



MEEGA+: Um Modelo para a Avaliação de Jogos Educacionais para o ensino de Computação

MEEGA+: A Model for the Evaluation of Educational Games for Computing Education

Giani Petri
Colégio Politécnico, Universidade
Federal de Santa Maria (UFSM)
gpetri@inf.ufsm.br

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

Adriano Ferreti Borgatto
Departamento de Informática e
Estatística (INE), Universidade
Federal de Santa Catarina (UFSC)
adriano.borgatto@ufsc.br

Resumo

Jogos educacionais estão sendo usados como uma estratégia instrucional inovadora para o ensino e aprendizagem em diversas áreas de conhecimento. No entanto, é essencial avaliar sistematicamente esses jogos a fim de obter evidências de suas contribuições. Atualmente, existem poucos modelos/métodos que forneçam um suporte sistemático para a avaliação de jogos educacionais. Neste contexto, um modelo que têm se destacado é o MEEGA (Model for the Evaluation of Educational Games). Porém, resultados de uma análise de sua versão inicial apontam limitações em termos de sua validade. Assim, o objetivo deste artigo é apresentar o modelo MEEGA+, evoluindo a versão inicial do modelo MEEGA. O modelo MEEGA+ é sistematicamente desenvolvido por meio da decomposição de um objetivo de avaliação em medidas e definindo um instrumento de medição para avaliar a qualidade de jogos educacionais. A análise da confiabilidade e validade, com base em dados coletados em 48 estudos de caso envolvendo uma população de 843 alunos, avaliando 18 diferentes jogos para o ensino de computação, indica que o instrumento de medição do modelo MEEGA+ é confiável (alfa de Cronbach $\alpha=,928$). Os resultados de uma análise fatorial exploratória indicam que a qualidade de jogos educacionais é avaliada por meio de dois fatores de qualidade (experiência do jogador e usabilidade). Os resultados também demonstram validade convergente por meio de um grau de correlação satisfatório encontrado entre a maioria dos itens do instrumento de medição. Assim, o modelo MEEGA+ fornece um instrumento de medição confiável e válido para desenvolvedores de jogos, instrutores e pesquisadores, a fim de avaliar a qualidade de jogos educacionais como base para sua melhoria e adoção na prática.

Palavras-Chave: Jogo Educacional; Avaliação; Modelo; Instrumento de medição.

Abstract

Educational games are being used as an innovative instructional strategy for teaching and learning in different knowledge areas. However, it is essential to systematically evaluate such games in order to obtain sound evidence on their contributions. Currently, there are a few models/methods that provide a systematic support for the evaluation of educational games. In this context, a prominent evaluation model is MEEGA (Model for the Evaluation of Educational Games). However, analyses of its initial version have identified limitations in terms of its validity. Thus, the objective of this article is to present its evolution, the MEEGA+ model. The MEEGA+ model has been systematically developed by decomposing evaluation goals into measures and defining a measurement instrument to evaluate the quality of educational games. The analysis of the reliability and validity based on data collected from 48 case studies involving a population of 843 students, evaluating 18 different games for computing education, indicates that the MEEGA+ measurement instrument can be considered reliable (Cronbach's alpha $\alpha=,928$). Results of an exploratory factor analysis indicate that the quality of educational games is evaluated through two quality factors (player experience and usability). In addition, the results demonstrate convergent validity through a satisfactory degree of correlation found between most of the items for all quality factors. Thus,

Cite as: Petri, G., von Wangenheim, C. G. & Borgatto, A. F. (2019). MEEGA+: A Model for the Evaluation of Educational Games for Computing Education (MEEGA+: Um Modelo para a Avaliação de Jogos Educacionais para o ensino de Computação). *Brazilian Journal of Computers in Education (Revista Brasileira de Informática na Educação - RBIE)*, 27(3), 52-81. DOI: 10.5753/RBIE.2019.27.03.52



the MEEGA+ model provides a reliable and valid measurement instrument for game creators, instructors and researchers in order to evaluate the quality of educational games as a basis for their improvement and adoption in practice.

Keywords: Educational game; Evaluation; Model; Measurement instrument.

1 Introdução

Nos últimos anos, os jogos educacionais têm sido utilizados como uma estratégia instrucional em diversas áreas do conhecimento, tais como: matemática, computação, saúde, nutrição, entre outras (Connolly et al., 2012; Backlund & Hendrix, 2013; Calderón & Ruiz, 2015). Esses jogos, além de possuir características de jogos de entretenimento, são projetados especificamente para ensinar as pessoas sobre um determinado assunto, expandir conceitos, reforçar o desenvolvimento, ou ajudá-las a explorar ou aprender uma habilidade ou uma mudança de atitude enquanto elas jogam (Dempsey et al., 1996). Na área de computação, os jogos educacionais são usados tipicamente para fornecer mais oportunidades práticas de aprendizagem para os estudantes por meio de um ambiente seguro e controlado (Backlund & Hendrix, 2013; Battistella & Gresse von Wangenheim, 2016).

Acredita-se que os jogos educacionais possam ser uma estratégia instrucional eficaz e eficiente. No entanto, na prática, os benefícios esperados dos jogos usados para o ensino ainda são questionáveis, devido à falta de estudos que forneçam evidências desses benefícios (Caulfield et al., 2011; Connolly et al., 2012; Petri & Gresse von Wangenheim, 2017). Uma razão para isso pode ser pelo fato que a maioria das avaliações de jogos educacionais são realizados de forma *ad-hoc* em termos de design de pesquisa, medição, coleta de dados e análise. Além disso, frequentemente, dados são coletados apenas em forma de comentários informais dos alunos e/ou por meio de questionários desenvolvidos de maneira *ad-hoc* (Petri & Gresse von Wangenheim, 2017; Boyle et al., 2016; Kosa et al., 2016; Calderón & Ruiz, 2015; Calderón et al., 2018). Portanto, essa falta de rigor científico, recorrente na comunidade de informática na educação, conforme destacado no princípio V do manifesto apresentado por Bittencourt & Isotani (2018), e o fato de a maioria dos estudos serem realizados com uma amostra pequena, deixa a confiabilidade e a validade de seus resultados e, assim, a qualidade e/ou a efetividade de tais jogos questionáveis.

Outra razão para a falta de rigor científico das avaliações de tais jogos pode ser pelo fato que, atualmente, existem poucas abordagens que forneçam um suporte sistemático para avaliações de jogos (Petri & Gresse von Wangenheim, 2016). Entre as abordagens para a avaliação de jogos encontradas na literatura, duas se destacam: MEEGA (*Model for the Evaluation of Educational Games*) (Savi et al., 2011) e EGameFlow (Fu et al., 2009). EGameFlow é uma escala que objetiva avaliar jogos *e-learning* do ponto de vista dos usuários por meio de um questionário (Fu et al., 2009). Embora a escala EGameFlow tenha sido desenvolvida de forma sistemática e avaliada em termos de confiabilidade e validade, parece ter sido descontinuada por seus autores (Petri & Gresse von Wangenheim, 2016). MEEGA é um modelo que avalia jogos educacionais em termos de motivação, experiência do usuário e aprendizagem por meio da reação dos estudantes. Atualmente, o modelo MEEGA parece ser uma das abordagens mais utilizadas na prática para a avaliação de jogos educacionais (Calderón & Ruiz, 2015; Petri & Gresse von Wangenheim, 2017). Embora o modelo MEEGA, inicialmente, apresente confiabilidade e validade, uma avaliação mais abrangente do modelo MEEGA com base em uma amostra de 723 respostas de alunos, identificou-se algumas oportunidades de melhoria em relação a sua validade. Essas oportunidades de melhoria estão relacionadas a uma sobreposição dos conceitos teóricos de motivação e experiência do usuário,



bem como à falta de compreensão da redação de alguns itens do questionário (Petri et al., 2017). Assim, avaliações usando o modelo MEEGA podem estar levando a resultados imprecisos sobre a qualidade do jogo, não identificando corretamente evidências de seus benefícios em relação aos conceitos sobrepostos (motivação e experiência do usuário). Isso, conseqüentemente, pode prejudicar a adoção efetiva e eficiente de jogos como uma estratégia instrucional. Portanto, o modelo MEEGA precisa ser melhorado para fornecer resultados válidos de avaliações de jogos educacionais. Assim, a questão de pesquisa abordada neste artigo é como avaliar jogos educacionais utilizando um modelo sistemático, válido e confiável.

Diante disso, a principal contribuição deste artigo é o desenvolvimento e a avaliação do modelo MEEGA+, um modelo sistemático para a avaliação de jogos usados para o ensino de computação, como uma evolução da versão inicial do modelo MEEGA (Savi et al., 2011). Os resultados deste estudo permitem que desenvolvedores de jogos, pesquisadores e instrutores avaliem a qualidade dos jogos usados para o ensino de computação. Assim, contribuindo para a análise de quais jogos proporcionam um aprendizado eficaz e eficiente, colaborando positivamente para a aprendizagem dos estudantes.

2 Método de Pesquisa

Com o objetivo de desenvolver e avaliar o modelo MEEGA+, é adotada uma pesquisa multi-método, como mostrado na Figura 1.

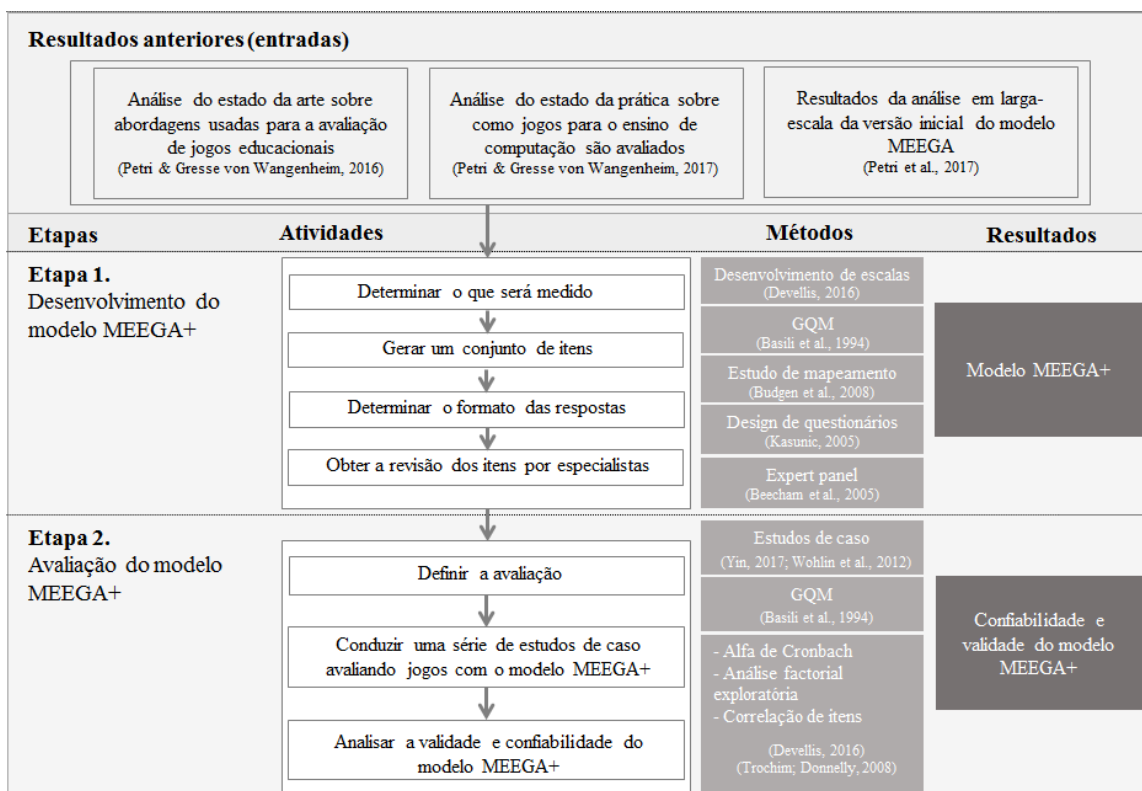


Figura 1: Método de pesquisa

Como etapas anteriores a este estudo (entradas), foi conduzido uma análise do estado da arte com o objetivo de identificar abordagens existentes (métodos, modelos, frameworks e escalas) para a avaliação sistemática de jogos educacionais (Petri & Gresse von Wangenheim, 2016). Adicionalmente, foi realizada uma análise do estado da prática, de modo a identificar



como os jogos usados para o ensino de computação são avaliados atualmente (Petri & Gresse von Wangenheim, 2017). Como resultado, identificou-se o modelo MEEGA (Savi et al., 2011) como uma abordagem de avaliação proeminente, sendo amplamente utilizado na prática.

