1 - 20 participants	NI
4	EleMental: The recurrence	1 - 20 participants	NI

5	DELIVER!	21 - 40 participants	Applied in two courses in same context
6	Saving Sera	21 - 40 participants	Two sequential studies in same context
7	The Catacombs	21 - 40 participants	Two sequential studies in same context
8	The Risk Management Game	> 120 participants	NI
9	Program your robot	21 - 40 participants	NI
10	SortingGame	1 - 20 participants	NI
11	SortingCasino	1 - 20 participants	NI
12	Secret rule	1 - 20 participants	NI
13	Draw and guess	1 - 20 participants	NI
14	SCRUMIA	61 - 80 participants	Applied several times in two courses in same context
15	Software engineering education game in Second Life	41 - 60 participants	NI
16	BOTS	NI	NI
17	SpITKom Project	41 - 60 participants	NI
18	Prog&Play	> 120 participants	Applied in 3 iterations in different universities
19	Light-Bot	NI	NI
20	Software Development Manager (SDM)	21 - 40 participants	NI
21	PM Master	21 - 40 participants	NI
22	Requirements Collection and Analysis Game (RCAG)	81 - 100 participants	5 applications in same context with different audience
23	Innov8	1 - 20 participants	NI
24	Kernel Panic	NI	NI
25	I-SEE	NI	NI
26	hACMEgame	41 - 60 participants	NI
27	Evaluators	61 - 80 participants	NI
28	Simsoft	41 - 60 participants	NI
29	The Requirements Island	41 - 60 participants	2 experiments in same context

30	Motherboard-assembly pairing game	> 120 participants	NI
31	Describing computer parts	> 120 participants	NI
32	Hands-on OS game	> 120 participants	NI
33	Bear-cub’s computer game	> 120 participants	NI
34	Maze	1 - 20 participants	NI
35	Coloring Map	1 - 20 participants	NI
36	Scrum LEGO Challenge	21 - 40 participants	NI
37	Star Chef	101 - 120 participants	NI
38	Object Access Tool (OAT)	1 - 20 participants	3 evaluations in same context
39	DesigMPS	41 - 60 participants	Applied in 2 parallel courses in same context
40	Data Defender	1 - 20 participants	NI
41	Ztech	21 - 40 participants	NI
42	InspectorX	21 - 40 participants	NI
43	SPARSE	1 - 20 participants	NI
44	SORTIA	1 - 20 participants	NI
45	InspSoft	1 - 20 participants	NI
46	SimMS	1 - 20 participants	NI
47	UbiRE	1 - 20 participants	NI
48	CyberCIEGE	1 - 20 participants	NI
49	ProGames	61 - 80 participants	Applied several times in different courses at same university
50	IT Billionare Game Board	1 - 20 participants	NI
51	CAPTAIN3	61 - 80 participants	NI
52	The Last Java Code	21 - 40 participants	NI
53	Tower-Defense	1 - 20 participants	3 studies with different audiences in same context
54	The aMAZEing Labyrinth	> 120 participants	NI
55	Quoridor	> 120 participants	NI
56	San Francisco Cable Cars	> 120 participants	NI
57	NeverWinter Nights 2	41 - 60 participants	NI
58	Problems and Programmers	21 - 40 participants	NI

59	SESAM	NI	NI
		1 - 20 participants	NI
		21 - 40 participants	Applied in 2 sessions with the same audience
60	Open Software Solution (OSS)	21 - 40 participants	Applied in 2 courses in same context
61	SimjavaSP	NI	NI
62	Incredible Manager	1 - 20 participants	Two studies in different contexts (university and laboratory for industrial software development)
		21 - 40 participants	NI
		1 - 20 participants	Two studies in different contexts (university and laboratory for industrial software development)
63	Requirements Game	81 - 100 participants	NI
64	Lecture Quiz	1 - 20 participants	NI
65	PlayScrum	1 - 20 participants	NI
66	Management in Software Companies: a Classroom Game	21 - 40 participants	NI
67	Age of Computers	61 - 80 participants	NI
		> 120 participants	NI
68	Software project simulator	1 - 20 participants	NI
69	System Dynamics (SD) simulation	1 - 20 participants	NI
		1 - 20 participants	Replicated once in other university
		1 - 20 participants	Replicated once in other university
		1 - 20 participants	NI
70	Groupthink exercise in Second Life	21 - 40 participants	Applied in 2 courses in same context
71	MO-SEProcess	21 - 40 participants	Applied in 2 courses in same context
72	SE•RPG	21 - 40 participants	NI
73	The MIS Game	NI	Replicated several times in same context

74	Information Systems Project Manager Game - ISPM Game	NI	Replicated several times in same context
75	Paths	41 - 60 participants	NI
76	Tower of Cubes	21 - 40 participants	NI
77	Software Development Game	> 120 participants	NI
78	DesigMPS	1 - 20 participants	NI
79	CounterMeasures	1 - 20 participants	NI
80	Algorithms Recursive Game	21 - 40 participants	NI
81	Detective Game – what killed the project?	41 - 60 participants	Applied in two courses in same context
82	Computer Architecture Mini-Game	41 - 60 participants	Applied several times in different schools
83	Computing Networks Mini-Game	1 - 20 participants	Applied several times in different schools
84	Programming Fundamentals Mini-Game	1 - 20 participants	Applied several times in different schools
85	PDConsole	1 - 20 participants	Applied more than one time in same context
86	Crossword Puzzles	NI	NI
87	Jeopardy!® Game	NI	NI
88	BattleThreads Game	NI	NI
89	Process State Transition Game	NI	NI
90	ERPsim	> 120 participants	Applied several times in same context
91	Security Protocol Game	61 - 80 participants	NI
92	U-Test	21 - 40 participants	Applied in two courses in same context
93	Operating systems role plays	NI	NI

94	Dealing with difficult people	41 - 60 participants	Applied in two courses in same context
95	Starter MMO	1 - 20 participants	NI
96	AMEISE	1 - 20 participants	NI
97	C-Jump	1 - 20 participants	NI
98	BINX	1 - 20 participants	NI
99	Anti-Phishing Phil	1 - 20 participants	NI
100	Lego Factory	1 - 20 participants	NI
101	Control-Alt-Hack	41 - 60 participants	NI
102	3DAR Lego Game	61 - 80 participants	NI
103	Database concurrency control card game	NI	NI
104	Digital System Game	1 - 20 participants	3 studies in same context
105	Mystery of Traffic Lights	21 - 40 participants	NI
106	Project Execution Game (PEG)	> 120 participants	Two studies in same context

Data analysis methods used in the evaluations

ID	Game	Data Analysis methods	
		Descriptive statistics	**Hypothesis testing**
1	Wu's Castle	Mean; Histogram; Line chart	
2	X-MED	Mean, Median, SD, Histogram	Mann-Whitney U
3	SimSE	Mean; Histogram; Box Plot; Scatter Plot	ANOVA
		Mean; Histogram; Scatter plot; Pearson correlation coefficient	
		Mean; Pearson correlation coefficient	

			Qualitative analysis	
4	EleMental: The recurrence		Qualitative analysis	
5	DELIVER!		Median, Frequency diagram	
6	Saving Sera	Qualitative analysis	Mean; SD	
7	The Catacombs	Qualitative analysis	Mean; SD	
8	The Risk Management Game	Qualitative analysis	Mean; Line chart	
9	Program your robot		Informal	
10	SortingGame	Qualitative analysis	Mean	
11	SortingCasino	Qualitative analysis	Mean	
12	Secret rule	Qualitative analysis	Mean	
13	Draw and guess	Qualitative analysis	Mean	
14	SCRUMIA		Median, Mean, Frequency diagram, Histogram	
15	Software engineering education game in Second Life		Qualitative analysis	
16	BOTS		Mean; SD	
17	SpITKom Project	Qualitative analysis	Histogram	
18	Prog&Play	Qualitative analysis	Mean; Median; Histogram	
19	Light-Bot	Qualitative analysis	Histogram	
20	Software Development Manager (SDM)		Qualitative analysis	
21	PM Master		Histogram, Frequency diagram	
22	Requirements Collection and Analysis Game (RCAG)		Mean; SD	Mann-Whitney U; Kruskal–Wallis test; Wilcoxon matched-pairs
23	Innov8		Qualitative analysis	
24	Kernel Panic		Qualitative analysis	
25	I-SEE		Informal	
26	hACMEgame	Qualitative analysis	Mean; Line chart	
27	Evaluators		Mean; Histogram	Mann-Whitney
28	Simsoft	Qualitative analysis	Mean; Histogram	

29	The Requirements Island	Informal		
30	Motherboard-assembly pairing game	Qualitative analysis	Mean; SD; Pearson correlation coefficient	ANOVA; T-test
31	Describing computer parts	Qualitative analysis	Mean; SD; Pearson correlation coefficient	ANOVA; T-test
32	Hands-on OS game	Qualitative analysis	Mean; SD; Pearson correlation coefficient	ANOVA; T-test
33	Bear-cub's computer game	Qualitative analysis	Mean; SD; Pearson correlation coefficient	ANOVA; T-test
34	Maze	Qualitative analysis	Mean	
35	Coloring Map	Qualitative analysis	Mean	
36	Scrum LEGO Challenge	Qualitative analysis	Mean; Box plot	
37	Star Chef		Mean; SD	ANOVA; F-test
38	Object Access Tool (OAT)	Qualitative analysis		
39	DesigMPS		Mean, SD, Median, Quartile, Box plot	Mann-Whitney U, T-test
40	Data Defender	Qualitative analysis	Mean	Z-test
41	Ztech	Qualitative analysis		
42	InspectorX		Mean; Mode; Quartile; SD; Histogram	
43	SPARSE	Qualitative analysis	Pie chart; Histogram	
44	SORTIA	Qualitative analysis	Histogram; Frequency diagram	
45	InspSoft	Qualitative analysis	Histogram; Frequency diagram	
46	SimMS	Qualitative analysis	Frequency diagram	
47	UbiRE	Qualitative analysis	Mean	
48	CyberCIEGE		Mean; SD; Histogram	T-test
49	ProGames	Qualitative analysis	Line chart	
50	IT Billionare Game Board	Qualitative analysis	Mean; Pie chart	
51	CAPTAIN3		Mean; Line chart	T-test