Portanto, foi realizado um estudo em larga-escala da versão inicial do modelo MEEGA analisando sua validade e confiabilidade. A análise foi realizada com base em dados coletados em 43 estudos de caso, avaliando 20 jogos diferentes para ensino de computação, envolvendo uma população de 723 estudantes (Petri et al., 2017). Como resultado, identificou-se que a versão inicial do modelo MEEGA é aceitável em termos de confiabilidade ($\alpha = 0,915$). No entanto, em termos de validade, uma sobreposição conceitual em relação aos fatores de motivação e experiência do usuário foi observada, indicando a necessidade de reestruturação dos fatores do modelo MEEGA.

Etapa 1 – Desenvolvimento do Modelo MEEGA+. Nesta etapa o modelo MEEGA+ é desenvolvido, como uma evolução da versão inicial do modelo MEEGA, levando em consideração o levantamento do estado da arte (Petri & Gresse von Wangenheim, 2016) e prática (Petri & Gresse von Wangenheim, 2017) e os resultados da reavaliação do modelo MEEGA (Petri et al., 2017). O desenvolvimento do modelo MEEGA+ segue os procedimentos do guia de desenvolvimento de escalas proposto por DeVellis (2016) e do guia para design de questionários proposto por Kasunic (2005):

Determinar o que será medido. Adotando a abordagem GQM (*Goal/Question/Metric*) (Basili et al., 1994) o objetivo da avaliação é definido e decomposto sistematicamente em fatores a serem medidos (modelo teórico). Os fatores são definidos de modo a apoiar o desenvolvimento do instrumento de medição (questionário), com base em um estudo de mapeamento de seus conceitos seguindo o procedimento proposto por Budgen et al. (2008).

Gerar um conjunto de itens. A medição é operacionalizada pela decomposição dos fatores/dimensões em itens de um instrumento de medição. A definição dos itens é baseada em outros questionários padronizados encontrados na literatura. O conjunto de itens é analisado em termos de similaridade e redundância e, então, selecionados. De modo a padronizar os itens selecionados, todos os itens são refinados e transformados em afirmações.

Determinar o formato das respostas. Nesta etapa, o formato de respostas para os itens do instrumento de medição é definido. Esta definição é baseada no formato de respostas normalmente usados em questionários padronizados na literatura e com base no guia de desenvolvimento de escalas proposto por DeVellis (2016).

Obter a revisão dos itens por especialistas. Nesta etapa, a validade aparente (Trochim & Donnelly, 2008) é analisada por um painel de especialistas (Beecham et al., 2005). O painel de especialistas é composto por um grupo multidisciplinar de pesquisadores seniores com experiência em computação e/ou estatística. A revisão visa analisar a clareza, relevância, consistência e completude dos itens do instrumento de medição do modelo MEEGA+. As sugestões dos especialistas, incluindo mudanças na redação e formatação do texto, foram consideradas no desenvolvimento do instrumento de medição MEEGA+.

Etapa 2 – Avaliação do Modelo MEEGA+. De modo a avaliar o modelo MEEGA+ em termos de confiabilidade e validade de seu instrumento de medição, é realizada uma série de estudos de caso (Yin, 2017; Wohlin et al., 2012).

Definir a avaliação. O objetivo do estudo é definido usando a abordagem GQM (Basili et al., 1994) e decomposto em aspectos de qualidade e questões de análise. O design da pesquisa e o processo de coleta de dados são definidos com base no modelo MEEGA+, adotado para as



avaliações dos jogos. Os dados são coletados usando o instrumento de medição do modelo MEEGA+.

Conduzir uma série de estudos de caso. Nesta fase, uma série de estudos de caso avaliando vários jogos para o ensino de computação é conduzida. Em cada estudo de caso, após a sessão de jogo (tratamento), o instrumento de medição do modelo MEEGA+ é usado para coletar dados sobre as percepções dos participantes sobre o jogo (pós-teste). Para seleção dos participantes é adotada uma técnica de amostragem não probabilística em cada estudo de caso aplicando o método de amostragem por conveniência (Trochim & Donnelly, 2008), em que a amostra é composta por alunos matriculados nas respectivas disciplinas de cursos de computação onde os jogos são aplicados.

Analisar a validade e confiabilidade. Na fase de análise, os dados coletados em cada estudo de caso são agrupados em uma única amostra para análise de dados e, então, responder às questões de análise definidas. Os dados são analisados em termos de confiabilidade e validade de constructo seguindo a definição de Trochim & Donnelly (2008) e o guia de desenvolvimento de escala proposto por DeVellis (2016). Em termos de confiabilidade, é medida a consistência interna por meio do coeficiente alfa de Cronbach (Cronbach, 1951). A validade de constructo é analisada utilizando uma análise fatorial exploratória e com base em evidências de validade convergente e discriminante, obtida por meio do grau de correlação entre os itens (DeVellis, 2016; Trochim & Donnelly, 2008). Os resultados da análise estatística são interpretados por pesquisadores no contexto de ensino de computação, a fim de identificar a confiabilidade e a validade do instrumento de medição do modelo MEEGA+.

3 Trabalhos Relacionados

Como resultado de uma revisão sistemática da literatura realizada previamente (Petri & Gresse von Wangenheim, 2016) analisando abordagens existentes para a avaliação de jogos educacionais, identificou-se poucas abordagens para avaliar jogos educacionais (Freitas & Oliver, 2006; Connolly et al., 2009; Savi et al., 2011; Carvalho, 2012, Ak, 2012; Mayer, 2012).

A maioria das abordagens são *frameworks*, ao invés de modelos ou métodos, como proposto por Freitas & Oliver (2006), Connolly et al. (2009) e Carvalho (2012). Em geral, esses *frameworks* normalmente definem um conjunto de dimensões, tipicamente avaliando fatores como motivação, usabilidade, engajamento etc., que podem ser adaptadas a requisitos específicos em uma avaliação de um jogo em uma área específica de conhecimento. No entanto, tais abordagens falham em fornecer suporte concreto sobre como operacionalizar as avaliações na prática e deixam dúvidas sobre sua validade e confiabilidade, pois apresentam somente avaliações com estudos pilotos envolvendo um pequeno número de alunos (Freitas & Oliver, 2006; Connolly et al., 2009; Carvalho, 2012; Abdellatif et al., 2018).

Fu et al. (2009) e Ak (2012) apresentam escalas para a avaliação de jogos *e-learning*, tipicamente avaliando fatores como imersão, interação social, desafio, apreciação, concentração, controle e aprendizagem/melhoria do conhecimento, utilizando um questionário para coleta de dados. A escala EGameFlow, proposta por Fu et al. (2009), foi desenvolvida utilizando uma metodologia sistemática e validada em termos de análise de itens, confiabilidade e validade por meio de quatro sessões de jogos. No entanto, embora a escala EGameFlow tenha apresentado validade e confiabilidade satisfatórias, parece ter sido descontinuada por seus autores. Por outro lado, a escala proposta por Ak (2012) não cita informações sobre seu processo de



desenvolvimento, coleta de dados e métodos de análise de dados, nem sua validação são relatadas, deixando sua validade e confiabilidade questionáveis.

Estudos mais recentes estão usando *learning analytics* em jogos educacionais adotando sistemas de rastreamento que passam despercebidos pelos jogadores (Freire et al., 2016; Serrano-Laguna et al., 2018). Freire et al. (2016) e Serrano-Laguna et al. (2018) propuseram uma metodologia para avaliar a eficácia dos jogos sérios e inferir os resultados de aprendizagem dos jogadores usando *learning analytics*, coletando dados sobre o comportamento dos jogadores durante o jogo. No entanto, uma limitação desses estudos é que eles tipicamente avaliam apenas a efetividade da aprendizagem, não considerando outros fatores como engajamento, motivação, diversão, etc., igualmente importantes para uma estratégia instrucional usada em um contexto educacional como comumente indicado (Abt, 2002; Prensky, 2007; Ritterfeld et al., 2010; Djaouti et al., 2011).

Analisando os fatores de qualidade utilizados para avaliar jogos educacionais pelas abordagens relacionadas, observa-se que existe uma grande diversidade de fatores. Observa-se também que não há consenso sobre quais fatores de qualidade avaliar. Tipicamente, os fatores são definidos de forma inconsistente ao longo dos estudos e não claramente baseados em constructos teóricos (Abdellatif et al., 2018). No entanto, ao analisar os trabalhos relacionados, observa-se similaridades nas abordagens em avaliar os jogos educacionais considerando, principalmente, fatores como motivação, experiência do usuário, usabilidade, engajamento, apreciação (*enjoyment*) e aprendizagem. A aprendizagem, aparece como sendo o fator mais avaliado pelas abordagens, já que o principal objetivo dos jogos educacionais é potencializar a aprendizagem dos alunos. A Tabela 1 apresenta uma síntese dos fatores de avaliação considerados pelos trabalhos relacionados.

Tabela 1. Fatores de qualidade considerados pelas abordagens relacionadas

Trabalho relacionado	Tipo de abordagem	Fatores de qualidade avaliados pela abordagem					
		Motivação	Experiência do usuário	Usabilidade	Engajamento	Apreciação	Aprendizagem
Savi et al. (2011)	Modelo	X	X				X
Freitas & Oliver (2006)	Framework		X	X			X
Connolly et al. (2009)	Framework	X	X				X
Carvalho (2012)	Framework	X		X	X		X
Ak (2012)	Escala				X	X	X
Mayer (2012)	Modelo		X	X			X
Fu et al. (2009)	Escala		X	X		X	X
Serrano-Laguna et al. (2018)	Metodologia						X

Apenas uma abordagem encontrada na literatura foi desenvolvida considerando os requisitos específicos para a avaliação de jogos no contexto de ensino de computação. O modelo MEEGA (Savi et al., 2011), mede três fatores dos jogos educacionais: motivação, experiência do usuário e aprendizagem. O MEEGA foi sistematicamente desenvolvido e define explicitamente o objetivo da avaliação e fornece um questionário padronizado para ser aplicado a fim de coletar dados sobre a percepção dos alunos usando um delineamento de pesquisa não-experimental. O modelo também foi avaliado estatisticamente em termos de confiabilidade e validade, com base em dados coletados em estudos de caso de diferentes autores. No entanto, os resultados de uma análise sistemática de uma avaliação em larga-escala da versão inicial do modelo MEEGA mostram que existem algumas limitações em termos de sua validade (Petri et al., 2017), indicando uma sobreposição conceitual entre os fatores de motivação e experiência do usuário.



Portanto, existem poucas abordagens para avaliar sistematicamente jogos educacionais, com apenas uma abordagem específica para a avaliação de jogos para o ensino de computação (MEEGA). Assim, observa-se uma falta de abordagens sistemáticas, válidas e confiáveis utilizadas para a avaliação de jogos para o ensino de computação que abordam tanto a avaliação da aprendizagem quanto a avaliação de aspectos importantes para proporcionar uma experiência positiva e envolvente aos estudantes. Portanto, observa-se a necessidade de evoluir a abordagem que tem sido utilizada na prática para a avaliação de jogos educacionais, a fim de melhorar sua validade e confiabilidade como base para fornecer resultados válidos que possam ser utilizados para uma decisão sobre a adoção de tais jogos e/ou sua melhoria contínua.

4 Desenvolvimento do Modelo MEEGA+

Com base nas oportunidades de melhoria identificadas, esta seção apresenta o modelo MEEGA+ desenvolvido como uma evolução da versão inicial do modelo MEEGA (Savi et al., 2011). O objetivo do modelo MEEGA+ é analisar jogos educacionais com o propósito de avaliar a percepção da qualidade do ponto de vista de alunos no contexto de cursos superiores da área de computação (Petri et al., 2018).

Seguindo a abordagem GQM, esse objetivo é sistematicamente decomposto em fatores/dimensões a serem medidos. A versão inicial do modelo MEEGA avalia jogos em termos de motivação, experiência do usuário e aprendizagem (Savi et al., 2011). No entanto, os resultados de sua avaliação em larga-escala mostram que existe uma sobreposição conceitual entre os fatores de motivação e a experiência do usuário (Petri et al., 2017). Adicionalmente, analisando os resultados da revisão da literatura (Petri & Gresse von Wangenheim, 2016; Petri & Gresse von Wangenheim, 2017), identificou-se uma tendência que abrange um conjunto de fatores utilizados para a avaliação de jogos, principalmente relacionados à motivação, experiência do usuário, usabilidade, engajamento, apreciação (*enjoyment*) e aprendizagem percebida (Petri & Gresse von Wangenheim, 2017). Esses fatores de qualidade são ainda fragmentados (explicados) em dimensões que se sobrepõem conceitualmente, como por exemplo, os fatores de imersão e atenção focada. Assim, algumas dimensões de diferentes fatores são semelhantes ou relacionadas a outra dimensão de outro fator. Portanto, é realizado um mapeamento de cada fator (colunas) e suas dimensões (linhas), analisando suas definições conceituais e semelhanças, conforme apresentado na Tabela 2.

Tabela 2. Mapeamento dos fatores de avaliação e suas dimensões

Fator/Dimensão	Motivação (Keller, 1987; Savi et al., 2011)	Experiência do usuário (Savi et al., 2011)	Usabilidade (ISO/IEC, 2014)	Engajamento (Wiebe et al., 2014)	Apreciação (Fu et al., 2009)	Aprendizagem percebida (Sindre & Moody, 2003; Savi et al., 2011)
Atenção	X			X (Atenção focada)	X (Concentração)	
Relevância	X					
Confiança	X					
Satisfação	X			X		
Imersão		X		X (Atenção focada)	X	
Interação social		X			X	
Desafio		X			X	
Diversão		X		X (Satisfação)		
Competência		X				



Controle/ Operabilidade/ Autonomia		X (Controle)	X (Operabilidade)		X (Autonomia)	
Aprendizibilidade			X			
Estética			X	X		
Acessibilidade			X			
Proteção de erro do usuário			X			
Usabilidade percebida				X		
Metas claras					X	X
Feedback					X	
Melhoria do conhecimento					X	X

Com base nesse mapeamento (Tabela 2), é definido um conjunto de dimensões a serem medidas pelo modelo MEEGA+ para as avaliações de jogos digitais e não-digitais usados no ensino de computação (Tabela 3).

Tabela 3. Definição das dimensões/subdimensões

Dimensão/Subdimensão	Definição	
Atenção focada	Avaliar a atenção, concentração focada, absorção e dissociação temporal dos alunos (Keller, 1987; Wiebe et al., 2014; Savi et al., 2011).	
Diversão	Avaliar a sensação de prazer, felicidade, relaxamento e distração dos alunos (Poels et al., 2007; Savi et al., 2011).	
Desafio	Avaliar quanto o jogo é suficientemente desafiador em relação ao nível de competência do aluno. Novos obstáculos e situações devem ser apresentados ao longo do jogo para minimizar a fadiga e manter os alunos interessados (Sweetser & Wyeth, 2005; Savi et al., 2011).	
Interação social	Avaliar se o jogo promove a sensação de um ambiente compartilhado e conexão com outras pessoas em atividades de cooperação ou competição (Fu et al., 2009; Savi et al., 2011).	
Confiança	Avaliar se os alunos são capazes de progredir no estudo do conteúdo educacional por meio de seu esforço e habilidade (e.g., por meio de tarefas com crescente nível de dificuldade) (Keller, 1987; Savi et al., 2011).	
Relevância	Avaliar se os alunos percebem que a proposta educacional é consistente com seus objetivos e que podem vincular o conteúdo ao futuro profissional ou acadêmico (Keller, 1987; Savi et al., 2011).	
Satisfação	Avaliar se os alunos sentem que o esforço dedicado resulta em aprendizagem (Keller, 1987; Savi et al., 2011).	
Usabilidade	Aprendizibilidade	Avaliar se o jogo permite que os usuários aprendam a jogá-lo de forma fácil e rápida (ISO/IEC, 2014).
	Operabilidade	Avaliar o grau em que um jogo possui atributos que facilitam a operação e o controle (ISO/IEC, 2014).
	Estética	Avaliar se a interface do jogo permite uma interação agradável e satisfatória com o usuário (ISO/IEC, 2014).
	Acessibilidade	Avaliar se o jogo pode ser usado por pessoas com deficiência visual baixa/moderada e/ou com daltonismo (ISO/IEC, 2014).
	Proteção contra erros do usuário	Avaliar se o jogo protege os usuários de cometer erros (ISO/IEC, 2014). Aplicado apenas para avaliação de jogos digitais.
Aprendizagem percebida	Avaliar as percepções do efeito geral do jogo na aprendizagem dos alunos na disciplina (Sindre & Moody, 2003; Savi et al., 2011).	

Para operacionalizar a medição dessas dimensões definidas, um design de pesquisa é definido.

Definição do design de pesquisa. Avaliações de jogos educacionais são tipicamente conduzidas por meio de estudos empíricos, que visam medir se o público-alvo atingiu os objetivos definidos (Branch, 2010). Esses estudos podem variar de estudos experimentais a não-experimentais (Wohlin et al., 2012). Para definir um design de pesquisa para um estudo empírico, suas limitações práticas e objetivo(s) precisam ser considerados. No contexto desse estudo, relacionado à avaliação de jogos para o ensino de computação, espera-se que a avaliação possa ser realizada de forma rápida, não intrusiva, a fim de não interromper o fluxo normal de uma aula e não prejudicar os participantes envolvidos no estudo. Portanto, opta-se pela realização de estudos de caso que permite uma pesquisa aprofundada de um indivíduo, grupo ou evento (Yin, 2017). O design de pesquisa para o modelo MEEGA+ é definido como não-experimental com pós-teste (*one-shot post-test only design*) com um único grupo, no qual o estudo de caso começa com a aplicação do tratamento (jogo educacional) e então um



instrumento de medição (questionário) é respondido pelos alunos (autoavaliação) para coletar dados sobre suas percepções sobre o jogo.

A condução de estudos de caso nesse contexto é justificada por restrições típicas ao aplicar os jogos em sala de aula. Embora os experimentos adotem design de pesquisa mais rigorosos, eles podem ter limitações significativas quando conduzidos em contextos educacionais, como no ensino de computação (Schanzenbach, 2012; All et al., 2016). Por exemplo, em relação à viabilidade de implementação de um experimento, os alunos podem se sentir prejudicados pela alocação aleatória em grupo(s) de controle e experimental(is). Além disso, as diferentes intervenções utilizadas no experimento podem ser consideradas inferiores (Sitzmann et al., 2010; All et al., 2016). Ameaças também podem ser introduzidas por meio de diferenças entre pré/pós testes e/ou o impacto de fatores causais adicionais nos resultados do teste. Além disso, para obter resultados estatísticos significativos de tais experimentos, é necessário um tamanho de amostra considerável (Sitzmann et al., 2010; Wohlin et al., 2012; All et al., 2016). No entanto, isso pode não ser viável devido ao número de alunos comumente matriculados em cursos de computação (Bowman, 2018). Assim, mesmo com essa quantidade substancial de esforço, o estudo pode não produzir resultados significativos (All et al., 2016). Portanto, um estudo de caso pode ser um design de pesquisa mais adequado, fornecendo uma compreensão profunda do objeto de estudo (jogo educacional) em seu contexto real (ensino de computação) (Wohlin et al., 2012), minimizando as interrupções do fluxo normal de uma aula.

A coleta de dados por meio da autoavaliação é bem aceita para medir diversos fatores, como engajamento, motivação ou usabilidade (Johnson & Christensen, 2014). No entanto, sua confiabilidade e validade ainda podem ser questionáveis quando usadas para medir a aprendizagem (Brown et al., 2015). A esse respeito, embora não haja consenso, vários estudos fornecem evidências de que a autoavaliação fornece informações confiáveis e válidas (Ross, 2006; Andrade & Valtcheva, 2009; Thomas et al., 2011; Brown et al., 2015; Sharma et al., 2015), especialmente quando se utilizam instrumentos de medição confiáveis e válidos (Brown et al., 2015). Portanto, um compromisso pode ser o desenvolvimento de instrumentos de medição padronizados e estatisticamente validados, aumentando a validade e a confiabilidade dos dados coletados nos estudos (Kasunic, 2005; Devellis, 2016).

Definição dos itens do instrumento de medição (questionário). Com base no design de pesquisa e nas dimensões definidas, é derivado um conjunto de itens, melhorando a versão inicial do questionário do modelo MEEGA, personalizando e unindo os questionários padronizados existentes encontrados na literatura (Savi et al., 2011; Keller, 1987; Tullis & Albert, 2008; Sindre & Moody, 2003; Sweetser & Wyeth, 2005; Poels et al., 2007; Gámez, 2009; Takatalo et al., 2010, O'Brien & Toms, 2010; Wiebe et al., 2014; Fu et al., 2009; Mohamed & Jaafar, 2010; Zaibon & Shiratuddin, 2010; Zaibon, 2015; Brooke, 1996; Davis, 1989). A Tabela 4 mostra os itens definidos para o questionário do modelo MEEGA+ para cada dimensão/subdimensão.

Tabela 4. Itens do questionário do modelo MEEGA+

Dimensão/Subdimensão	Item	Descrição do Item	
Usabilidade	Estética	1	O design do jogo é atraente (interface, gráficos, tabuleiro, cartas, etc.).
		2	Os textos, cores e fontes combinam e são consistentes.
	Aprendizibilidade	3	Eu precisei aprender poucas coisas para poder começar a jogar o jogo.
		4	Aprender a jogar este jogo foi fácil para mim.
		5	Eu acho que a maioria das pessoas aprenderiam a jogar este jogo rapidamente.
	Operabilidade	6	Eu considero que o jogo é fácil de jogar.
		7	As regras do jogo são claras e compreensíveis.



	Acessibilidade	8	As fontes (tamanho e estilo) utilizadas no jogo são legíveis.
		9	As cores utilizadas no jogo são compreensíveis.
		10	O jogo permite personalizar a aparência (fonte e/ou cor) conforme a minha necessidade.
	Proteção contra erros do usuário	11	O jogo me protege de cometer erros.
		12	Quando eu cometo um erro é fácil de me recuperar rapidamente.
Confiança		13	Quando olhei pela primeira vez o jogo, eu tive a impressão de que seria fácil para mim.
		14	A organização do conteúdo me ajudou a estar confiante de que eu iria aprender com este jogo.
Desafio		15	Este jogo é adequadamente desafiador para mim.
		16	O jogo oferece novos desafios (oferece novos obstáculos, situações ou variações) com um ritmo adequado.
		17	O jogo não se torna monótono nas suas tarefas (repetitivo ou com tarefas chatas).
Satisfação		18	Completar as tarefas do jogo me deu um sentimento de realização.
		19	É devido ao meu esforço pessoal que eu consigo avançar no jogo.
		20	Me sinto satisfeito com as coisas que aprendi no jogo.
		21	Eu recomendaria este jogo para meus colegas.
Interação social		22	Eu pude interagir com outras pessoas durante o jogo.
		23	O jogo promove momentos de cooperação e/ou competição entre os jogadores.
		24	Eu me senti bem interagindo com outras pessoas durante o jogo.
Diversão		25	Eu me diverti com o jogo.
		26	Aconteceu alguma situação durante o jogo (elementos do jogo, competição, etc.) que me fez sorrir.
Atenção focada		27	Houve algo interessante no início do jogo que capturou minha atenção.
		28	Eu estava tão envolvido no jogo que eu perdi a noção do tempo.
		29	Eu esqueci sobre o ambiente ao meu redor enquanto jogava este jogo.
Relevância		30	O conteúdo do jogo é relevante para os meus interesses.
		31	É claro para mim como o conteúdo do jogo está relacionado com a disciplina.
		32	O jogo é um método de ensino adequado para esta disciplina.
		33	Eu prefiro aprender com este jogo do que de outra forma (outro método de ensino).
Aprendizagem percebida		34	O jogo contribuiu para a minha aprendizagem na disciplina.
		35	O jogo foi eficiente para minha aprendizagem, em comparação com outras atividades da disciplina.

Os itens apresentados na Tabela 4 compõem o questionário do modelo MEEGA+ para avaliar jogos digitais e não-digitais. No entanto, os itens 10, 11 e 12 estão medindo apenas a usabilidade de jogos digitais, referente à personalização dos jogos e proteção contra erros do usuário. Assim, ao avaliar jogos não-digitais, esses itens devem ser desconsiderados do questionário.