52	The Last Java Code	Qualitative analysis	Mean; SD	T-test
53	Tower-Defense	Qualitative analysis		
54	The aMAZEing Labyrinth	Qualitative analysis		
55	Quoridor	Qualitative analysis		
56	San Francisco Cable Cars	Qualitative analysis		
57	NeverWinter Nights 2		Mean; SD; Pearson correlation coefficient	Z-test; Mann-Whitney U
58	Problems and Programmers	Qualitative analysis	Mean	
59	SESAM	Qualitative analysis		
		Qualitative analysis	Histogram; Line chart	
		Qualitative analysis		
60	Open Software Solution (OSS)	Qualitative analysis	Mean; Box plot	
61	SimjavaSP	Qualitative analysis	Histogram	
62	Incredible Manager	Qualitative analysis		
		Qualitative analysis		
		Qualitative analysis	Mean	
63	Requirements Game	Qualitative analysis		
64	Lecture Quiz	Qualitative analysis	Mean	
65	PlayScrum	Qualitative analysis	Mean	
66	Management in Software Companies: a Classroom Game	Qualitative analysis		
67	Age of Computers		Mean; SD; Scatter plot	F-test; ANOVA
		Qualitative analysis	Line chart; Histogram	
68	Software project simulator	Qualitative analysis		
69	System Dynamics (SD) simulation		Mean; Median; SD; Range	ANCOVA; T-test
			Mean; Median; SD; Range	ANCOVA; T-test
			Mean; Median; SD; Range	ANCOVA; T-test; Z-test
		Qualitative analysis	Mean; Median; SD	T-test
70	Groupthink exercise in Second Life	Qualitative analysis		
71	MO-SEProcess	Qualitative analysis		
72	SE•RPG		Mean; Histogram	T-test

73	The MIS Game	Informal		
74	Information Systems Project Manager Game - ISPM Game	Informal		
75	Paths	Qualitative analysis		
76	Tower of Cubes	Qualitative analysis		
77	Software Development Game	Qualitative analysis		
78	DesigMPS	Qualitative analysis		
79	CounterMeasures	Qualitative analysis	Histogram; Box plot	
80	Algorithms Recursive Game	Qualitative analysis	Histogram	
81	Detective Game – what killed the project?	Qualitative analysis	Mean; Frequency diagram	
82	Computer Architecture Mini-Game	Qualitative analysis		
83	Computing Networks Mini-Game	Qualitative analysis		
84	Programming Fundamentals Mini-Game	Qualitative analysis		
85	PDConsole	Informal		
86	Crossword Puzzles	Informal		
87	Jeopardy!® Game	Informal		
88	BattleThreads Game	Informal		
89	Process State Transition Game	Informal		
90	ERPsim	Informal		
91	Security Protocol Game	Qualitative analysis		
92	U-Test	Qualitative analysis		Mann-Whitney U
93	Operating systems role plays	Informal		
94	Dealing with difficult people	Qualitative analysis	Frequency diagram	
95	Starter MMO	Qualitative analysis	Mean; Histogram	
96	AMEISE	Qualitative analysis		
97	C-Jump	Qualitative analysis	Mean; SD	
98	BINX	Qualitative analysis		
99	Anti-Phishing Phil	Qualitative analysis	Mean; Histogram; Spearman rho	ANOVA; T-test; F-test
100	Lego Factory	Informal		

<table>
<tr><td>101</td><td>Control-Alt-Hack</td><td colspan="3">Informal</td></tr>
<tr><td>102</td><td>3DAR Lego Game</td><td colspan="3">Mean</td></tr>
<tr><td>103</td><td>Database concurrency control card game</td><td colspan="3">Informal</td></tr>
<tr><td>104</td><td>Digital System Game</td><td>Qualitative analysis</td><td colspan="2">Mean</td></tr>
<tr><td>105</td><td>Mystery of Traffic Lights</td><td colspan="3">Qualitative Analysis</td></tr>
<tr><td>106</td><td>Project Execution Game (PEG)</td><td>Qualitative analysis</td><td>Mean; Pearson correlation coefficient; Histogram</td><td>T-test</td></tr>
</table>

References of the selected studies of the state of the practice

Adamo-Villani, N., Haley-Hermiz, T., & Cutler, R. (2013). Using a Serious Game Approach to Teach 'Operator Precedence' to Introductory Programming Students. *Proc. of the 17th Int. Conf. on Information Visualization*, (pp.523-526). London, GB.

Baker, A., Navarro, E. O., & van der Hoek, A. (2003). Problems and Programmers: An Educational Software Engineering Card Game. *Proc. of the 25th Int. Conf. on Software Engineering*, (pp.614-619). Portland: Oregon.

Baker, A., Navarro, E. O., & van der Hoek, A. (2005). An experimental card game for teaching software engineering processes. *Journal of Systems and Software*, 75(1–2), 3-16.

Battistella P. & Gresse von Wangenheim, C. (2016). Games for teaching computing in higher education – a systematic review. IEEE Technology and Engineering Education, 9(1), 8-30.

Battistella, P., Gresse von Wangenheim, C. & Carvalho, O. P. (2013). Teach management teams in Disciplines of Software Project Management. Brazilian Magazine of Informatics in Education, 21(1), p. 16.

Batistella, P., Wangenheim, A. & Gresse von Wangenheim, C. (2012). A Game for Teaching Sorting Algorithms: A Case Study in a Datastructure Course. In Brazilian Symposium on Informatics in Education, Rio de Janeiro/RJ, Brazil (in portuguese).

Barnes, T., Powell, E., Chaffin, A., & Lipford, H. (2008). Game2Learn: improving the motivation of CS1 students. *Proc. of the 3rd Int. Conf. on Game Development in Computer Science Education*, (pp. 1-5). New York, USA.

Barros, M. O., Dantas, A. R., Veronese, G. O., & Werner, C. M. L. (2006). Model-driven Game Development: Experience and Model Enhancements in Software Project Management Education. *Software Process: Improvement and Practice*, 11(4), 411 – 421.

Bavota, G., Lucia, A., Fasano, F., Oliveto, R., & Zottoli, C. (2012). Teaching software engineering and software project management: an integrated and practical approach. *Proc. of the 34th Int. Conf. on Software Engineering*, (pp. 1155-1164). IEEE Press, Piscataway, NJ, USA.

Benitti, F. B. V. & Molléri, J. S. (2008). Use of an RPG in teaching management and software development process. *Proc. of Workshop on Education in Computing*, (pp.258-267). Belém, Brazil.

Bezakova, I., Heliotis, J. E., & Strout, S. P. (2013). Board game strategies in introductory computer science. *Proc. of the 44th ACM Technical Symposium on Computer Science Education*, (pp.17-22). ACM, New York, NY, USA.

Caulfield, C., Veal, D., & Maj, S. P. (2011). Teaching Software Engineering Project Management – A Novel Approach for Software Engineering Programs. *Modern Applied Science*, 5(5), 87-104.

Chaffin, A., Doran, K., Hicks, D., & Barnes, T. (2009). Experimental evaluation of teaching recursion in a video game. In Stephen N. Spencer (Ed.), *Proc. of the 2009 ACM SIGGRAPH Symposium on Video Games*, (pp. 79-86). ACM, New York, NY, USA.

Chaves, R.O., Gresse von Wangenheim, C., Furtado, J. C. C., Oliveira, S. R. B., Santos, A. & Favero, E. L. (2015). Experimental Evaluation of a Serious Game for Teaching Software Process Modeling, IEEE Transactions on Education, pp.1-8.

Chaves, R. O., Miranda, T. C., Tavares, E. M. C., Oliveira, S. R. B., & Favero, E. L. (2011). DESIGMPS: A game to support the teaching of software process quality models, based in conceptual maps. *Proc. of the Brazilian Congress of Engineering Education*. Blumenau, Brazil.

Collofello, J. S. (2000). University/industry collaboration in developing a simulation based software project management training course. *Proc. of the 13th Conf. on Software Engineering Education & Training, IEEE Computer Society*, (pp. 161-168). Washington, DC, USA.

Dantas, A. R., Barros, M. O., & Werner, C. M. L. (2004a). A Simulation-Based Game for Project Management Experiential Learning. *Proc. of the 6th Inter. Conf. on Software Engineering & Knowledge Engineering*. Banff, Canada.

Dantas, A. R., Barros, M. O., & Werner, C. M. L. (2004b). Experimental training with Simulation Game for Software Project Managers. *Proc. of the 28th Brazilian Symposium on Software Engineering*, (pp. 23-38). Brasília, Brazil.

Denning, T., Kohno, T., & Shostack, A. (2012). *Control-Alt-HackTM: A Card Game for Computer Security Outreach, Education, and Fun*. Department of Computer Science and Engineering University of Washington, Technical Report UW-CSE-12-07-01.

Do, T. V. & Lee, J. W. (2009). A Multiple-Level 3D-LEGO Game in Augmented Reality for Improving Spatial Ability. In Julie A. Jacko (Ed.), *Human-Computer Interaction, Part IV*, (pp. 296-303). Springer-Verlag, Berlin, Heidelberg.

Drappa, A. & Ludewig, J. (1999). Quantitative Modeling for the Interactive Simulation of Software Project. *Journal of Systems and Software*. 46, 113–122.

Drappa, A. & Ludewig, J. (2000). Simulation in Software Engineering Training. *Proc. of the 22nd Int. Conf. on Software Engineering*, (pp. 199-208). Limerick, Ireland.

Eagle, M., & Barnes, T. (2009). Experimental evaluation of an educational game for improved learning in introductory computing. *Proc. of the 40th ACM Technical Symposium on Computer Science Education*, (pp. 321-325).

Fernandes, J. M. & Sousa, S. M. (2010). PlayScrum - A Card Game to Learn the Scrum Agile Method. *Proc. of the 2nd Int. Conf. on Games and Virtual Worlds for Serious Applications*, (pp. 52-59). Braga, Portugal.

Fu, F., Su, R., & Yu, S. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, 52(1), 101-112.

Gouws, L. A., Bradshaw, K., & Wentworth, P. (2013). Computational thinking in educational activities: an evaluation of the educational game light-bot. *Proc. of the 18th ACM Conf. on Innovation and Technology in Computer Science Education*, (10-15). ACM, New York, NY, USA.

Graven, O. H., Hansen, H. A., & MacKinnon, L. (2009). A Computer Game Modelling Routing in Computer Networks as Abstract Learning Material in a Blended Learning Environment. *Journal of Emerging Technologies in Learning.* Supplement, Issue S3, p18.

Gresse von Wangenheim, C., Kochanski, D., & Savi, R. (2009). Systematic Review on evaluation of games for software engineering learning in Brazil. Software Engineering Education Forum. Fortaleza, Brazil (in portuguese).