Adicionalmente, itens relacionados aos objetivos de aprendizagem de cada jogo avaliado devem ser incluídos no questionário, na dimensão de aprendizagem percebida, e customizados de acordo com os objetivos de aprendizagem específicos de cada jogo educacional. Tipicamente, os jogos para o ensino de computação são usados para construir conhecimentos nos níveis cognitivos de lembrança, compreensão e aplicação (ACM/IEEE-CS, 2013) de acordo com a versão revisada da taxonomia de Bloom (Anderson et al., 2001), pois representam o que pode ser aprendido durante um curso de graduação (ACM/IEEE-CS, 2013). No entanto, o modelo MEEGA+ também é flexível para cobrir objetivos em níveis cognitivos superiores, como análise, avaliação e criação. Além do conhecimento técnico, os jogos educacionais também podem contribuir para desenvolver habilidades, como resolução de problemas, comunicação interpessoal, trabalho em equipe, liderança, gerenciamento, entre outros (ACM/IEEE-CS, 2013). Esses objetivos de aprendizagem se concentram em uma mudança e/ou desenvolvimento de comportamentos ou habilidades. Eles podem ser classificados de acordo com a taxonomia da Bloom do domínio psicomotor (Simpson, 1972). Além disso, os jogos podem também contribuir para desenvolver atitudes profissionais, como um comportamento maduro do aluno, considerando questões profissionais e legais, bem como uma atitude ética na profissão (ACM/IEEE-CS, 2013). Os objetivos de aprendizagem relacionados a isso visam



tipicamente a conscientização e o desenvolvimento de atitudes, emoções e sentimentos e podem ser classificados de acordo com a taxonomia de Bloom do domínio afetivo (Krathwohl et al., 1973). Assim, de acordo com a taxonomia dos objetivos de aprendizagem, para cada objetivo(s) de aprendizagem do jogo a ser avaliado, a afirmação seguinte deve ser customizada e incluída no instrumento de medição do modelo MEEGA+: “O jogo contribuiu para <verbo conforme nível do objetivo de aprendizagem (cognitivo, psicomotor, afetivo)> <objetivo/conceito>”. Por exemplo, de acordo com os objetivos de aprendizagem do jogo SCRUMIA (Gresse von Wangenheim et al., 2013), um jogo para reforçar a compreensão dos conceitos SCRUM e para exercitar a aplicação do processo SCRUM, tal afirmação seria: “O jogo contribuiu para lembrar os conceitos sobre o Planejamento de uma Sprint”. Essas afirmações relacionadas aos objetivos de aprendizagem do jogo compõem o questionário MEEGA+ que será aplicado após a sessão do jogo, a fim de coletar a percepção dos alunos sobre seu nível de concordância (ou discordância) no cumprimento dos objetivos de aprendizagem do jogo avaliado.

Formato das respostas. O formato de resposta definido para o questionário do modelo MEEGA+ é uma escala Likert de 5 pontos com alternativas de resposta que variam de discordo totalmente a concordo totalmente (DeVellis, 2016; Malhotra & Birks, 2008). O uso da escala Likert, em seu formato original de 5 pontos, permite expressar a opinião do indivíduo (aluno) sob o objeto de estudo (jogo educacional) com maior precisão, além de permitir que o indivíduo fique mais confortável para expressar sua opinião, usando um ponto neutro e, assim, aumentando a qualidade das respostas (Dawes, 2008).

Revisão dos itens por especialistas. A validade aparente do questionário MEEGA+ foi analisada por um grupo multidisciplinar de quatro especialistas (pesquisadores seniores) das áreas de informática e estatística. A revisão dos itens do questionário visou avaliar sua clareza, relevância, consistência e completude. As sugestões dos especialistas, incluindo mudanças na redação e formatação de texto, foram consideradas no desenvolvimento do questionário MEEGA+ (Tabela 4).

Análise dos dados coletados. De modo a alcançar o objetivo da avaliação definido no modelo MEEGA+, os dados coletados usando o questionário são analisados por meio de métodos de estatística descritiva (Trochim & Donnelly, 2008; Wohlin et al., 2012). Métodos estatísticos descritivos são usados para descrever e apresentar graficamente aspectos interessantes dos dados coletados (Wohlin et al., 2012). Assim, os dados coletados são analisados em termos de medidas de tendência central (mediana, média e frequência de respostas) e visualização gráfica (gráficos de frequência). Com o intuito de operacionalizar a análise dos dados coletados e auxiliar os pesquisadores a alcançar o objetivo da avaliação, o modelo MEEGA+ oferece uma planilha de análise de dados padronizada utilizando o Microsoft Excel (versão desktop). Além disso, um guia de instruções de uso das planilhas foi preparado. Este guia descreve o passo a passo para o uso da planilha de análise de dados do modelo MEEGA+, descrevendo como os dados devem ser inseridos e preparados para suportar a geração automática da análise descritiva. O material completo do modelo MEEGA+ está disponível em português, inglês e espanhol, sob *Creative Commons License*, no seguinte endereço: <<http://www.gqs.ufsc.br/meega-a-model-for-evaluating-educational-games/>>.

5 Avaliação do Modelo MEEGA+

De modo a avaliar a confiabilidade e a validade de constructo do instrumento de medição (questionário) do modelo MEEGA+ é realizada uma série de estudos de caso aplicando jogos educacionais para o ensino de computação.



5.1 Definição da Avaliação

Considerando que a confiabilidade e a validade de constructo são indicadores fundamentais em relação a instrumentos de medição como questionários (Carmines & Zeller, 1982; DeVellis, 2016), o objetivo dessa avaliação é analisar o instrumento de medição do modelo MEEGA+ para avaliar sua confiabilidade e validade de constructo do ponto de vista de pesquisadores no contexto de ensino de computação. Os resultados desse estudo são interpretados a partir da perspectiva dos pesquisadores, membros do Grupo de Pesquisa em Qualidade de Software, com formação em computação e estatística.

A confiabilidade refere-se ao grau de consistência dos itens do instrumento no mesmo fator de qualidade. A consistência interna é estimada para avaliar a consistência dos resultados entre os itens de um instrumento de medição (Carmines & Zeller, 1982; Trochim & Donnelly, 2008) por meio do coeficiente alfa de Cronbach (Cronbach, 1951). A validade de constructo é definida como a sua capacidade de medir o que ele pretende medir, envolvendo validade convergente e discriminante, que é obtida por meio do grau de correlação entre os itens do instrumento (Carmines & Zeller, 1982; Trochim & Donnelly, 2008).

Seguindo a abordagem GQM (Basili et al., 1994), o objetivo do estudo é decomposto em aspectos de qualidade e questões a serem analisadas com base na avaliação do instrumento de medição (Carmines & Zeller, 1982; DeVellis, 2016; Trochim & Donnelly, 2008):

Confiabilidade

QA1: Há evidência de consistência interna no instrumento de medição do modelo MEEGA+?

Validade de constructo

QA2: Quantos fatores subjacentes influenciam as respostas dos itens do instrumento de medição do modelo MEEGA+?

QA3: Há evidência de validade convergente e discriminante no instrumento de medição do modelo MEEGA+?

5.2 Execução

Entre julho de 2016 a dezembro de 2017, foi realizada uma série de 48 estudos de caso (38 estudos avaliando jogos não-digitais e 10 estudos avaliando jogos digitais), totalizando 18 diferentes jogos educacionais (4 jogos digitais e 14 jogos não-digitais) de quatro áreas de conhecimento da computação: Engenharia de Software, Fundamentos de Desenvolvimento de Software, Algoritmos e Complexidade e Interação Humano-Computador. Em cada um desses estudos de caso, o instrumento de medição do modelo MEEGA+ foi utilizado após a sessão de jogo (tratamento) para coletar as percepções dos alunos. No total, as respostas de 843 alunos em 6 instituições de ensino do Brasil foram coletadas. Detalhes dos estudos de caso conduzidos são apresentados no Apêndice A. A execução dos estudos de caso foi formalmente aprovada pelo Comitê de Ética da Universidade Federal de Santa Catarina (Certificado N° 1.601.297 / 2016).

Os dados coletados nos estudos de caso foram agrupados em uma única amostra, usando-os cumulativamente para avaliar o modelo MEEGA+ (e não um jogo específico). O agrupamento dos dados foi possível devido à similaridade dos estudos de caso e padronização dos dados coletados pelo modelo MEEGA+. Os estudos de caso são semelhantes em termos de definição (com o objetivo de avaliar um jogo educacional), design de pesquisa (estudos de caso) e contexto (ensino de computação). Além disso, todos os estudos de caso são padronizados em



termos de medidas, método de coleta de dados (questionário MEEGA+) e formato de resposta (escala Likert de 5 pontos).

5.3 Análise

Para avaliar a confiabilidade e validade do modelo MEEGA+ foi realizada uma avaliação estatística. A avaliação é baseada no guia de desenvolvimento de escalas proposto por DeVellis (2016) e nos procedimentos para a avaliação da consistência interna e validade de construto de um instrumento de medição (Trochim & Donnelly, 2008). A análise de dados foi realizada usando o software estatístico IBM SPSS versão 23.

Confiabilidade

QA1: Há evidência de consistência interna no instrumento de medição do modelo MEEGA+?

Avaliação dos itens padronizados

Para responder a esta questão de análise, a consistência interna do instrumento de medição do modelo MEEGA+ foi medida por meio do coeficiente alfa de Cronbach (DeVellis, 2016; Trochim & Donnelly, 2008). O coeficiente alfa de Cronbach indica indiretamente o grau em que um conjunto de itens está medindo um único fator de qualidade. Deste modo, neste estudo, o objetivo é avaliar se o instrumento de medição do modelo MEEGA+ está medindo o mesmo fator de qualidade, que corresponde a percepção da qualidade de um jogo educacional. Tipicamente, um coeficiente alfa de Cronbach entre $0,8 > \alpha \geq 0,7$ é considerado aceitável, entre $0,9 > \alpha \geq 0,8$ é considerado bom, e $\alpha \geq 0,9$ é considerado excelente (DeVellis, 2016), indicando assim a consistência interna do instrumento de medição.

O coeficiente alfa de Cronbach foi analisado separadamente (Tabela 5) considerando a amostra com dados coletados das avaliações apenas de jogos digitais (n=205), com dados coletados apenas de jogos não-digitais (n=638) e agrupando as amostras (n=843), descartando apenas os itens específicos para a avaliação dos jogos digitais (itens 10, 11 e 12).

Tabela 5. Coeficientes alfa de Cronbach

Dimensão/ Subdimensão	Item	Descrição	Alfa de Cronbach se o item for deletado			
			Somente jogos digitais (n=205)	Somente jogos não-digitais (n=638)	Agrupando jogos digitais e não-digitais (n=843)	
Usabilidade	Estética	1	O design do jogo é atraente (interface, gráficos, tabuleiro, cartas, etc.).	,910	,933	,926
		2	Os textos, cores e fontes combinam e são consistentes.	,913	,935	,927
	Aprendizibilidade	3	Eu precisei aprender poucas coisas para poder começar a jogar o jogo.	,913	,936	,928
		4	Aprender a jogar este jogo foi fácil para mim.	,913	,935	,927
		5	Eu acho que a maioria das pessoas aprenderiam a jogar este jogo rapidamente.	,913	,935	,927
	Operabilidade	6	Eu considero que o jogo é fácil de jogar.	,913	,934	,927
		7	As regras do jogo são claras e compreensíveis.	,911	,935	,926
	Acessibilidade	8	As fontes (tamanho e estilo) utilizadas no jogo são legíveis.	,914	,937	,928
		9	As cores utilizadas no jogo são compreensíveis.	,914	,936	,928
		10	O jogo permite personalizar a aparência (fonte e/ou cor) conforme a minha necessidade.	,913	Itens específicos de jogos digitais	
	Proteção contra erros do usuário	11	O jogo me protege de cometer erros.	,913		
		12	Quando eu cometo um erro é fácil de me recuperar rapidamente.	,915		
Confiança	13	Quando olhei pela primeira vez o jogo, eu tive a impressão de que seria fácil para mim.	,914	,935	,928	



	14	A organização do conteúdo me ajudou a estar confiante de que eu iria aprender com este jogo.	,910	,935	,925
Desafio	15	Este jogo é adequadamente desafiador para mim.	,909	,933	,925
	16	O jogo oferece novos desafios (oferece novos obstáculos, situações ou variações) com um ritmo adequado.	,910	,936	,926
	17	O jogo não se torna monótono nas suas tarefas (repetitivo ou com tarefas chatas).	,911	,932	,925
Satisfação	18	Completar as tarefas do jogo me deu um sentimento de realização.	,908	,931	,924
	19	É devido ao meu esforço pessoal que eu consigo avançar no jogo.	,910	,934	,926
	20	Me sinto satisfeito com as coisas que aprendi no jogo.	,909	,934	,924
	21	Eu recomendaria este jogo para meus colegas.	,907	,932	,923
Interação social	22	Eu pude interagir com outras pessoas durante o jogo.	,911	,932	,926
	23	O jogo promove momentos de cooperação e/ou competição entre os jogadores.	,910	,934	,926
	24	Eu me senti bem interagindo com outras pessoas durante o jogo.	,910	,933	,925
Diversão	25	Eu me diverti com o jogo.	,908	,934	,923
	26	Aconteceu alguma situação durante o jogo (elementos do jogo, competição, etc.) que me fez sorrir.	,910	,933	,925
Atenção focada	27	Houve algo interessante no início do jogo que capturou minha atenção.	,909	,935	,925
	28	Eu estava tão envolvido no jogo que eu perdi a noção do tempo.	,909	,935	,925
	29	Eu esqueci sobre o ambiente ao meu redor enquanto jogava este jogo.	,910	,933	,925
Relevância	30	O conteúdo do jogo é relevante para os meus interesses.	,909	,934	,925
	31	É claro para mim como o conteúdo do jogo está relacionado com a disciplina.	,913	,933	,926
	32	O jogo é um método de ensino adequado para esta disciplina.	,901	,933	,925
	33	Eu prefiro aprender com este jogo do que de outra forma (outro método de ensino).	,914	,934	,926
Aprendizagem percebida	34	O jogo contribuiu para a minha aprendizagem na disciplina.	,911	,934	,925
	35	O jogo foi eficiente para minha aprendizagem, em comparação com outras atividades da disciplina.	,912	,934	,925
	Total		,914	,936	,928

Analisando os 35 itens padronizados do instrumento de medição do modelo MEEGA+, o coeficiente alfa de Cronbach é excelente considerando os dados dos jogos digitais ($\alpha = 0,914$). Da mesma forma, o coeficiente do alfa de Cronbach considerando dados de jogos não-digitais também é excelente ($\alpha = 0,936$). Agrupando os dados de jogos digitais e não-digitais (excluindo apenas os itens para avaliar os jogos digitais), o valor do alfa de Cronbach também é considerado excelente ($\alpha = 0,928$). Deste modo, conclui-se que as respostas aos itens estão consistentes e precisas, indicando a confiabilidade dos itens padronizados do instrumento de medição do modelo MEEGA+.

Avaliação dos itens customizados

Os resultados da análise da confiabilidade dos itens customizados, relacionados aos objetivos de aprendizagem específicos de um determinado jogo educacional, mostram que o coeficiente alfa de Cronbach também é excelente para os itens customizados ($\alpha = 0,953$). Isto indica uma excelente confiabilidade também desta parte do instrumento de medição do modelo MEEGA+.

Validade de Constructo

QA2: Quantos fatores subjacentes influenciam as respostas dos itens do instrumento de medição do modelo MEEGA+?

Para identificar o número de fatores de qualidade que representam as respostas dos 35 itens padronizados do instrumento de medição do modelo MEEGA+, foi realizada uma análise fatorial exploratória.



De modo a analisar se os itens do instrumento de medição do modelo MEEGA+ podem ser submetidos a uma análise fatorial, utilizou-se o índice de Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett, sendo os testes mais utilizados para este fim (Brown, 2006). Esses testes indicam quanto a realização de uma análise fatorial é apropriada para o conjunto de itens (Brown, 2006). O índice KMO mede a adequação da amostra com valores entre 0 e 1. Um índice próximo de 1 suporta a análise fatorial, e, qualquer valor inferior a 0,5 indica que a análise fatorial pode não ser adequada para o conjunto de itens (Dziuban & Shirkey, 1974). O teste de esfericidade de Bartlett também indica se a análise fatorial é adequada com valores de nível de significância $<0,05$ sendo considerados aceitáveis (Dziuban & Shirkey, 1974). Analisando o conjunto de itens do instrumento de medição do modelo MEEGA+, obteve-se um índice KMO de 0,918 e um nível de significância de 0,000. Consequentemente, indicando que a análise fatorial é apropriada para analisar o número de fatores que representam as respostas do instrumento de medição do modelo MEEGA+.

Aplicando a análise fatorial, o número de fatores retidos na análise é decidido (Glorfeld, 1995; Brown, 2006). Para esta decisão foi utilizado o teste *scree* de Cattell (Cattell, 1966; Cattell, 1978), uma das estratégias mais utilizadas para determinar o número de componentes a serem retidos em uma análise fatorial (Raiche et al., 2006). O teste *scree* de Cattell consiste na observação do gráfico de *scree* (Figura 2), no qual é apresentado o número de itens do instrumento de medição (aqui chamado de fatores) (eixo x) e seus autovalores correspondentes (eixo y). Por meio da análise do gráfico é possível observar quais fatores apresentam maiores autovalores, sendo assim, responsáveis por uma maior variância explicada. O objetivo é encontrar o ponto com uma inclinação íngreme (comumente chamada de cotovelo) onde os autovalores apresentam uma tendência descecente linear (nivelamento). O intervalo entre a inclinação íngreme e um nivelamento indica o número de fatores significativos para a retenção na análise fatorial. Os fatores após o início do cotovelo devem ser desconsiderados na análise fatorial (Cattell, 1966; Cattell, 1978). Com base nessa definição, o gráfico de *scree* (Figura 2) mostra que a inclinação íngreme da curva (o cotovelo) aparece no terceiro fator. Além disso, a linha pontilhada vermelha (Figura 2) mostra que a queda da curva é menos abrupta (nivelamento) a partir do terceiro fator. Portanto, justificando a retenção de dois fatores na análise fatorial deste estudo.

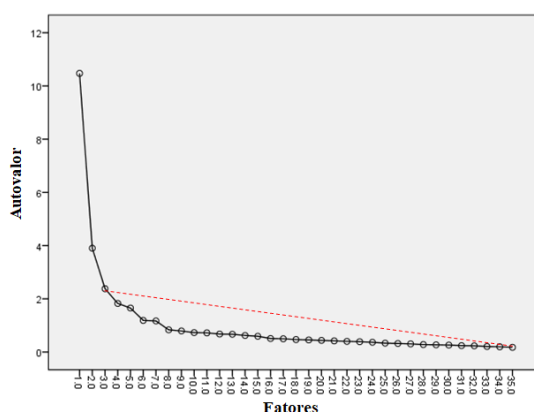


Figura 2. Gráfico de *scree*

No entanto, às vezes, o teste de Cattell é criticado, pois um viés do pesquisador pode ser introduzido devido à subjetividade envolvida na determinação do número de componentes a serem retidos usando um dispositivo de inspeção visual (gráfico *scree*) (Garson, 2013). Para minimizar esse viés, também foi analisado o percentual de variância acumulada dos fatores como um método complementar para a decisão do número de fatores a serem retidos na análise.



Deste modo, os dois fatores retidos na análise fatorial explicam 41,07% da variância acumulada. Essa variância é classificada como adequada para a decisão na retenção dos fatores, considerando 40% o limiar da variância como aceitável (Cuesta, 1996). Com relação ao modelo MEEGA+, isso significa que as respostas do instrumento de medição representam dois conceitos subjacentes (fatores de qualidade).

Uma vez identificado o número de fatores subjacentes, outra questão é determinar quais itens do instrumento de medição contribuem para qual fator. A fim de identificar as cargas fatoriais dos itens, um método de rotação é usado (Brown, 2006; Tabachnick & Fidel, 2007). Para isso, utilizou-se o método de rotação Varimax com normalização de Kaiser, por ser o método de rotação mais amplamente aceito e utilizado (Tabachnick & Fidel, 2007). A Tabela 6 mostra as cargas fatoriais dos itens associado com os dois fatores retidos. A maior carga fatorial de cada item, indicando a qual fator o item está mais relacionado, é marcado em negrito.

Tabela 6. Cargas fatoriais

Dimensão/Subdimensão		Item	Descrição	Fatores	
				1	2
Usabilidade	Estética	1	O design do jogo é atraente (interface, gráficos, tabuleiro, cartas, etc.).	,349	,490
		2	Os textos, cores e fontes combinam e são consistentes.	,160	,494
	Aprendizabilidade	3	Eu precisei aprender poucas coisas para poder começar a jogar o jogo.	,028	,681
		4	Aprender a jogar este jogo foi fácil para mim.	-,001	,793
		5	Eu acho que a maioria das pessoas aprenderiam a jogar este jogo rapidamente.	,059	,757
	Operabilidade	6	Eu considero que o jogo é fácil de jogar.	,080	,794
		7	As regras do jogo são claras e compreensíveis.	,260	,641
	Acessibilidade	8	As fontes (tamanho e estilo) utilizadas no jogo são legíveis.	,007	,555
		9	As cores utilizadas no jogo são compreensíveis.	,060	,506
		10	O jogo permite personalizar a aparência (fonte e/ou cor) conforme a minha necessidade.	,305	-,241
	Proteção contra erros do usuário	11	O jogo me protege de cometer erros.	,240	-,119
		12	Quando eu cometo um erro é fácil de me recuperar rapidamente.	,156	-,111
Confiança	13	Quando olhei pela primeira vez o jogo, eu tive a impressão de que seria fácil para mim.	,067	,584	
	14	A organização do conteúdo me ajudou a estar confiante de que eu iria aprender com este jogo.	,544	,361	
Desafio	15	Este jogo é adequadamente desafiador para mim.	,673	,000	
	16	O jogo oferece novos desafios (oferece novos obstáculos, situações ou variações) com um ritmo adequado.	,703	-,065	
	17	O jogo não se torna monótono nas suas tarefas (repetitivo ou com tarefas chatas).	,621	,073	
Satisfação	18	Completar as tarefas do jogo me deu um sentimento de realização.	,741	,114	
	19	É devido ao meu esforço pessoal que eu consigo avançar no jogo.	,461	,231	
	20	Me sinto satisfeito com as coisas que aprendi no jogo.	,722	,220	
	21	Eu recomendaria este jogo para meus colegas.	,762	,219	
Interação social	22	Eu pude interagir com outras pessoas durante o jogo.	,600	-,057	
	23	O jogo promove momentos de cooperação e/ou competição entre os jogadores.	,622	,018	
	24	Eu me senti bem interagindo com outras pessoas durante o jogo.	,721	,039	
Diversão	25	Eu me diverti com o jogo.	,783	,185	
	26	Aconteceu alguma situação durante o jogo (elementos do jogo, competição, etc.) que me fez sorrir.	,621	,138	
Atenção focada	27	Houve algo interessante no início do jogo que capturou minha atenção.	,664	,128	
	28	Eu estava tão envolvido no jogo que eu perdi a noção do tempo.	,707	,041	
	29	Eu esqueci sobre o ambiente ao meu redor enquanto jogava este jogo.	,676	,050	
Relevância	30	O conteúdo do jogo é relevante para os meus interesses.	,568	,239	
	31	É claro para mim como o conteúdo do jogo está relacionado com a disciplina.	,389	,335	
	32	O jogo é um método de ensino adequado para esta disciplina.	,572	,324	
	33	Eu prefiro aprender com este jogo do que de outra forma (outro método de ensino).	,489	,227	
Aprendizagem percebida	34	O jogo contribuiu para a minha aprendizagem na disciplina.	,635	,250	
	35	O jogo foi eficiente para minha aprendizagem, em comparação com outras atividades da disciplina.	,597	,246	



Analizando as cargas fatoriais dos itens (Tabela 6), pode-se observar que o primeiro fator (fator 1), consiste em um conjunto de 25 itens (itens nº 10 a 12 e 14 a 35), incluindo itens das dimensões: atenção focada, diversão, desafio, interação social, confiança, relevância, satisfação e aprendizagem percebida. Assim, esse resultado parece sugerir que esses itens (e suas dimensões originais), de fato, estão relacionadas e integradas para mensurar a qualidade dos jogos educacionais (como mostra o mapeamento (Tabela 2)). Como esse resultado mostra que essas dimensões estão integradas em um único fator de qualidade. Deste modo, define-se esse fator de qualidade como sendo a experiência do jogador. Assim, neste estudo, define-se a experiência do jogador como um fator de qualidade que abrange um envolvimento profundo do aluno/jogador nas tarefas do jogo, incluindo sua percepção de aprendizagem, sentimentos, prazeres e interações com o jogo, ambiente e outros jogadores (Savi et al., 2011; O'Brien & Toms, 2010; Wiebe et al., 2014; Sweetser & Wyeth, 2005; Fu et al., 2009; Tullis & Albert, 2008; Keller, 1987; ISO/IEC, 2014; Savi et al., 2011; Sindre & Moody, 2003). Além disso, o primeiro fator também inclui os itens 10, 11 e 12, que são específicos para a avaliação de jogos digitais, em termos de customização dos jogos e proteção contra erros do usuário. No entanto, esses itens foram analisados apenas com uma amostra de 205 pontos de dados (amostra de jogos digitais). Assim, os resultados da análise fatorial em termos desses itens podem ser influenciados pelo tamanho da amostra, gerando uma classificação diferente desses itens. Por esse motivo, esses itens precisam ser reanalisados com um tamanho de amostra maior de jogos digitais para confirmar esse resultado.

Com relação ao segundo fator (Tabela 6), um conjunto de 10 itens (itens nº 1 a 9 e 13) está relacionado. Este resultado, claramente, sugere que o fator 2 está relacionado ao conceito de usabilidade, medindo a estética, a capacidade de aprendizado, a operabilidade e a acessibilidade dos jogos educacionais. Embora o item 13 está originalmente relacionado a confiança, o resultado da análise deste item parece indicar que este item está medindo também a usabilidade do jogo (e não a confiança).

Portanto, com base nos resultados da análise fatorial, a estrutura do modelo MEEGA+ resulta em um modelo de avaliação de jogos com dois fatores de qualidade e suas dimensões, conforme apresentado na Figura 3.