Gresse von Wangenheim, C., Savi, R. & Borgatto, A. F. (2012). DELIVER! – An educational game for teaching Earned Value Management in computing courses, Information and Software Technology, 54(3), 286-298.

Gresse von Wangenheim, C., Savi, R. & Borgatto, A. F. (2013). SCRUMIA - An educational game for teaching SCRUM in computing courses. Journal of Systems and Software, 86(10), 2675-2687.

Gresse von Wangenheim, C. & Shull, F. (2009). To Game or Not to Game? Software, IEEE, 26(2), 92-94.

Gresse von Wangenheim, C., Thiry, M. & Kochanski, D. (2009). Empirical evaluation of an educational game on software measurement. Empirical Software Engineering. 14, 418–452.

Gresse von Wangenheim, C., Thiry, M., Kochanski, D., Steil, L. & Lino, J. (2009). Development of a game to software measurement teaching. In: SBQS – Brazilian Symposium on Software Quality, Ouro Preto/MG, Brazil.

Hainey, T., Connolly, T. M., & Boyle, L. (2009). Development and evaluation of a game to tech requirements collection and analysis in software engineering at tertiary education level. *Proc. of the 3rd European Conf. on Games-based Learning*. Graz, Austria.

Hainey, T., Connolly, T. M., Stansfield, M., & Boyle, E. A. (2011). Evaluation of a game to teach requirements collection and analysis in software engineering at tertiary education level, *Computers & Education*, 56(1), 21-35.

Hakulinen, L. (2011). Using serious games in computer science education. *Proc. of the 11th Int. Conf. on Computing Education Research*, (pp. 83-88). ACM, New York, NY, USA.

Hamey, L. G. C. (2003). Using the Security Protocol Game to teach computer network security. *Proc. of Improving Learning Outcomes Through Flexible Science Teaching*, (pp. 96-101). Sydney, Australia.

Hatzilygeroudis, I., Grivokostopoulou, F., & Perikos, I. (2012). Using game-based learning in teaching CS algorithms. *Proc. of the IEEE Int. Conf. on Teaching, Assessment and Learning for Engineering*, (pp.H2C-9-H2C-12). Hong Kong, China.

Hicks, A. (2010). Towards social gaming methods for improving game-based computer science education. *Proc. of the 5th Int. Conf. on the Foundations of Digital Games*, (pp. 259-261)). ACM, New York, NY, USA.

Hijon-Neira, R., Velazquez-Iturbide, A., Pizarro-Romero, C., & Carrico, L. (2014). Serious games for motivating into programming. *Proc. of the IEEE Frontiers in Education Conf.*, (pp.1-8). Madrid, Spain.

Hill, J. M. D., Ray, C. K., Blair, J. R. S., & Carver Jr., C. A. (2003). Puzzles and games: addressing different learning styles in teaching operating systems concepts. *Proc. of the 34th Technical Symposium on Computer Science Education*, (pp. 182-186). ACM, New York, NY, USA.

Jordan, C., Knapp, M., Mitchell, D., Claypool, M., & Fisler, K. (2011). CounterMeasures: a game for teaching computer security. *Proc. of the 10th Annual Workshop on Network and Systems Support for Games*, Article 7, 6 pages. Piscataway, NJ, USA.

Kazimoglu, C., Kiernan, M., Bacon, L., & Mackinnon, L. (2012a). A Serious Game for Developing Computational Thinking and Learning Introductory Computer Programming. *Procedia - Social and Behavioral Sciences*, 47, 1991-1999.

Kazimoglu, C., Kiernan, M., Bacon, L., & Mackinnon, L. (2012b). Learning Programming at the Computational Thinking Level via Digital Game-Play. *Procedia Computer Science*, 9, 522-531.

Keller, J. (1987). Development and Use of the ARCS Model of motivational Design. *Journal of Instructional Development*, 10(3), 2-10.

Kern, V. M., Stotz, M. R., & Bento, M. (2008). A game-playing experience in the learning of database concurrency control. *Merlot*.

Kohwalter, T.C., Clua, E.W.G., & Murta, L.G.P. (2011). SDM - An Educational Game for Software Engineering, *Brazilian Symposium on Games and Digital Entertainment*, (pp.222-231). Salvador, Brazil.

Leger, P-M., Charland, P., Feldstein, H. D., Robert, J., Babin, G., & Lyle, D. (2011). Business Simulation Training in Information Technology Education: Guidelines for New Approaches in IT Training. *Journal of Information Technology Education*, 10, 39-53.

Leong, B., Koh, Z. H., & Razeen, A. (2011). *Teaching Introductory Programming as an Online Game*. Teaching Report. Singapore.

Leverington, M., Yuksel, M., & Robinson, M. (2009). Using role play for an upper level CS course. *Journal of Computing Sciences in Colleges*. 24(4), 259-266.

Lima, T., Campos, B., Santos, R. & Werner, C. (2012). UbiRE: A game for teaching requirements in the context of ubiquitous systems. *Proc. of the 38th Latin-American Conf. in Informatics*, (pp.1-10). Medellín, Colombia.

Liu, T., Chu, Y., & Tan, T. (2012). Using computer games in a computer course to improve learning, Teaching, *Proc. of the IEEE Int. Conf. on Assessment and Learning for Engineering*, (pp.W2C-16-W2C-19). Hong Kong, China.

Livovský, J. & Porubän, J. (2014). Learning object-oriented paradigm by playing computer games: concepts first approach. *Central European Journal of Computer Science*, 4(3), 171-182.

Lopes, A. C., Marques, A. B., & Conte, T. U. (2012). Evaluation of InspSoft game: a game for software inspection teaching. *Software Engineering Education Forum*. Natal, Brazil.

Mandl-Striegnitz, P. (2001). How to Successfully Use Software Project Simulation for Educating Software Project Managers. *Proc. of the Frontiers in Education Conference*, (pp. 19-24). Reno, NV, USA.

Martin, A. (2000). The design and evolution of a simulation/game for teaching information systems development. *Journal Simulation and Gaming*. 31(4), 445-463.

Melero, J., Hernández-Leo, D., & Blat, J. (2012). Considerations for the design of mini-games integrating hints for puzzle solving ICT-related concepts. *Proc. of the 12th IEEE Int. Conf. on Advanced Learning Technologies*, (pp. 154-158). Rome, Italy.

Mittermeir, R. T., Hochmüller, E., Bollin, A., Jäger, S., & Nusser, M. (2003). AMEISE – A Media Education Initiative for Software Engineering Concepts, the Environment and Initial Experiences. *Proc. of the Interactive Computer aided Learning*, (pp. 1-17). Villach, Austria.

Morsi, R. & Richards, C. (2012). BINX: An XNA/XBox 360 Educational Game for Electrical and Computer Engineers. *Consumer Electronics Times*, 1(3), 33-42.

MUMMS. (2015). *Measuring the Usability of Multi-Media Systems*. Available at: http://www.ucc.ie/hfrg/questionnaires/mumms/index.html. Access on: 31 Aug. 2015.

Muratet, M., Delozanne, E., Torguet, P., & Viallet, F. (2012). Serious game and students' learning motivation: effect of context using prog&play. In Stefano A. Cerri, William J. Clancey, Giorgos Papadourakis, & Kitty Panourgia (Eds.), *Proc. of the 11th Int. Conf. on Intelligent Tutoring Systems*, (pp. 123-128). Springer-Verlag, Berlin, Heidelberg.

Muratet, M., Torguet, P., Jessel, J-P., & Viallet, F. (2009). Towards a serious game to help students learn computer programming. *International Journal of Computer Games Technology - Game Technology for Training and Education*, Article 3, 12 pages.

Muratet, M., Torguet, P., Viallet, F., & Jessel, J-P. (2011). Experimental feedback on Prog & Play: a serious game for programming practice. *Computer Graphics Forum*, 30(1), 61-73.

Natvig, L. & Line, S. (2004). Age of computers: game-based teaching of computer fundamentals. *Proc. of the 9th Annual Conf. on Innovation and Technology in Computer Science Education*, (pp. 107-111). ACM, New York, NY, USA.

Natvig, L., Line, S. & Djupdal, A. (2004). "Age of computers"; an innovative combination of history and computer game elements for teaching computer fundamentals. *Proc. Of the 34th Annual Frontiers in Education*, (pp.S2F,1-6).

Navarro, E. O. (2006). *SimSE: A Software Engineering Simulation Environment for Software Process Education*. Unpublished Thesis. University of California, Irvine, USA.

Navarro, E. O., Baker, A., & van der Hoek, A. (2004). Teaching Software Engineering Using Simulation Games. *Proc. of the Int. Conf. on Simulation in Education*. San Diego, California, USA.

Navarro, E. O. & van der Hoek, A. (2005). Design and Evaluation of an Educational Software Process Simulation Environment and Associated Model. *Proc. of the 18th Conf. on Software Engineering Education and Training*. Ottawa, Canada.

Navarro, E. & van der Hoek, A. (2007). Comprehensive Evaluation of an Educational Software Engineering Simulation Environment. *Proc. of the 20th Conf. on Software Engineering Education & Training*, (pp.195-202). Dublin, Ireland.

Navarro, E. O. & van der Hoek, A. (2008). On the Role of Learning Theories in Furthering Software Engineering Education. In H. J. C. Ellis, S. A. Demurjian, & J. F. Naveda (Eds.), *Software Engineering: Effective Teaching and Learning Approaches and Practices*, (pp. 1-22). IGI Global.

Navarro, E. & van der Hoek, A. (2009). Multi-site evaluation of SimSE. *Proc. of the 40th ACM Technical Symposium on Computer Science Education*, (pp.326-330). New York, NY, USA.

Nerbraten, O. & Rostad, L. (2009). hACMEgame: A Tool for Teaching Software Security. *Proc. of the Int. Conf. on Availability, Reliability and Security*, (pp. 811-816). Fukuoaka, Japan.

Nunohiro, E., Matsushita, K., Mackin, K. J., & Ohshiro, M. (2013). Development of game-based learning features in programming learning support system. *Artificial Life and Robotics*, 17(3-4), 373-377.

Oania, M. C. (2011). *A serious game to teach secure coding in introductory programming: design, development and initial evaluation*. Master Thesis. Purdue University, 139 pages, West Lafayette, Indiana, USA.

Paasivaara, M., Heikkilä, V., Lassenius, C., & Toivola, T. (2014). Teaching students Scrum using LEGO blocks. *Proc. of the 36th Int. Conf. on Software Engineering*, (pp. 382-391). ACM, New York, NY, USA.