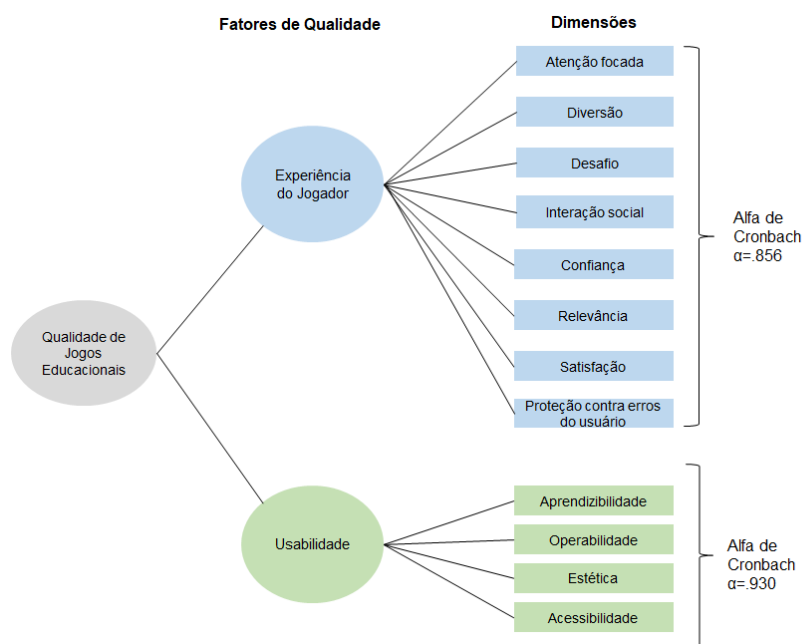


Figura 3. Estrutura do modelo MEEGA+ como resultado da análise fatorial



QA3: Há evidência de validade convergente e discriminante no instrumento de medição do modelo MEEGA+?

De modo a obter evidências da validade convergente e discriminante dos itens padronizados do instrumento de medição do modelo MEEGA+, as correlações dos itens são calculadas (DeVellis, 2016). A validade convergente refere-se ao grau que dois itens de um mesmo fator de qualidade que teoricamente devem estar relacionados, de fato, estão relacionados. Em contrapartida, validade discriminante testa se conceitos ou medidas que supostamente não devem ter relação, de fato, não estão relacionados (Carmines & Zeller, 1982; Trochim & Donnelly, 2008).

Intercorrelação dos itens padronizados

Para obter evidência de validade convergente é esperado que os itens do mesmo fator de qualidade (ex., usabilidade ou experiência do jogador) tenham uma alta correlação (Carmines & Zeller, 1982; Trochim & Donnelly, 2008). Por outro lado, para obter evidência de validade discriminante é esperado que os itens de diferentes fatores de qualidade tenham uma baixa correlação (Carmines & Zeller, 1982; Trochim & Donnelly, 2008). Por exemplo, é esperado que os itens de diferentes fatores de qualidade (ex., usabilidade e experiência do jogador) tenham uma baixa correlação, pois em teoria, os itens estão medindo diferentes fatores de qualidade.

Para analisar a correlação entre os itens, utilizou-se uma matriz de correlação não-paramétrica de Spearman para cada fator de qualidade (Tabelas 7 e 8). As matrizes mostram o coeficiente de correlação de Spearman, indicando o grau de correlação entre dois itens (par de itens). Este coeficiente é utilizado por ser o coeficiente mais apropriado para análises de escalas Likert (Chen & Popovich, 2002). Os coeficientes de correlação entre os itens de um mesmo fator de qualidade estão marcados em cores. De acordo com Cohen (Cohen, 1988), uma correlação entre itens é considerada satisfatória se o coeficiente de correlação for maior que 0,29, indicando que há uma correlação moderada, ou uma correlação alta se o coeficiente for maior que 0,50 (Cohen, 1988). Uma baixa correlação entre os itens é considerada se o coeficiente estiver próximo de 0,10. Coeficientes de correlação satisfatórios estão destacados em negrito.

Tabela 7. Matriz de correlação de Spearman do fator de qualidade usabilidade

Item/ Dimensão	1 Estética	2 Estética	3 Aprendizibilidade	4 Aprendizibilidade	5 Aprendizibilidade	6 Operabilidade	7 Operabilidade	8 Acessibilidade	9 Acessibilidade	13
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	
13	,27	,19	,37	,44	,45	,43	,32	,28	,24	1,00

Analisando as correlações dos itens do fator de qualidade usabilidade (Tabela 7), pode-se observar que a maioria dos pares de itens apresenta uma correlação moderada ($>0,29$) ou alta ($>0,50$). Este resultado indica que, embora a usabilidade esteja fragmentada em suas dimensões, os itens apresentam uma correlação satisfatória para medir o que se propõe a medir (a usabilidade dos jogos educacionais). Em relação ao item 13 (originalmente relacionado a dimensão confiança), os resultados mostram que este item tem uma correlação moderada com itens das dimensões de aprendizagem e operabilidade. Assim, de fato, o item 13 parece estar medindo também a usabilidade do jogo. Por outro lado, poucos pares de itens (11) apresentaram correlação insatisfatória de acordo com o coeficiente de Cohen. Mesmo assim, esses pares de



itens têm um coeficiente de correlação muito próximo ao nível de correlação moderada. Portanto, com base nos coeficientes de correlação, pode-se estabelecer uma validade convergente no fator de qualidade usabilidade.

Com relação ao fator qualidade da experiência do jogador (Tabela 8), pode-se observar novamente que a maioria dos pares de itens apresenta um coeficiente de correlação moderado ou alto. Esse resultado indica que, de fato, a experiência do jogador em jogos educacionais é medida em termos de confiança, desafio, satisfação, interação social, diversão, atenção focada, relevância e aprendizagem percebida, como sugerido pelos altos coeficientes de correlação entre os itens.

No entanto, conforme mostra a Tabela 8, também pode-se observar que os itens 10, 11 e 12 apresentam uma pequena correlação com outros itens desse fator de qualidade. Assim, embora o resultado da análise fatorial indique que esses itens compõem o fator qualidade da experiência do jogador, os coeficientes de correlação parecem indicar que esses itens não estão medindo a experiência do jogador, quando comparados a outros itens desse fator de qualidade. Uma explicação para esse resultado pode ser que o tamanho da amostra da análise desses itens (n=205) é menor que o tamanho da amostra dos demais itens (n=843) e, assim, pode ter influenciado os resultados. Portanto, esses itens (10, 11 e 12) devem ser reanalisados com uma amostra maior para confirmar se esses itens, de fato, estão compondo o fator qualidade da experiência do jogador ao avaliar jogos para o ensino de computação. No entanto, considerando a maioria dos coeficientes de correlação moderada e alta, pode-se observar que há evidências de validade convergente no fator qualidade da experiência do jogador.

A fim de obter evidência de validade discriminante, analisou-se os coeficientes de correlação dos diferentes fatores de qualidade. Nesse caso, é avaliado se os itens de diferentes fatores de qualidade têm uma pequena correlação (Carmines & Zeller, 1982; Trochim & Donnelly, 2008). Como resultado dessa análise observou-se que a maioria dos pares de itens de diferentes fatores de qualidade (por exemplo, 3-22 e 9-28) apresenta um baixo coeficiente de correlação, indicando assim, evidência de validade discriminante.

Tabela 8. Matriz de correlação de Spearman do fator de qualidade experiência do jogador

Item	10	11	12	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		
		Proteção contra erros do usuário		Confiança		Desafio			Satisfação			Interação social		Diversão		Atenção focada			Relevância				Aprendizagem percebida				
10	1,0																										
11	,54	1,0																									
12	,26	,42	1,0																								
14	,04	,10	,10	1,0																							
15	,14	,15	,10	,40	1,0																						
16	,18	,17	,17	,36	,59	1,0																					
17	,09	,09	,06	,27	,46	,50	1,0																				
18	,12	,09	,09	,40	,48	,47	,46	1,0																			
19	,02	,03	,02	,38	,36	,31	,27	,46	1,0																		
20	,13	,10	,08	,48	,44	,44	,35	,56	,48	1,0																	
21	,16	,19	,10	,47	,44	,43	,46	,57	,39	,68	1,0																
22	,21	,14	,11	,28	,32	,34	,34	,37	,20	,37	,40	1,0															
23	,15	,13	,03	,29	,37	,39	,37	,45	,26	,39	,42	,62	1,0														
24	,21	,17	,06	,36	,39	,40	,40	,48	,26	,45	,49	,69	,69	1,0													
25	,09	,09	,04	,41	,39	,43	,48	,58	,31	,49	,62	,50	,56	,63	1,0												
26	,09	,06	,01	,31	,32	,30	,39	,45	,22	,38	,47	,43	,44	,55	,67	1,0											
27	,13	,13	,01	,38	,36	,38	,34	,43	,26	,40	,50	,31	,33	,42	,53	,46	1,0										
28	,16	,13	,04	,38	,36	,38	,43	,49	,25	,38	,49	,35	,40	,48	,58	,49	,58	1,0									
29	,17	,14	,08	,28	,31	,35	,38	,47	,24	,35	,48	,39	,40	,48	,53	,44	,51	,76	1,0								
30	,06	,06	,01	,39	,38	,29	,29	,45	,36	,52	,51	,27	,30	,34	,44	,29	,36	,34	,34	1,0							
31	-,05	-,01	-,03	,29	,25	,20	,20	,27	,30	,39	,34	,22	,25	,27	,34	,29	,22	,20	,18	,48	1,0						
32	-,01	,03	,01	,43	,33	,31	,27	,39	,37	,50	,48	,22	,29	,34	,45	,37	,34	,29	,30	,48	,65	1,0					
33	,02	,05	-,04	,29	,26	,24	,26	,31	,29	,38	,38	,20	,25	,30	,38	,31	,28	,26	,32	,32	,41	,62	1,0				
34	,03	,01	-,01	,46	,37	,39	,26	,43	,37	,58	,49	,26	,30	,34	,44	,33	,36	,33	,30	,53	,53	,65	,53	1,0			
35	-,03	,00	,02	,41	,34	,37	,28	,41	,36	,53	,47	,22	,26	,33	,44	,32	,34	,31	,28	,44	,43	,56	,55	,73	1,0		

Em síntese, pode-se observar que há uma correlação entre os itens dentro de um mesmo fator de qualidade. Isso indica que a validade convergente pode ser estabelecida para os dois fatores de qualidade (usabilidade e experiência do jogador). Por outro lado, poucos pares de itens (principalmente relacionados aos itens 11, 12 e 13) apresentam correlações baixas ou negativas, o que indica que esses itens precisam ser revisados e/ou, principalmente, reanalisados



com uma amostra maior, em relação aos jogos digitais, para confirmar este resultado. Em geral, itens de diferentes fatores de qualidade apresentaram uma baixa correlação, assim, também pode-se identificar evidência de validade discriminante. Com base nesses resultados, pode-se observar que o instrumento de medição do modelo MEEGA+, de fato, mede o que ele pretende medir (a qualidade dos jogos para o ensino de computação em termos de usabilidade e experiência do jogador).

6 Discussão

Os resultados da análise do instrumento de medição do modelo MEEGA+ mostram evidências suficientes para considerar a confiabilidade e a validade de constructo do MEEGA+ aceitável como um modelo para a avaliação da qualidade de jogos usados para o ensino de computação. Em termos de confiabilidade (QA1), os resultados da análise indicam um coeficiente alfa de Cronbach excelente para todos os fatores de qualidade (alfa de Cronbach $\alpha = 0,928$), indicando a consistência interna do instrumento de medição do modelo MEEGA+. Além disso, comparando a confiabilidade com a versão inicial do modelo MEEGA (alfa de Cronbach $\alpha = .915$) (Petri et al., 2017), obteve-se um aumento na confiabilidade com o instrumento de medição do modelo MEEGA+. Assim, indicando que os itens do instrumento de medição do modelo MEEGA+ são consistentes e precisos em relação à avaliação da qualidade dos jogos e que há uma evolução em relação a versão inicial do modelo. Também comparando nossos resultados com a confiabilidade relatada por algumas das abordagens relacionadas, a avaliação do MEEGA+ demonstrou uma confiabilidade maior do que a apresentada em cada fator da escala EGameFlow (alfa de Cronbach $\alpha = 0,80$) (Fu et al., 2009).