Pfahl, D., Ruhe, G., & Koval, N. (2001). An Experiment for Evaluating the Effectiveness of Using a System Dynamics Simulation Model in Software Project Management Education. *Proc. of the 7th Int. Symposium on Software Metrics*, (pp.97-109). London, GB.

Pfahl, D., Laitenberger, O., Dorsch, J., & Ruhe, G. (2003). An Externally Replicated Experiment for Evaluating the Learning Effectiveness of Using Simulations in Software Project Management Education. *Empirical Software Engineering*, 8(4), 367-395.

Pfahl, D., Laitenberger, O., Ruhe, G., Dorsch, J., & Krivobokova, T. (2004). Evaluating the learning effectiveness of using simulations in software project management education: results from a twice replicated experiment. *Information and Software Technology*, 46(2), 127-147.

Poels, K., Kort, Y. D., & Ijsselsteijn, W. (2007). It is always a lot of fun!: exploring imensions of digital game experience using focus group methodology. *Proc. of Conf. on Future Play*, (pp. 83-89). Toronto, Canada.

Potter, H., Schots, M., Duboc, L., & Werneck, V. (2014). InspectorX: A game for software inspection training and learning. *Proc. of the IEEE 27th Conf. on Software Engineering Education and Training*, (pp.55-64). Klagenfurt, Austria.

Raman, R., Lal, A., & Achuthan, K. (2014). Serious games based approach to cyber security concept learning: Indian context. *Proc. of Int. Conf. on Green Computing Communication and Electrical Engineering*, (pp.1-5). Coimbatore, Indian.

Rodriguez, D., Sicilia, M. A., Cuadrado-Gallego, J. J., & Pfahl, D. (2006). e-Learning in Project Management Using Simulation Models: A Case Study Based on the Replication of an Experiment. *IEEE Transactions on Education*, 49(4), 451 – 463.

Rodriguez-Cerezo, D., Sarasa-Cabezuelo, A., Gomez-Albarran, M., & Sierra, J-L. (2012). Facilitating comprehension of basic concepts in computer language implementation courses: A game-based approach. *Proc. of the Int. Symposium on Computers in Education*, (pp.1-6). Andorra la Vella, Andorra.

Rodriguez-Cerezo, D., Sarasa-Cabezuelo, A., Gomez-Albarran, M., & Sierra, J-L. (2014). Serious games in tertiary education: A case study concerning the comprehension of basic concepts in computer language implementation courses. *Computers in Human Behavior*, 31, 558-570.

Roodt, S. & Joubert, P. (2009). Evaluating Serious Games in Higher Education: A Theory-based Evaluation of IBMs Innov8. *Proc. of the 3rd European Conf. on Games-based learning*, (pp. 332-338). Graz, Austria.

Rossiou, E. & Papadakis, S. (2007). Educational Games in Higher Education: a case study in teaching recursive algorithms. *Proc. of the Int. Conf. Education in a Changing Environment*, (pp. 149-157). Salford, UK.

Rusu, A., Russell, R., Burns, E. & Fabian, A. (2011). Employing software maintenance techniques via a tower-defense serious computer game. In Maiga Chang, Wu-Yuin Hwang, Ming-Puu Chen, & Wolfgang Müller (Eds.), *Proc. of the 6th Int. Conf. on E-learning and Games, Edutainment Technologies*, (pp. 176-184). Springer-Verlag, Berlin, Heidelberg.

Ryoo, J., Techatassanasoontorn, A. A., Lee, D., & Lothian, J. (2011). Game-Based InfoSec Education Using OpenSim. *Proc. of the Colloquium for Information Systems Security Education*. Fairborn, Ohio.

Savi, R., Gresse von Wangenheim, C. & Borgatto, A. F. (2011). A Model for the Evaluation of Educational Games for Teaching Software Engineering. In 25th Brazilian Symposium on Software Engineering (SBES), São Paulo/SP, Brazil, pp.194-203 (in portuguese).

Schmitz, B., Czauderna, A., Klemke, R., & Specht, M. (2011). Game based learning for computer science education. *Proc. of the Conference on Computer Science Education Research*, (pp. 81-86). Heerlen, Netherlands.

Seng, W. Y. & Yatim, M. H. M. (2014). Computer Game as Learning and Teaching Tool for Object Oriented Programming in Higher Education Institution. *Procedia - Social and Behavioral Sciences*, 123, 215-224.

Sharp, H. & Hall, P. (2000). An Interactive Multimedia Software House Simulation for Postgraduate Software Engineers. *Proc. of the 22nd Int. Conf. on Software Engineering*, (pp. 1-12). Limerick, Ireland.

Shaw, K. & Dermoudy, J. (2005). Engendering an Empathy for Software Engineering. *Proc. of the 7th Australasian Conf. on Computing Education*, (pp. 135-144). Newcastle, New South Wales, Australia.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L., Hong, J., & Nunge, E. (2007). Anti-Phishing Phil: The Design and Evaluation of a Game that Teaches People Not to Fall for Phish. *Proc. of the Symposium on Usable Privacy and Security*, (pp. 88-99). Pittsburgh/PA, USA.

Simão, D. D., Correa, D. F., & Parreira Júnior, P. A. (2014). SimMS - Um Jogo Educacional de apoio ao Ensino de Manutenção de Software. *Proc. of the 2nd Workshop on Software Visualization, Evolution and Maintenance*, (pp.70-77). Maceió, AL, Brazil.

Sindre, G., Natvig, L., & Jahre, M. (2009). Experimental validation of the learning effect for a pedagogical game on computer fundamentals. *IEEE Transactions on Education*, 52(1), 10–18.

Singh, J., Dorairaj, S. K. & Woods, P. (2007). *Learning Computer Programming Using a Board Game - Case Study On C-Jump*. Masters thesis, Multimedia University, Cyberjaya, Malaysia.

Soares, G. M. & Rausis, B. Z. (2011). *Development of an educational game for teaching project management in undergraduate computing programs*. Project thesis, Information Systems Program, Federal University of Santa Catarina, Brazil (in portuguese).

Soflano, M., Connolly, T. M., & Hainey, T. (2015a). An application of adaptive games-based learning based on learning style to teach SQL. *Computers & Education*, 86, 192-211.

Soflano, M., Connolly, T. M., & Hainey, T. (2015b). Learning style analysis in adaptive GBL application to teach SQL. *Computers & Education*, 86, 105-119.

Solingen, R., Dullemond, K., & Gameren, B. (2011). Evaluating the Effectiveness of Board Game Usage to Teach GSE Dynamics, *Proc. of the 6th IEEE Int. Conf. on Global Software Engineering*, (pp.166-175). Helsinki, Finland.

Souza, M. M., Resende R. F., Prado, L. S., Fonseca, E. F., Carvalho, F. A., & Rodrigues, A. D. (2010). SPARSE - A Teaching and Learning Environment of Software Engineering Based on Games and Simulation. *Proc. of the 21st Brazilian Symposium on Informatics in Education*. João Pessoa, PB, Brazil.

Srinivasan, V., Butler-Purry, K., & Pedersen, S. (2008). Using Video Games to Enhance Learning in Digital Systems. *FuturePlay*, Toronto, Ontario, Canada.

Sweetser, P. & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment*, 3(3), 1-24.

Syberfeldt, A. & Syberfeldt, S. (2010). A serious game for understanding artificial intelligence in production optimization. *Proc. of the IEEE Symposium on Computational Intelligence and* Games, (pp.443-449). Dublin, IR.

Takatalo, J., Häkkinen, J., Kaistinen, J., & Nyman, G. (2010). Presence, Involvement, and Flow in Digital Games. In: Bernhaupt, R. (Ed.). *Evaluating User Experience in Games: Concepts and Methods*, (pp. 23-46). Springer.

Tan, B. & Seng, J. L. K. (2010). Game-based Learning for Data Structures: A case study, *Proc. of the 2nd Int. Conf. on Computer Engineering and Technology*, (pp.V6-718-V6-721). Singapore, Singapore.

Tang, Y., Shetty, S., & Chen, X. (2012). Educational Effectiveness of Virtual Reality Games Promoting Metacognition and Problem-Solving. *Proc. of the Annual Conf. of American Society for Engineering Education*. San Antonio, USA.

Taran, G. (2007). Using Games in Software Engineering Education to Teach Risk Management. *Proc. of the 20th Conf. Software Engineering Education & Training*, (pp.211-220). Dublin, IR.

Thiry, M. & Gonçalvez, R. Q. (2010) Development of a game to support the teaching of requirements engineering: the requirements island. *Proc. of the Brazilian Symposium on Games and Digital Entertainment*, (pp. 358-361). Florianópolis, SC, Brazil.

Thiry, M., Zoucas, A., & Silva, A. C. (2011). Empirical study upon software testing learning with support from educational game. *Proc. of the 23rd Int. Conf. on Software Engineering & Knowledge Engineering*, (pp.482-484). Miami Beach, USA.

Wang, A. I., Fsdahl, T., & Morch-Storstein, O. K. (2008). An Evaluation of a Mobile Game Concept for Lectures. *Proc. of the 21st Conf. on Software Engineering Education and Training*, (pp. 197-204). Charleston, SC.

Wang, T. & Zhu, Q. (2009). A Software Engineering Education Game in a 3-D Online Virtual Environment. *Proc. of the 1st Int. Workshop on Education Technology and Computer Science*, 2, (pp. 708-710). Wuhan, Hubei.

Wein, J., Kourtchikov, K., Cheng, Y., Gutierez, R., Khmelichek, R., Topol, M., & Sherman, C. (2009). Virtualized games for teaching about distributed systems. *SIGCSE Bull*, 41(1), 246-250.

Ye, E., Liu, C., & Polack-Wahl, J. A. (2007). Enhancing software engineering education using teaching aids in 3-D online virtual worlds. *Proc. of the 37th Annual Conf. on Frontiers in Education, IEEE Computer Society*, (pp. T1E-8-T1E-13). Milwaukee, WI.

Zapata, C. (2009). Teaching Software development by means of a classroom game: the software development game. *Developments in Business Simulation and Experiential Learning*, 36, 156-164.

Zapata, C. M. (2010). A Classroom Game for Teaching Management of Software Companies. *Dyna*. 77(163), 290–299.

Zapata, C. M. & Awad-Aubad, G. (2007). Requirements Game: Teaching Software Project Management. *CLEI Electronic Journal*, 10(1), paper 3.

Zhang, J., Smith, E., Caldwell, E. R., & Perkins, M. (2014). Learning and practicing decision structures in a game. *Journal of Computing Sciences in Colleges*, 29(4), 60-67.