Em termos de validade de constructo, com base nos resultados da análise fatorial exploratória (QA2), identificou-se que as respostas do instrumento de medição do modelo MEEGA+ são explicadas por dois conceitos subjacentes (fatores de qualidade), representando 41,07% da variância acumulada. Este resultado indica que a qualidade dos jogos para o ensino de computação é avaliada por meio de dois fatores de qualidade (experiência do jogador e usabilidade). O primeiro fator que mede a experiência do jogador consiste em um conjunto de 25 itens das dimensões de atenção focada, diversão, desafio, interação social, confiança, relevância, satisfação e aprendizagem percebida. Além disso, esse fator também inclui os itens 10, 11 e 12, específicos para a avaliação dos jogos digitais. No entanto, como esses itens apresentaram apenas uma baixa correlação com outros itens desse fator, indicando que esses itens podem não estar medindo a experiência do jogador, esses itens precisam ser reanalisados com um tamanho de amostra maior. Itens relacionados ao fator 2 que medem a usabilidade do jogo, medem as dimensões de estética, aprendizibilidade, operabilidade e a acessibilidade dos jogos educacionais.

Diferente dos resultados da análise fatorial da versão inicial do modelo MEEGA (Petri et al., 2017), que indicaram uma sobreposição conceitual em termos de motivação e experiência do usuário, nossos resultados indicam uma estrutura bem definida composto por dois fatores de qualidade (experiência do jogador e usabilidade) (Figura 3), a fim de avaliar jogos para o ensino de computação. Além disso, em comparação com as abordagens relacionadas, que normalmente usam uma ampla variedade de fatores e alguns deles se sobrepõem conceitualmente, nossos resultados fornecem evidências de uma estrutura conceitual bem definida. As dimensões foram definidas seguindo uma tendência identificada em uma revisão de literatura realizada previamente (Petri & Gresse von Wangenheim, 2017), e um estudo de mapeamento, minimizando, assim, a sobreposição de conceitos.



Analisando os coeficientes de correlação entre os itens dos dois fatores de qualidade (QA3), pode-se observar que, em geral, existe uma correlação moderada e alta entre a maioria dos itens dentro de cada fator de qualidade do modelo MEEGA+. Isso indica que uma validade convergente pode ser estabelecida para os dois fatores de qualidade (usabilidade e experiência do jogador). Da mesma forma, itens de diferentes fatores de qualidade apresentam uma baixa correlação e, assim, evidenciam a validade discriminante. Além disso, comparado com a versão inicial do modelo MEEGA, novamente, o modelo MEEGA+ apresenta uma evolução, pois, embora o modelo MEEGA tenha demonstrado evidências de validade convergente, não apresentou evidências de validade discriminante, indicando uma sobreposição conceitual (Petri et al., 2017). Em contraste, no que diz respeito ao modelo MEEGA+, pode-se identificar evidências de validade convergente e discriminante, apresentando uma estrutura bem definida e correlacionada, composta por dois fatores de qualidade (experiência do jogador e usabilidade).

Portanto, com base nesses resultados, conclui-se que o instrumento de medição do modelo MEEGA+ possui alta confiabilidade e mede o que pretende medir: a qualidade de jogos para o ensino de computação em termos de usabilidade e experiência do jogador.

6.1 Ameaças à validade

Devido as características deste tipo de pesquisa, esse trabalho está sujeito a ameaças à validade. Assim, foram identificadas possíveis ameaças e aplicadas estratégias de mitigação de modo a minimizar o seu impacto nesta pesquisa. Algumas ameaças estão relacionadas ao tipo de estudo. Para mitigar essa ameaça, foi definido e documentado uma metodologia sistemática para o estudo. O modelo MEEGA+ foi definido adotando a abordagem GQM, decompondo sistematicamente o objetivo da avaliação em fatores para serem medidos por um instrumento de medição. O instrumento de medição foi desenvolvido seguindo métodos de desenvolvimento de escalas e questionários. Além disso, para a avaliação do instrumento de medição do modelo MEEGA+, uma série de estudos de caso foi sistematicamente definida e documentada.

Outro risco refere-se à qualidade dos dados agrupados em uma única amostra para avaliação, em termos de padronização de dados (formato de resposta) e adequação ao modelo MEEGA+. Como este estudo se limita exclusivamente às avaliações que utilizaram o modelo MEEGA+, esse risco é minimizado, pois em todos os estudos foi utilizado o mesmo instrumento de coleta de dados. Outra questão refere-se aos dados agrupados de diferentes contextos. Para minimizar essa ameaça, todos os estudos de caso de avaliação de jogos foram conduzidos em um mesmo contexto: ensino superior de computação.

Uma limitação do estudo refere-se à avaliação da aprendizagem. Adotando um design de pesquisa não-experimental (estudos de caso), apenas um pós-teste usando autoavaliação foi aplicado para avaliar a aprendizagem percebida pelos alunos. Um pré-teste não foi aplicado e, portanto, não foi possível avaliar o impacto na aprendizagem promovida pelos jogos. Entretanto, em relação à autoavaliação, apesar de não haver consenso, há evidências de que a autoavaliação fornece informações confiáveis, válidas e úteis para esse tipo de estudo (Sitzmann et al., 2010; Thomas et al., 2011; Sharma et al., 2010), principalmente quando se utiliza um modelo de avaliação sistemático, confiável e válido como o MEEGA+.

Em termos de validade externa, uma ameaça a possibilidade de generalizar os resultados está relacionada ao tamanho da amostra e a diversidade dos dados utilizados para a avaliação. Em relação ao tamanho da amostra, a avaliação utilizou dados coletados de 48 estudos de caso que avaliaram 18 diferentes jogos educacionais, envolvendo uma população de 843 alunos de 6 diferentes instituições de ensino. Em termos de significância estatística, este é um tamanho de



amostra considerado satisfatório, permitindo a geração de resultados significativos (Kasunic, 2005, Devellis, 2016).

Em termos de confiabilidade, uma ameaça refere-se a quanto os dados e a análise dependem dos pesquisadores específicos. Para mitigar essa ameaça, foi sistematicamente documentada a avaliação do modelo MEEGA+, definindo claramente o objetivo do estudo, o processo de coleta de dados e os métodos estatísticos utilizados para a análise dos dados. Outra questão refere-se à escolha correta dos testes estatísticos para análise dos dados. Para minimizar essa ameaça, foi realizada uma avaliação estatística baseada no guia de construção de escalas proposta por DeVellis (2016), que está alinhada aos procedimentos de avaliação da consistência interna e validade de construto de um instrumento de medição (Trochim & Donnelly, 2008).

7 Conclusões

Esse artigo apresenta o desenvolvimento e a avaliação do modelo MEEGA+, como uma evolução do modelo MEEGA proposto por Savi et al. (2011), a fim de avaliar a qualidade de jogos educacionais usados para o ensino de computação. Seguindo uma metodologia sistemática, o modelo MEEGA+ foi desenvolvido com base em uma revisão de literatura e na análise da versão inicial do modelo MEEGA. Um conjunto de fatores foi definido e um instrumento de medição foi desenvolvido para ser usado na avaliação de jogos educacionais usados em estudos com um design de pesquisa não-experimental com pós-teste.

O modelo MEEGA+ foi avaliado em termos de confiabilidade e validade de construto com base em uma amostra de dados de 843 participantes em 48 estudos de casos que avaliaram 18 diferentes jogos digitais e não-digitais. Os resultados indicam que o instrumento de medição do modelo MEEGA+ possui confiabilidade e validade de construto satisfatórias. Com relação à confiabilidade, um alfa de Cronbach = 0,928 indica uma excelente consistência interna, o que significa que as respostas entre os itens são consistentes e precisas. Os resultados da análise fatorial exploratória indicam que a qualidade dos jogos para o ensino de computação é avaliada por meio de dois fatores de qualidade (experiência do jogador e usabilidade). Além disso, os resultados mostram evidências de validade convergente apresentando de um grau satisfatório de correlação encontrado entre a maioria dos itens dos fatores de qualidade. Assim, indicando que o modelo MEEGA+ é confiável e válido para avaliar a qualidade de jogos para o ensino de computação.

E, embora a ênfase deste trabalho está na avaliação de jogos usados para o ensino de computação, o modelo MEEGA+ pode ser usado e adaptado para a avaliação de jogos digitais e não-digitais para o ensino em outras áreas de conhecimento, como já realizado por Gomes (2016), Herpich et al. (2017), Pereira et al. (2017), Silva et al. (2017a), Silva et al. (2017b), e Moosa, Al-Maadeed, and AlJa'am (2018). No entanto, quando transferido o modelo para outra área de conhecimento, estudos empíricos adicionais são necessários de modo a confirmar a validade e confiabilidade do modelo MEEGA+ também nessas áreas de conhecimento.

Agradecimentos

Um agradecimento a todos os alunos e professores que participaram dos estudos de caso de avaliação de jogos. Este trabalho é apoiado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), uma entidade do governo brasileiro focada no desenvolvimento científico.



Referências

- Abdellatif, A. J., McCollum, B., & McMullan, P. (2018). Serious Games: Quality Characteristics Evaluation Framework and Case Study. *Proc. of IEEE Integrated STEM Education Conference*, (pp. 112-119). Princeton, USA. doi: [10.1109/ISECon.2018.8340460](https://doi.org/10.1109/ISECon.2018.8340460) [GS Search]
- Abt, C. C. (2002). *Serious Games*. Lanhan: University Press of America.
- ACM/IEEE-CS. (2013). *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science, 2013*. Disponível em: <<https://www.acm.org/education/CS2013-final-report.pdf>> Access: 06 Nov. 2018.
- Ak, O. (2012). A Game Scale to Evaluate Educational Computer Games. *Procedia – Social and Behavioral Sciences*, 46, 2477-2481. doi: [10.1016/j.sbspro.2012.05.506](https://doi.org/10.1016/j.sbspro.2012.05.506) [GS Search]
- All, A., Castellar, E. P. N., & Looy, J. V. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, 92–93, 90-103. doi: [10.1016/j.compedu.2015.10.007](https://doi.org/10.1016/j.compedu.2015.10.007) [GS Search]
- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. Longman. [GS Search]
- Andrade, H. & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48, 12-19. doi: [10.1080/00405840802577544](https://doi.org/10.1080/00405840802577544) [GS Search]
- Backlund, P. & Hendrix, M. (2013). Educational games - Are they worth the effort? A literature survey of the effectiveness of serious games. *Proc. of the 5th Int. Conf. on Games and Virtual Worlds for Serious Applications*, Poole, GB. doi: [10.1109/VS-GAMES.2013.6624226](https://doi.org/10.1109/VS-GAMES.2013.6624226) [GS Search]
- Basili, V. R., Caldiera, G., & Rombach, H. D. (1994). Goal, Question Metric Paradigm. In J. J. Marciniak, *Encyclopedia of Software Engineering*, (pp. 528-532). New York: Wiley-Interscience. [GS Search]
- Battistella, P. & Gresse von Wangenheim, C. (2016). Games for teaching computing in higher education – a systematic review. *IEEE Technology and Engineering Education Journal*, 9(1), 8-30. [GS Search]
- Battistella, P. E., Gresse von Wangenheim, C., von Wangenheim, A., & Martina, J. E. (2017). Design and Large-scale Evaluation of Educational Games for Teaching Sorting Algorithms. *Informatics in Education*, 17(2), 141-164. doi: [10.15388/infedu.2017.08](https://doi.org/10.15388/infedu.2017.08) [GS Search]
- Beecham, S., Hall, T., Britton, C., Cottee, M., & Rainer, A. (2005). Using an Expert Panel to Validate a Requirements Process Improvement Model. *Journal of Systems and Software*, 76(3), 251-275. doi: [10.1016/j.jss.2004.06.004](https://doi.org/10.1016/j.jss.2004.06.004) [GS Search]
- Bittencourt, I. I. & Isotani, S. (2018). Informática na Educação baseada em Evidências: Um Manifesto. *Revista Brasileira de Informática na Educação - RBIE*, 26(3), 108-119. doi: [10.5753/RBIE.2018.26.03.108](https://doi.org/10.5753/RBIE.2018.26.03.108) [GS Search]
- Bowman, D. D. (2018). Declining Talent in Computer Related Careers. *Journal of Academic Administration in Higher Education*, 14(1), 1-4. [GS Search]
- Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (2016). An update to the systematic literature review of empirical



evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, (94), 178-192. doi: [10.1016/j.compedu.2015.11.003](https://doi.org/10.1016/j.compedu.2015.11.003) [GS Search]

Branch, R. M. (2010). *Instructional Design: The ADDIE Approach*. New York: Springer New York Dordrecht Heidelberg London. [GS Search]

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4-7. [GS Search]

Brown, G., Andrade, H., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education Principles Policy and Practice*, 22(4), 1-26. doi: [10.1080/0969594X.2014.996523](https://doi.org/10.1080/0969594X.2014.996523) [GS Search]

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press. [GS Search]

Budgen, D., Turner, M., Brereton, P., & Kitchenham, B. (2008). Using mapping studies in Software Engineering. *Evidence-based Software Engineering*. Lancaster University, 195-204. [GS Search]