Zwikael, O. & Gonen, A. (2007). Project execution game (PEG): training towards managing unexpected events. *Journal of European Industrial Training*, 31(6), 495-512

APPENDIX C – Spearman correlation matrix of the initial version of the MEEGA model

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
	Motivation										User Experience																Learning		
	Attention			Relevance			Confidence		Satisfaction		Immersion			Social Interaction			Challenge		Fun				Competence		Digital Game		Short-term		Long-term
1	1.00																												
2	**0.39**	1.00																											
3	**0.38**	**0.46**	1.00																										
4	0.23	0.26	**0.32**	1.00																									
5	0.29	0.27	**0.35**	**0.36**	1.00																								
6	0.14	0.23	0.22	0.29	0.29	1.00																							
7	0.20	0.16	0.15	0.18	0.22	0.28	1.00																						
8	0.21	0.27	**0.38**	0.27	**0.44**	0.24	0.28	1.00																					
9	0.26	**0.30**	**0.39**	**0.40**	**0.38**	**0.30**	0.20	**0.48**	1.00																				
10	0.17	0.25	0.26	0.23	0.21	0.25	0.17	0.26	**0.35**	1.00																			
11	0.25	**0.35**	**0.36**	0.21	0.22	0.15	0.13	0.24	0.24	0.28	1.00																		
12	0.30	**0.37**	**0.42**	0.22	0.23	0.18	0.12	0.23	0.27	0.23	**0.68**	1.00																	
13	0.27	**0.38**	**0.39**	0.20	0.28	0.19	0.16	0.28	0.27	0.24	**0.62**	**0.67**	1.00																
14	0.18	0.25	**0.31**	0.19	0.19	0.13	0.15	0.20	0.28	0.20	0.26	0.28	0.27	1.00															
15	0.26	**0.33**	**0.37**	0.20	0.22	0.15	0.17	0.23	**0.33**	0.21	**0.40**	**0.42**	**0.40**	**0.64**	1.00														
16	0.24	0.27	**0.31**	0.24	0.22	0.14	0.20	0.21	**0.32**	0.25	**0.33**	**0.35**	**0.31**	**0.54**	**0.56**	1.00													
17	0.25	**0.36**	**0.30**	0.27	0.28	0.22	0.15	**0.31**	**0.32**	0.30	**0.30**	**0.32**	**0.31**	0.24	**0.30**	**0.33**	1.00												
18	0.29	**0.39**	**0.45**	0.28	**0.31**	0.18	0.15	**0.35**	**0.35**	0.23	**0.42**	**0.45**	**0.46**	0.29	**0.42**	**0.35**	**0.47**	1.00											
19	**0.35**	**0.44**	**0.45**	**0.32**	**0.34**	0.26	0.25	**0.31**	**0.37**	0.20	**0.45**	**0.52**	**0.46**	**0.34**	**0.56**	**0.42**	**0.39**	**0.57**	1.00										
20	0.25	**0.35**	**0.30**	0.24	0.26	0.22	0.17	**0.30**	0.29	0.20	**0.39**	**0.46**	**0.47**	0.16	**0.33**	0.21	**0.33**	**0.40**	**0.47**	1.00									
21	**0.36**	**0.40**	**0.40**	**0.41**	**0.43**	0.23	0.19	**0.42**	**0.41**	0.23	**0.38**	**0.41**	**0.41**	0.20	**0.34**	0.27	**0.40**	**0.47**	**0.52**	**0.48**	1.00								
22	**0.33**	**0.39**	**0.36**	**0.36**	**0.37**	0.20	0.19	**0.34**	**0.38**	0.22	**0.38**	**0.46**	**0.42**	0.18	**0.32**	0.27	**0.37**	**0.43**	**0.49**	**0.50**	**0.68**	1.00							
23	0.20	0.25	0.26	0.26	**0.31**	0.29	0.24	**0.38**	**0.31**	**0.44**	0.26	0.28	**0.32**	0.14	0.18	0.19	**0.27**	**0.31**	**0.32**	0.26	**0.32**	**0.31**	1.00						
24	0.29	**0.35**	**0.39**	**0.35**	**0.42**	0.28	0.22	**0.48**	**0.37**	0.28	**0.33**	**0.34**	**0.35**	0.19	0.29	0.28	**0.38**	**0.40**	**0.45**	**0.36**	**0.43**	**0.42**	**0.49**	1.00					
25	0.08	-0.08	-0.10	-0.08	-0.11	-0.01	0.07	-0.15	-0.07	-0.01	-0.06	-0.04	-0.05	-0.18	-0.08	-0.09	-0.12	-0.15	-0.03	0.08	-0.05	-0.01	-0.02	-0.06	1.00				
26	0.06	-0.08	-0.01	-0.05	-0.09	-0.04	0.11	-0.18	-0.11	0.01	-0.07	-0.04	-0.06	-0.17	-0.08	-0.06	-0.10	-0.14	-0.11	0.07	-0.06	-0.01	-0.03	-0.09	**0.79**	1.00			
27	0.18	0.20	**0.32**	**0.34**	**0.41**	0.19	0.15	**0.44**	**0.40**	0.17	0.16	0.19	0.18	0.16	0.20	0.22	0.28	0.27	0.28	0.20	**0.36**	**0.34**	**0.31**	**0.41**	0.10	-0.12	1.00		
28	0.16	0.17	0.29	0.22	**0.38**	0.18	0.12	**0.39**	**0.34**	0.19	0.17	0.22	0.24	0.16	0.19	0.16	0.25	0.26	0.24	0.21	**0.34**	**0.34**	0.27	**0.35**	-0.11	-0.13	**0.64**	1.00	
29	0.21	**0.30**	**0.30**	**0.38**	**0.34**	0.19	0.15	**0.35**	**0.45**	0.20	0.30	**0.30**	**0.31**	0.17	0.27	0.25	0.30	**0.35**	0.34	0.28	**0.43**	**0.43**	0.22	**0.36**	-0.10	-0.12	**0.49**	**0.45**	1.00

APPENDIX D – Comparison between the initial version of the MEEGA model and the MEEGA+ model

MEEGA model (SAVI, 2011)			MEEGA+ model		
Quality factor	**Dimension**	**Item description**	**Quality factor**	**Dimension**	**Item description**
Motivation	Attention	The game design is attractive.	Usability	Aesthetics	The game design is attractive (interface, graphics, cards, boards, etc.).
		There was something interesting at the beginning of the game that captured my attention.	Player experience	Focused Attention	There was something interesting at the beginning of the game that captured my attention.
		The variation (form, content or activities) helped me to keep attention to the game.	--		
	Relevance	The game content is relevant to my interests.	Player experience	Relevance	The game contents are relevant to my interests.
		The way the game works suits my way of learning.	--		
		The game content is connected to other knowledge I already had.	--		
	Confidence	It was easy to understand the game and start using it as study material.	--		
		Passing through the game, I felt confident that I was learning.	--		
	Satisfaction	I am satisfied because I know I will have opportunities to use in practice things I learned playing this game.	--		
		It is due to my personal effort that I manage to advance in the game.	Player experience	Satisfaction	It is due to my personal effort that I managed to advance in the game.

User Experience	Immersion	Temporarily I forgot about my daily; I have been fully concentrated on the game.	Player experience	Focused Attention	I forgot about my immediate surroundings while playing this game.
		I did not notice the time pass while playing; when I saw the game had already ended.	Player experience	Focused Attention	I was so involved in my gaming task that I lost track of time.
		I felt myself more in the game context than real life, forgetting what was around me.	--		
	Social Interaction	I was able to interact with others during the game.	Player experience	Social Interaction	I was able to interact with other players during the game.
		I had fun with other people.	--		
		The game promotes cooperation and/or competition among the players.	Player experience	Social Interaction	The game promotes cooperation and/or competition among the players.
	Challenge	This game is appropriately challenging for me, the tasks are not too easy nor too difficult.	--		
		The game progresses at an adequate pace and does not become monotonous - offers new obstacles, situations or variations in its tasks.	Player experience	Challenge	The game does not become monotonous as it progresses (repetitive or boring tasks).
			Player experience	Challenge	The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace.
	Fun	I had fun with the game.	Player experience	Fun	I had fun with the game.
		When interrupted at the end of the class, I was disappointed that the game was over.	--		
		I would recommend this game to my colleagues.	Player experience	Satisfaction	I would recommend this game to my colleagues.
		I would like to play this game again.	--		

	Competence	I achieved the goals of the game applying my knowledge.	--		
		I had positive feelings on the efficiency of this game.	--		
	Control	The controls to perform actions in the game responded well.	--		
		It's easy to learn how to use the interface and game controls.	Usability	Learnability	Learning to play this game was easy for me.
Learning	Short-term learning	The game contributed to my learning in this course.	Player experience	Perceived Learning	The game contributed to my learning in this course.
		The game was efficient for my learning, comparing it with other activities of the course.	Player experience	Perceived Learning	The game allowed for efficient learning compared with other activities in the course.
	Long-term learning	The experience with the game will contribute to my professional performance in practice.	--		
			Usability	Aesthetics	The text font and colors are well blended and consistent.
				Learnability	I needed to learn a few things before I could play the game.
					I think that most people would learn to play this game very quickly.
				Operability	I think that the game is easy to play.
					The game rules are clear and easy to understand.
				Accessibility	The fonts (size and style) used in the game are easy to read.
					The colors used in the game are meaningful.

Player experience	Confidence	The contents and structure helped me to become confident that I would learn with this game.
	Challenge	This game is appropriately challenging for me.
	Satisfaction	Completing the game tasks gave me a satisfying feeling of accomplishment.
		I feel satisfied with the things that I learned from the game.
	Social Interaction	I felt good interacting with other players during the game.
	Fun	Something happened during the game (game elements, competition, etc.) which made me smile.
	Relevance	It is clear to me how the contents of the game are related to the course.
		This game is an adequate teaching method for this course.
		I prefer learning with this game to learning through other ways (e.g. other teaching methods).
	Perceived Learning	The game contributed to <*verb related to the level of the learning goal (cognitive, psychomotor, and affective)*> <*goal/concept*>.

APPENDIX E – Definition of the work products (outputs) of the MEEGA+ Process

Phase 1. Scoping

Output: P1.1.1 – Educational game

Game's name	Indicates the game's name.
Platform	(Digital or non-digital)
Computing knowledge area	Indicates the computing knowledge area following the ACM and IEEE curriculum guidelines for undergraduate degree programs in Computer Science (ACM; IEEE-CS, 2013). The Knowledge Areas are (ACM; IEEE-CS, 2013): AL - Algorithms and Complexity AR - Architecture and Organization CN - Computational Science DS - Discrete Structures GV - Graphics and Visualization HCI - Human-Computer Interaction IAS - Information Assurance and Security IM - Information Management IS - Intelligent Systems NC - Networking and Communications OS - Operating Systems PBD - Platform-based Development PD - Parallel and Distributed Computing PL - Programming Languages SDF - Software Development Fundamentals SE - Software Engineering SF - Systems Fundamentals SP - Social Issues and Professional Practice
Time	Indicates the duration of a game session in minutes.
Learning objectives	Indicates the learning objectives of the selected game.

Output: P1.2.2 - Evaluation goal

Evaluation goal	Analyse the <*name of the selected game*> for the purpose of *evaluate the quality* in terms of *usability and player experience* from the *students'* point of view in the context of *higher computing education*.

Phase 2. Planning

Output: P2.1.1 – Evaluation context

Institution	Indicates the institution's name.
Program	Indicates the name of the program.
Course	Indicates the name of the course.
Population	Indicates the number of students enrolled in the course.

Output: P2.2.1 – Evaluation schedule

Date and hour	Indicates the date and hour that the evaluation will be conducted.
Room	Indicates the room/lab number and/or address that the evaluation will be conducted.

Output: P2.3.1 – MEEGA+ data collection instrument

The MEEGA+ questionnaire is available in English, Spanish, and Brazilian Portuguese at http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/.

Questionnaire for quality evaluation of educational games

Game: ____

Please, help us improve the game answering the following questions. All information is collected anonymously and will be used only in a summarized way in the context of this game evaluation.

Researcher: ____

Place and date: ____

Demographic Information	
Institution:	
Program:	
Course:	
Age group:	☐ Under 18 years ☐ 18 to 28 years ☐ 29 to 39 years ☐ 40 to 50 years ☐ Over 50 years
Gender:	☐ Male ☐ Female
How often do you play digital games?	☐ Never ☐ Rarely: from time to time ☐ Monthly: at least once a month ☐ Weekly: at least once a week ☐ Daily: every day.
How often do you play non-digital games (card or board games, etc.)?	☐ Never ☐ Rarely: from time to time ☐ Monthly: at least once a month

	☐ Weekly: at least once a week ☐ Daily: every day.

Please, **select an option** according to how much you agree or disagree with each statement below.

Usability					
Statements	**Select an option as your evaluation**				
	Strongly disagree	**Disagree**	**Neither disagree nor agree**	**Agree**	**Strongly agree**
The game design is attractive (interface, graphics, etc.).	☐	☐	☐	☐	☐
The text font and colors are well blended and consistent.	☐	☐	☐	☐	☐
I needed to learn a few things before I could play the game.	☐	☐	☐	☐	☐
Learning to play this game was easy for me.	☐	☐	☐	☐	☐
I think that most people would learn to play this game very quickly.	☐	☐	☐	☐	☐
I think that the game is easy to play.	☐	☐	☐	☐	☐
The game rules are clear and easy to understand.	☐	☐	☐	☐	☐
The fonts (size and style) used in the game are easy to read.	☐	☐	☐	☐	☐
The colors used in the game are meaningful.	☐	☐	☐	☐	☐

Player Experience					
Statements	**Select an option as your evaluation**				
	Strongly disagree	**Disagree**	**Neither disagree nor agree**	**Agree**	**Strongly agree**
The contents and structure helped me to become confident that I would learn with this game.	☐	☐	☐	☐	☐
This game is appropriately challenging for me.	☐	☐	☐	☐	☐
The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace.	☐	☐	☐	☐	☐
The game does not become monotonous as it progresses (repetitive or boring tasks).	☐	☐	☐	☐	☐
Completing the game tasks gave me a satisfying feeling of accomplishment.	☐	☐	☐	☐	☐
It is due to my personal effort that I managed to advance in the game.	☐	☐	☐	☐	☐
I feel satisfied with the things that I learned from the game.	☐	☐	☐	☐	☐
I would recommend this game to my colleagues.	☐	☐	☐	☐	☐
I was able to interact with other players during the game.	☐	☐	☐	☐	☐
The game promotes cooperation and/or competition among the players.	☐	☐	☐	☐	☐
I felt good interacting with other players during the game.	☐	☐	☐	☐	☐
I had fun with the game.	☐	☐	☐	☐	☐
Something happened during the game (game elements, competition, etc.) which made me smile.	☐	☐	☐	☐	☐

There was something interesting at the beginning of the game that captured my attention.	☐	☐	☐	☐	☐
I was so involved in my gaming task that I lost track of time.	☐	☐	☐	☐	☐
I forgot about my immediate surroundings while playing this game.	☐	☐	☐	☐	☐
The game contents are relevant to my interests.	☐	☐	☐	☐	☐
It is clear to me how the contents of the game are related to the course.	☐	☐	☐	☐	☐
This game is an adequate teaching method for this course.	☐	☐	☐	☐	☐
I prefer learning with this game to learning through other ways (e.g. other teaching methods).	☐	☐	☐	☐	☐
The game contributed to my learning in this course.	☐	☐	☐	☐	☐
The game allowed for efficient learning compared with other activities in the course	☐	☐	☐	☐	☐
This statement is repeated for each learning goal of the game. The game contributed to <*verb as level of the learning goal (cognitive, psychomotor, affective)*> <*goal/concept*>. An example in accordance with the learning goals of SCRUMIA (Gresse von Wangenheim et al., 2013): The game contributed to *recall* the concepts from *Sprint Planning*.	☐	☐	☐	☐	☐

Please list three strong aspects of the game:____
____
____
______

Please give three suggestions to improve the game:____
____
____
______

Any further comment?____
____
___

Thanks a lot for your contribution!

Output: P2.4.1 – Approval of the Ethics Committee (optional)

There is no a template for this output. This output is defined by the Ethics Committee of the institution that the evaluation will be conducted.

Phase 3. Execution

Output: P3.1.2 – Consent form printed (optional)

There is no a template for this output. A template for this output is typically provided by the Ethics Committee of the institution that the evaluation will be conducted.

Output: P3.2.1 – Participants' consent (optional)

A document presenting the formal consent signed by the participants, following the template provided in the Output: P3.1.2 – Consent form printed.

Output: P3.3.1 – Game executed

Duration:	Indicates the duration of the game execution
Execution:	Describes the execution of the game and any deviations from plan. If the game is a digital game, describes if technological problems occurred. If the game is a non-digital game, describes any problem with the room, for example, the tables available were small for the board used in the game.

Output: P3.4.1 – Data collected

All data collection instruments/electronic forms filled out by the participants (following the template of the output P2.3.1 – MEEGA+ model data collection instruments) and checked/validated by the researcher.

Sample size:	Number of students that participated of the evaluation and filled out the questionnaire in a valid way.

Phase 4. Analysis

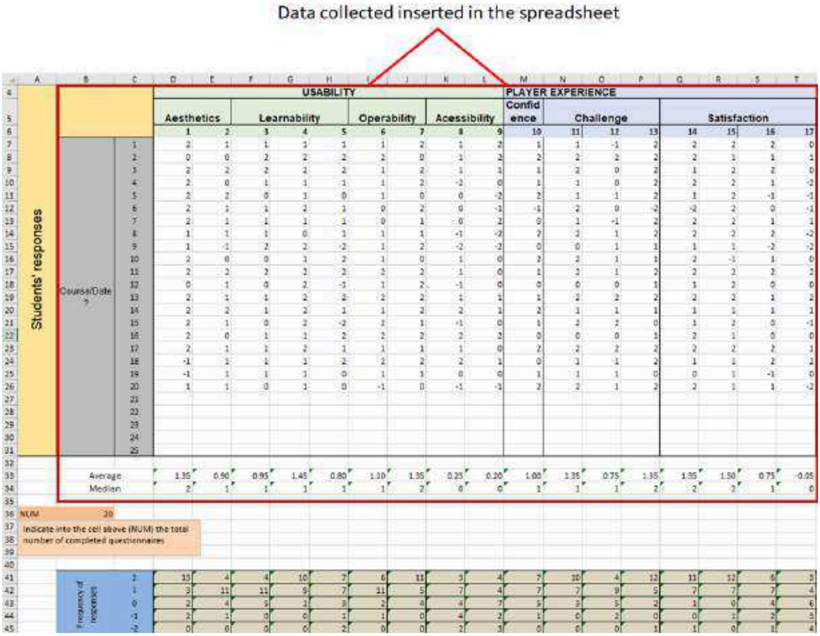
Output: P4.1.1 – MEEGA+ data analysis spreadsheets

The MEEGA+ data analysis spreadsheets are available in English, Spanish, and Brazilian Portuguese at http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/.

The MEEGA+ data analysis spreadsheet contains five internal spreadsheets (Data; Demographic data; Graphs, Demographic graphs; and Items).

The Data spreadsheet (Figure 69) aims to organize the data collected for a descriptive analysis. Each row of the spreadsheet represents an answer of one student for the game evaluation. The columns represent the MEEGA+ measurement instrument items. After inserting the data collected in the spreadsheet, the average, median and frequency of responses are automatically calculated.

Figure 69 - Data analysis spreadsheet (evaluation data)



Source: developed by the author.

The Demographic data spreadsheet (Figure 70) aims to organize the demographics data collected through the MEEGA+ measurement instrument, in order to identify the characteristics of the sample. The demographics information collected are age group, gender and how often students play digital and/or non-digital games.

Figure 70 - Data analysis spreadsheet (demographic data)

Students' responses			DEMOGRAPHIC INFORMATION			
			Age group: 1 - Under 18 years 2 - 18 to 28 years 3 - 29 to 39 years 4 - 40 to 50 years 5 - Over 50 years	Gender: M - Male F - Female	How often do you play digital games? 1 - Never 2 - Rarely: from time to time 3 - Monthly: at least once a month 4 - Weekly: at least once a week 5 - Daily: every day	How often do you play non-digital games (card or board games, etc.)? 1 - Never 2 - Rarely: from time to time 3 - Monthly: at least once a month 4 - Weekly: at least once a week 5 - Daily: every day
	INE 5617 - 31/03/2017	1	2	M	4	3
		2	2	M	4	3
		3	2	M	4	2
		4	2	M	5	3
		5	2	M	5	2
		6	2	M	5	2
		7	2	M	5	2
		8	2	F	4	2
		9	1	F	2	1
		10	1	M	2	2
		11	2	F	3	2
		12	3	M	2	2
		13	2	F	3	2
		14	2	M	4	3
		15	2	M	5	3
		16	2	M	2	2
		17	2	M	4	2
		18	2	M	4	3
		19	2	M	3	4
		20	2	M	5	4
		21	2	F	5	3
		22	2	M	5	4

Source: developed by the author.

The Items spreadsheet (Figure 71) is a standardized spreadsheet that contains the description of the MEEGA+ measurement instrument items used in the graphs. The customized items related to the learning goals of the evaluated game should be inserted in this spreadsheet in the corresponding dimension (Learning goals) to be used in the graphs.

Figure 71 - Items spreadsheet

Quality factor	Dimension	No. Item	Item Description
Usability	Aesthetics	1	The game design is attractive (board, cards, interface, graphics, etc.).
		2	The text font and colors are well blended and consistent.
	Learnability	3	I needed to learn a few things before I could play the game.
		4	Learning to play this game was easy for me.
		5	I think that most people would learn to play this game very quickly.
	Operability	6	I think that the game is easy to play.
		7	The game rules are clear and easy to understand.
	Accessibility	8	The fonts (size and style) used in the game are easy to read.
		9	The colors used in the game are meaningful.
Player experience	Confidence	10	The contents and structure helped me to become confident that I would learn with this game.
	Challenge	11	This game is appropriately challenging for me.
		12	The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace.
		13	The game does not become monotonous as it progresses (repetitive or boring tasks).
	Satisfaction	14	Completing the game tasks gave me a satisfying feeling of accomplishment.
		15	It is due to my personal effort that I managed to advance in the game.
		16	I feel satisfied with the things that I learned from the game.
		17	I would recommend this game to my colleagues.
	Social Interaction	18	I was able to interact with other players during the game.
		19	The game promotes cooperation and/or competition among the players.
		20	I felt good interacting with other players during the game.
	Fun	21	I had fun with the game.
		22	Something happened during the game (game elements, competition, etc.) which made me smile.
	Focused Attention	23	There was something interesting at the beginning of the game that captured my attention.
		24	I was so involved in my gaming task that I lost track of time.
		25	I forgot about my immediate surroundings while playing this game.
	Relevance	26	The game contents are relevant to my interests.
		27	It is clear to me how the contents of the game are related to the course.
		28	This game is an adequate teaching method for this course.
		29	I prefer learning with this game to learning through other ways (e.g. other teaching methods).
	Perceived Learning	30	The game contributed to my learning in this course.
		31	The game allowed for efficient learning compared with other activities in the course.
		32	The game contributed to *<verb as level of the learning goal (cognitive, psychomotor, affective)>* <goal/concept>.
		33	The game contributed to *<verb as level of the learning goal (cognitive, psychomotor, affective)>* <goal/concept>.
		34	...
		35	...

Source: developed by the author.

Output: P4.2.1 – Descriptive statistics results

Frequency of responses, average and median for each measurement instrument item (Figure 72).

Figure 72 - Data analysis spreadsheet (descriptive statistics)

Students' responses	USABILITY									PLAYER EXPERIENCE							
Course/Date ?	Aesthetics		Learnability			Operability		Acessibility		Confidence	Challenge			Satisfaction			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	2	1	1	1	1	1	2	1	2	1	1	-1	2	2	2	2	0
2	0	0	2	2	2	2	0	1	2	2	2	2	2	2	1	1	1
3	2	2	2	2	2	1	2	1	1	1	2	0	2	1	2	2	0
4	2	0	1	1	1	1	2	-2	0	1	1	0	2	2	2	1	-2
5	2	2	0	1	0	1	0	0	-2	2	1	1	2	1	2	-1	-1
6	2	1	1	2	1	0	2	0	-1	-1	2	0	-2	-2	2	0	-1
7	2	1	1	1	1	0	1	0	2	0	1	-1	2	2	2	1	1
8	1	1	1	0	1	1	1	-1	-2	2	2	1	2	2	2	2	-2
9	1	-1	2	2	-2	1	2	-2	-2	0	0	1	1	1	1	-2	-2
10	2	0	0	1	2	1	0	1	0	2	2	1	1	2	-1	1	0
11	2	2	2	2	2	2	2	1	0	1	2	1	2	2	2	2	2
12	0	1	0	2	-1	1	2	-1	0	0	0	0	1	1	2	0	0
13	2	1	1	2	2	2	2	1	1	1	2	2	2	2	2	1	2
14	2	2	1	2	1	1	2	2	1	2	1	1	1	1	1	1	1
15	2	1	0	2	-2	2	1	-1	0	1	2	2	0	1	2	0	-1
16	2	0	1	1	2	2	2	2	2	0	0	0	1	2	1	0	0
17	2	1	1	2	1	1	1	1	0	2	2	2	2	2	2	2	1
18	-1	1	1	1	2	2	2	2	1	0	1	1	2	1	1	2	2
19	-1	1	1	1	0	1	1	0	0	1	1	1	0	0	1	-1	0
20	1	1	0	1	0	-1	0	-1	-1	2	2	1	2	2	1	1	-2
21																	
22																	
23																	
24																	
25																	
Average	1.35	0.90	0.95	1.45	0.80	1.10	1.35	0.25	0.20	1.00	1.35	0.75	1.35	1.35	1.50	0.75	-0.05
Median	2	1	1	1	1	1	2	0	0	1	1	1	2	2	2	1	0

NUM 20

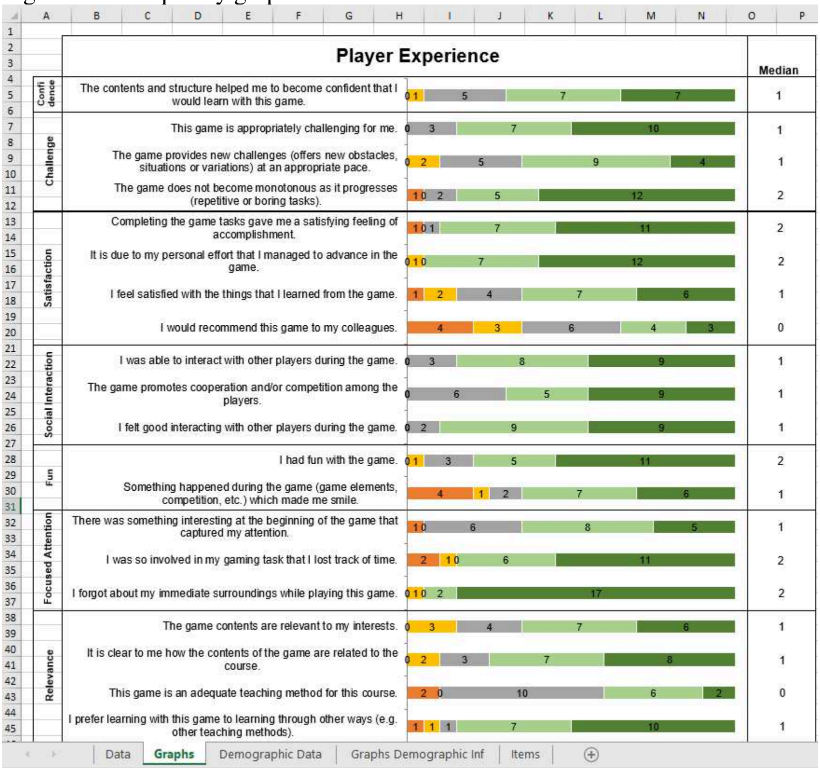
Indicate into the cell above (NUM) the total number of completed questionnaires

Frequency of responses	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	13	4	4	10	7	6	11	3	4	7	10	4	12	11	12	6	3
1	3	11	11	9	7	11	5	7	4	7	7	9	5	7	7	7	4
0	2	4	5	1	5	1	4	4	7	5	3	5	2	1	0	4	6
-1	2	1	0	0	1	1	0	4	2	1	0	2	0	0	1	2	3
-2	0	0	0	0	2	0	0	2	3	0	0	0	1	1	0	1	4

Source: developed by the author.

Graphics of frequency are presented in accordance with the quality factors and dimensions of the MEEGA+ model (Figure 73).

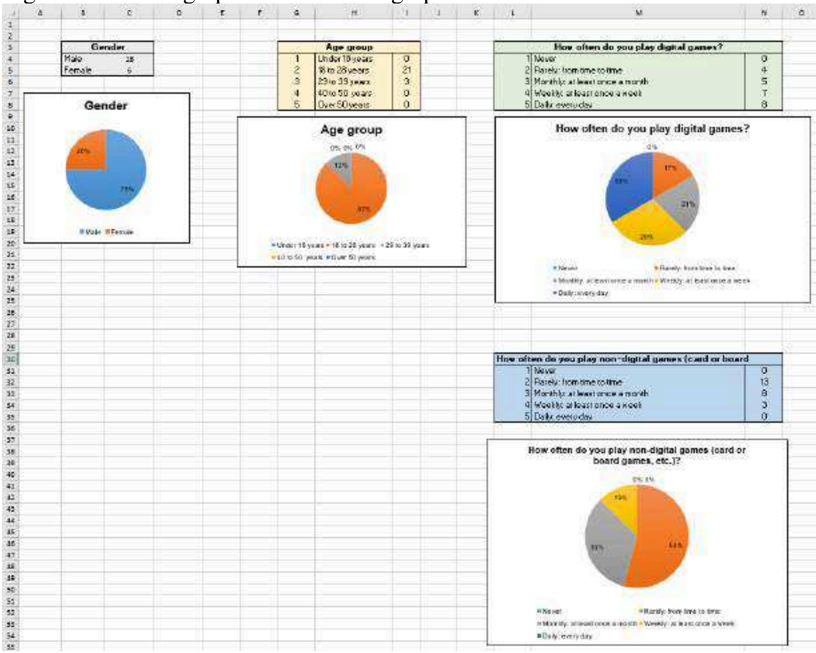
Figure 73 - Frequency graphs



Source: developed by the author.

Graphics presenting the demographic information are generated and presented in the Demographic graph spreadsheet (Figure 74).

Figure 74 - Demographic information graphs



Source: developed by the author.

Output: P4.3.2 – Game quality level

Game quality level:	Describes the characteristics of the evaluated game based on their quality level, as described in Table P.4.3.2.1.

Table P4.3.2.1. Game quality levels

Quality level	Level description
Low quality ($\theta < 42.5$)	At this level, the game rarely provides social interaction and hardly ever produces moments of fun among the players. The game does not capture the student's focused attention, does not arouse the confidence that he/she will learn from the game, nor does it produce feelings of satisfaction. The game rarely presents challenges, has monotonous tasks and does not contribute to student learning. Although a game at this level has a low relevance to the students' interests, a student recognizes that the game's content is related to the course. In terms of usability, a game at this level sometimes exhibits operability features, which may have some clear rules and be easy to play.
Good quality ($42.5 <= \theta < 65$)	At this level, the game sometimes presents challenging activities, offering new challenges for students. It provides moderately focused attention to the players, although students do not forget about their surroundings. Sometimes the game also provides feelings of confidence and satisfaction in the players. Frequently the game presents moments of social

		interaction and fun among the players. Often the game is considered relevant to the students' interests and, usually, the students recognize that the game's content is related to the course. Frequently the game contributes efficiently to student learning. In terms of usability, the game usually has the clear rules and is easy to play, although, usually does not present a fully attractive design.
	Excellent quality (θ >= 65)	At this level, the game is challenging for students and has no monotonous activities. It is highly relevant to students' interests and provides excellent focused attention, satisfaction, fun, and social interaction. It allows the student to be confident that he/she will learn from the game and contribute to and efficient student learning. In terms of usability, the game presents excellent operability and learnability, that is, it has clear rules and is easy to learn to play. Even so, a game at this level may present improvements in terms of aesthetics, not presenting a fully attractive design.

Output: P4.4.1 – Evaluation results

Analysis question 1: Does the <*name of the evaluated game*> have a good usability?	Analyses the main results for each dimension of the quality factor usability, answering the defined analysis question.
Analysis question 2: Does the <*name of the evaluated game*> provides a positive player experience?	Analyses the main results for each dimension of the quality factor player experience, answering the defined analysis question.
Analysis question 3: How old are the students that compose the sample of the study?	Analyses the frequency of responses to each age group.
Analysis question 4: What is the gender of the students that compose the sample of the study?	Analyses the frequency of responses to each gender.
Analysis question 5: What is the frequency that the students play digital and/or non-digital games?	Analyses the frequency of each response category indicating how often the students play digital and/or non-digital games.

Output: P4.5.1 – Discussion

Discussion:	Discussion of findings and, optionally, compared them with similar researches.
Threats to validity:	Indicates the limitations of the study and its threats to validity, as well as mitigation strategies adopted in order to minimize the impact in the study.

Phase 5. Presentation

Output: P5.1.1 – Evaluation report

Section	Contents/Scope	
Title	Evaluation of the <*object of study*> for the <*computing knowledge area*> education in <*evaluation context*>. E.g.: Evaluation of the Kahoot! PMQuiz for the Project Management education in Computing courses.	
Authorship	Names, contacts and affiliations.	
Abstract	Summarizes the paper under headings of background or context, objectives, method, main results, and conclusions.	
Introduction	Problem Statement	Indicates what the problem is; where is occurs, and who observe it.
	Research Objective	Defines the evaluation using the formalized style used in the MEEGA+ model.
	Context	Indicates environmental factors such as institution, course, and participants involved in the evaluation.
Related Work	How this research relates to existing research (studies)?	
Research Method	Reports the methods used in the research, such as the MEEGA+ method (PETRI; GRESSE VON WANGENHEIM, BORGATTO, 2018a), case studies (YIN, 2017; WOHLIN et al., 2012), GQM (BASILI et al., 1994).	
Evaluation Planning	Object of study	Indicates the game selected fort the evaluation.
	Evaluation goal	Presents the defined evaluation goal and the analysis questions following the MEEGA+ model. Analyse the <*name of the selected game*> for the purpose of *evaluate the quality* in terms of *usability and player experience* from the *students'* point of view in the context of *higher computing education*.
	Analysis questions	AQ1: Does the <*name of the evaluated game*> have a good usability? AQ2: Does the <*name of the evaluated game*> provide a positive player experience? AQ3: How old are the students that compose the sample of the study? AQ4: What is the gender of the students that compose the sample of the study? AQ5: What is the frequency that the students play digital and/or non-digital games?
	Context details	Indicates the place that the evaluation took place, such as institution and course.
	Research design	Indicates the research design applied, following the definition of the MEEGA+ model. Case study design (one-shot post-test only).
	Schedule	Indicates the schedule of the evaluation such as date and time.

	Number of the Ethics Committee approval	Indicates the number of the approval provided by the Ethics Committee (if necessary).
Execution	Sample	Description of the sample characteristics (demographic information).
	Preparation	What has been done to prepare the execution of the evaluation (i.e., schedule, materials)?
	Game applied	Indicates how game application took place and any deviations from plan.
	Data collection performed	How data collection took place and any deviations from plan.
Analysis	Answer the analysis questions	Summarizes the data collected and describes how it was analysed and answers each of the analysis questions defined.
	Game quality level	Indicates the quality level of the evaluated game, obtained from the MEEGA+ scale.
Discussion	Evaluation of results	Interprets and explains the findings from the Analysis section.
	Threats to validity	Discusses the main threats to validity and mitigation strategies applied.
Conclusions and Future Work	Summary	Provides a concise summary of the research objective and evaluation execution.
	Findings	Identifies the most important results of the study.
	Improvement opportunities	Suggestions for other studies to further investigate.
Acknowledgements	Identifies any sponsors, participants, and contributors who do not fulfil authorship criteria.	
References	Lists all cited literature in the format requested by the publisher.	
Appendices	Includes supplementary data and/or detailed analyses which might help others to use the results.	

APPENDIX F – Spearman correlation matrix of the MEEGA+ model

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	Usability									Player Experience																					
1	1.00																														
2	0.58	1.00																													
3	0.29	0.24	1.00																												
4	0.31	0.28	0.68	1.00																											
5	0.36	0.25	0.59	0.71	1.00																										
6	0.36	0.31	0.55	0.70	0.72	1.00																									
7	0.36	0.29	0.47	0.54	0.54	0.61	1.00																								
8	0.38	0.50	0.27	0.33	0.30	0.38	0.38	1.00																							
9	0.42	0.63	0.20	0.28	0.27	0.34	0.33	0.65	1.00																						
10	0.36	0.25	0.22	0.24	0.28	0.29	0.38	0.26	0.24	1.00																					
11	0.23	0.17	0.03	0.04	0.10	0.08	0.22	0.01	0.09	0.40	1.00																				
12	0.21	0.14	-	- 0.06	0.20	0.20	0.17	0.03	0.07	0.36	0.59	1.00																			
13	0.29	0.19	0.06	0.08	0.14	0.16	0.20	0.01	0.13	0.27	0.46	0.50	1.00																		
14	0.26	0.10	0.14	0.15	0.17	0.19	0.24	0.05	0.09	0.40	0.48	0.47	0.46	1.00																	
15	0.21	0.12	0.19	0.22	0.22	0.18	0.30	0.14	0.14	0.38	0.36	0.31	0.27	0.46	1.00																
16	0.26	0.19	0.17	0.18	0.22	0.22	0.32	0.14	0.14	0.48	0.44	0.44	0.35	0.56	0.48	1.00															
17	0.34	0.21	0.21	0.22	0.23	0.25	0.34	0.09	0.13	0.47	0.44	0.43	0.46	0.57	0.39	0.68	1.00														
18	0.19	0.14	0.08	0.08	0.09	0.14	0.20	0.10	0.06	0.28	0.32	0.34	0.34	0.37	0.20	0.37	0.40	1.00													
19	0.28	0.15	0.12	0.17	0.18	0.21	0.27	0.12	0.11	0.29	0.37	0.39	0.37	0.45	0.26	0.39	0.42	0.62	1.00												
20	0.31	0.20	0.15	0.13	0.13	0.19	0.25	0.11	0.11	0.36	0.39	0.40	0.40	0.48	0.26	0.45	0.49	0.69	0.69	1.00											
22	0.42	0.24	0.13	0.20	0.22	0.30	0.34	0.13	0.15	0.41	0.39	0.43	0.48	0.58	0.31	0.49	0.62	0.50	0.56	0.63	1.00										
22	0.34	0.19	0.09	0.15	0.18	0.22	0.27	0.11	0.15	0.31	0.32	0.30	0.39	0.45	0.22	0.38	0.47	0.43	0.44	0.55	0.67	1.00									
23	0.33	0.18	0.10	0.07	0.16	0.16	0.22	0.08	0.12	0.38	0.36	0.38	0.34	0.43	0.26	0.40	0.50	0.31	0.33	0.42	0.53	0.46	1.00								
24	0.30	0.12	0.09	0.07	0.16	0.15	0.23	0.02	0.08	0.38	0.36	0.38	0.43	0.49	0.25	0.38	0.49	0.35	0.40	0.48	0.58	0.49	0.58	1.00							
25	0.26	0.13	0.01	0.05	0.16	0.13	0.19	0.05	0.08	0.28	0.31	0.35	0.38	0.47	0.24	0.35	0.48	0.39	0.40	0.48	0.53	0.44	0.51	0.76	1.00						
26	0.18	0.19	0.15	0.17	0.19	0.24	0.29	0.16	0.18	0.39	0.38	0.29	0.29	0.45	0.36	0.52	0.51	0.27	0.30	0.34	0.44	0.29	0.36	0.34	0.34	1.00					
27	0.20	0.22	0.15	0.20	0.16	0.22	0.28	0.25	0.21	0.29	0.25	0.20	0.20	0.27	0.30	0.39	0.34	0.22	0.25	0.27	0.34	0.29	0.22	0.20	0.18	0.48	1.00				
28	0.30	0.27	0.15	0.20	0.18	0.23	0.28	0.26	0.27	0.43	0.33	0.31	0.27	0.39	0.37	0.50	0.48	0.22	0.29	0.34	0.45	0.37	0.34	0.29	0.30	0.48	0.65	1.00			
29	0.25	0.20	0.09	0.13	0.13	0.16	0.20	0.14	0.17	0.29	0.26	0.24	0.26	0.31	0.29	0.38	0.38	0.20	0.25	0.30	0.38	0.31	0.28	0.26	0.32	0.32	0.41	0.62	1.00		
30	0.23	0.17	0.13	0.16	0.17	0.19	0.30	0.19	0.18	0.46	0.37	0.39	0.26	0.43	0.37	0.58	0.49	0.26	0.30	0.34	0.44	0.33	0.36	0.33	0.30	0.53	0.53	0.65	0.53	1.00	
31	0.24	0.17	0.12	0.14	0.18	0.21	0.26	0.17	0.16	0.41	0.34	0.37	0.28	0.41	0.36	0.53	0.47	0.22	0.26	0.33	0.44	0.32	0.34	0.31	0.28	0.44	0.43	0.56	0.55	0.73	1.00