Calderón, A. & Ruiz M. (2015). A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education*, 87, 396-422. doi: [10.1016/j.compedu.2015.07.011](https://doi.org/10.1016/j.compedu.2015.07.011) [GS Search]

Calderón, A., Ruiz M., & O'Connor, R. (2018). A multivocal literature review on serious games for software process standards education. *Computer Standards & Interfaces*, 57, 36-48. doi: [10.1016/j.csi.2017.11.003](https://doi.org/10.1016/j.csi.2017.11.003) [GS Search]

Caulfield, C., Xia, J., Veal, D., & Maj, S. P. (2011). A systematic survey of games used for software engineering education. *Modern Applied Science*, 5(6), 28-43. doi: [10.5539/mas.v5n6p00](https://doi.org/10.5539/mas.v5n6p00) [GS Search]

Carmines, E. G., & Zeller, R. A. (1982). *Reliability and validity assessment* (5th ed.). Beverly Hills: Sage Publications Inc. [GS Search]

Carvalho, C. V. (2012). Is game-based learning suitable for engineering education? *Proc. of the Global Engineering Education Conf.*, (pp.1-8). Marrakech, Morocco. doi: [10.1109/EDUCON.2012.6201140](https://doi.org/10.1109/EDUCON.2012.6201140) [GS Search]

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

Cattell, R. B. (1966). The Scree Test for the Number of Factors. *Multivariate Behavioral Research*, 1(2), 245-276. [GS Search]

Chen, P. Y. & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks: Sage Publications. [GS Search]

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge Academic. [GS Search]

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661-686. doi: [10.1016/j.compedu.2012.03.004](https://doi.org/10.1016/j.compedu.2012.03.004) [GS Search]

Connolly, T. M., Stansfield, M. H., & Hainey, T. (2009). Towards the development of a games-based learning evaluation framework. In T. M. Connolly, M. H. Stansfield, & E. Boyle (Eds.),



Games-based learning advancement for multisensory human computer interfaces: Techniques and effective practices. Hershey: Idea-Group Publishing. [[GS Search](#)]

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. [[GS Search](#)]

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

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340. doi: [10.2307/249008](https://doi.org/10.2307/249008) [[GS Search](#)]

Dawes, J. (2008). 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*, 50(1), 61-77. [[GS Search](#)]

Dempsey, J. V., Lucassen, B. A., Haynes, L. L., & Casey, M. S. (1996). *Instructional applications of computer games*. Paper presented at the American Education Research Association, NY. [[GS Search](#)]

DeVellis, R. F. (2016). *Scale development: theory and applications* (4th. ed.). Thousand Oaks: SAGE Publications. [[GS Search](#)]

Djaouti, D., Alvarez J., Jessel J. P., & Rampoux O. (2011). Origins of Serious Games. In: Ma M., Oikonomou A., Jain L. (Eds). *Serious Games and Edutainment Applications*. London: Springer. [[GS Search](#)]

Dziuban, C. D. & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, 81, 358-361. doi: [10.1037/h0036316](https://doi.org/10.1037/h0036316) [[GS Search](#)]

Freitas, S. D., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, 46(3), 249-264. doi: [10.1016/j.compedu.2005.11.007](https://doi.org/10.1016/j.compedu.2005.11.007) [[GS Search](#)]

Freire M., Serrano-Laguna Á., Iglesias B. M., Martínez-Ortiz I., Moreno-Ger P., & Fernández-Manjón B. (2016). Game Learning Analytics: Learning Analytics for Serious Games. In: Spector M., Lockee B., Childress M. (Eds) *Learning, Design, and Technology*. Cham: Springer. [[GS Search](#)]

Fu, F., Su, R., & Yu, S. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, 52(1), 101-112. doi: [10.1016/j.compedu.2008.07.004](https://doi.org/10.1016/j.compedu.2008.07.004) [[GS Search](#)]

Gámez, E. H. (2009). *On the Core Elements of the Experience of Playing Video Games* (Dissertation). UCL Interaction Centre, Department of Computer Science, London.

Garson, D. (2013). *Factor Analysis*. Raleigh: Statistical Associates Publishing.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55(3), 377-393. doi: [10.1177/0013164495055003002](https://doi.org/10.1177/0013164495055003002) [[GS Search](#)]

Gomes, M. N. (2016). Desenvolvimento de uma Atividade Gamificada Voltada ao Turismo de Caxias Do Sul. Trabalho de Conclusão de Curso (Curso de Ciência da Computação), Universidade de Caxias do Sul. [[GS Search](#)]



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. doi: [10.1016/j.jss.2013.05.030](https://doi.org/10.1016/j.jss.2013.05.030) [GS Search]

Herpich, F., Nunes, F. B., Voss, G. B., Sindeaux, P., Tarouco, L. M. R., & Lima, J. V. (2017). Realidade Aumentada em Geografia: uma atividade de orientação no ensino fundamental. *Revista Novas Tecnologias na Educação*, 15(2), 1–10. [GS Search]

International Standard Organization (ISO). (2014). ISO/IEC 25010: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models, Technical Report.

Johnson, R. B. & Christensen, L. (2014). *Educational Research – Quantitative, Qualitative, and Mixed Approaches* (5th ed.). Thousand Oaks: Sage Publications, Inc. [GS Search]

Kasunic, M. (2005). Designing an effective survey. Handbook CMU/SEI-2005-HB-004, Software Engineering Institute. Pittsburgh: Carnegie Mellon University. [GS Search]

Keller, J. (1987). Development and Use of the ARCS Model of motivational Design. *Journal of Instructional Development*, 10(3), 2-10. doi: [10.1007/BF02905780](https://doi.org/10.1007/BF02905780) [GS Search]

Kosa, M., Yilmaz, M., O'Connor, R., & Clarke, P. (2016). Software engineering education and games: a systematic literature review. *Journal of Universal Computer Science*, 22(12), 1558-1574. doi: [10.3217/jucs-022-12-1558](https://doi.org/10.3217/jucs-022-12-1558) [GS Search]

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1973). *Taxonomy of Educational Objectives, the Classification of Educational Goals*. Handbook II: Affective Domain. Philadelphia: David McKay Company, Inc.

Malhotra, N. K., & Birks, D. F. (2008). *Marketing Research: An Applied Approach* (3rd ed.). Philadelphia: Trans-Atlantic Publications. [GS Search]

Mayer, I. (2012). Towards a Comprehensive Methodology for the Research and Evaluation of Serious Games. *Procedia Computer Science*, 15, 233-247. doi: [10.1016/j.procs.2012.10.075](https://doi.org/10.1016/j.procs.2012.10.075) [GS Search]

Mohamed, H. & Jaafar, A. (2010). Development and Potential Analysis of Heuristic Evaluation for Educational Computer Game (PHEG). *Proc. of the 5th Int. Conf. on Computer Sciences and Convergence Information Technology*, (pp. 222-227). Seoul, South Korea. doi: [10.1109/ICCIT.2010.5711061](https://doi.org/10.1109/ICCIT.2010.5711061) [GS Search]

Moosa, A. M., Al-Maadeed, N., & Alja'am, J. M. (2018). A Simple Health-Based Game for Children. *Proc. of the International Conference on Computer and Applications*. Beirut, Líbano, p. 309-312. doi: [10.1109/COMAPP.2018.8460213](https://doi.org/10.1109/COMAPP.2018.8460213) [GS Search]

O'Brien, H. L. & Toms, E. G. (2010). 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. doi: [10.1002/asi.21229](https://doi.org/10.1002/asi.21229) [GS Search]

Pereira, C. X., Hoffmann, Z., Castro, C. P., Santos, G. S., Aires, T. A., & Francisco, R. E. (2017). CHEMIS3: A game for learning chemical concepts through elements of nature. *Proc. of the 12nd Iberian Conference on Information Systems and Technologies*, Lisboa, Portugal, p. 1-5. doi: [10.23919/CISTI.2017.7975801](https://doi.org/10.23919/CISTI.2017.7975801) [GS Search]

Petri, G. & Chiavegatti, N. C. (2015). Um Role-Playing Game para o Ensino de Elicitação e Análise de Requisitos. *Revista Novas Tecnologias na Educação*, 13(1), 1-10. doi: [10.22456/1679-1916.57599](https://doi.org/10.22456/1679-1916.57599) [GS Search]



Petri, G. & Gresse von Wangenheim, C. (2016). How to evaluate educational games: a systematic literature review. *Journal of Universal Computers Science*, 22(7), 992-1021. doi: [10.3217/jucs-022-07-0992](https://doi.org/10.3217/jucs-022-07-0992) [GS Search]

Petri, G. & Gresse von Wangenheim, C. (2017). How games for computing education are evaluated? A systematic literature review. *Computers & Education*, 107, 68-90. doi: [10.1016/j.compedu.2017.01.004](https://doi.org/10.1016/j.compedu.2017.01.004) [GS Search]

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. *Anais do 27º Simpósio Brasileiro de Informática na Educação*, pp. 320-329. Uberlândia/MG, Brasil. Doi: [10.5753/cbie.sbie.2016.320](https://doi.org/10.5753/cbie.sbie.2016.320) [GS Search]

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. *Proc. of the IEEE/ACM 39th Int. Conf. on Software Engineering: Software Engineering Education and Training Track*, (pp. 180-189). Buenos Aires/Argentina. doi: [10.1109/ICSE-SEET.2017.11](https://doi.org/10.1109/ICSE-SEET.2017.11) [GS Search]

Petri, G., Gresse von Wangenheim, C., & Borgatto, A. F. (2018). MEEGA+, Systematic Model to Evaluate Educational Games. In Newton Lee (Eds) *Encyclopedia of Computer Graphics and Games*, (pp. 1-7). Cham: Springer. doi: [10.1007/978-3-319-08234-9](https://doi.org/10.1007/978-3-319-08234-9) [GS Search]

Poels, K., Kort, Y. D., & Ijsselsteijn, W. (2007). It is always a lot of fun!: exploring dimensions of digital game experience using focus group methodology. *Proc. of Conf. on Future Play*, (pp. 83-89). Toronto, Canada. doi: [10.1145/1328202.1328218](https://doi.org/10.1145/1328202.1328218) [GS Search]

Prensky, M. (2007). *Digital Game-Based Learning*. New York: Paragon House. [GS Search]

Raiche, G., Riopel, M., & Blais, J.-G. (2006). Nongraphical solutions for the Cattell's scree test. *Proc. of the Annual meeting of the Psychometric Society*. Montreal, Canada. [GS Search]

Ritterfeld, U., Cody, M., & Vorderer, P. (Eds.). (2010). *Serious Games*. New York: Routledge.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation*, 11(10), 1-13. [GS Search]

Savi, R., Gresse von Wangenheim, C., & Borgatto, A. F. (2011). A model for the evaluation of educational games for teaching software engineering. *Anais do XXV Simpósio Brasileiro de Engenharia de Software*, pp. 194-203. São Paulo, Brasil. doi: [10.1109/SBES.2011.27](https://doi.org/10.1109/SBES.2011.27) [GS Search]

Schanzenbach, D. W. (2012). Limitations of Experiments in Education Research. *Education Finance and Policy*, 7(2), 219-232. doi: [10.1162/EDFP_a_00063](https://doi.org/10.1162/EDFP_a_00063) [GS Search]

Sharma, R., Jain, A., Gupta, N., Garg, S., Batta, M., & Dhir, S. K. (2016). Impact of self-assessment by students on their learning. *International Journal of Applied and Basic Medical Research*, 6(3), 226-229. doi: [10.4103/2229-516X.186961](https://doi.org/10.4103/2229-516X.186961) [GS Search]

Serrano-Laguna, A., Manero, B., Freire, M., & Fernández-Manjón, B. (2018). A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes. *Multimedia Tools and Applications*, 77(2), 2849-2871. [GS Search]

Silva, J. P., Carvalho, F. C., Silvestre, A., Luiz, M. F., Alencar, R. L., & Silveira, I. F. (2017). Força dos Sinônimos: Um Jogo da Força Multiplayer para o ensino de Tesouros em Língua Portuguesa como Ferramenta de busca na Web. *Anais do Simpósio Brasileiro de Jogos e Entretenimento Digital*, Curitiba/PR, Brasil, p. 1128-1131. [GS Search]



- Silva, J. P., Jorge, A. A., Gaia, C. C., & Costa, G. C. (2017). Quais as chances? Um jogo de dados e cartas para o ensino do cálculo de probabilidades. *Anais do 8º Congresso de Inovação, Ciência e Tecnologia do IFSP*, Cubatão/SP, Brasil, p. 1-3. [[GS Search](#)]
- Simpson, E. J. (1972). *The classification of educational objectives, psychomotor domain*. Washington: Gryphon House. [[GS Search](#)]
- Sindre, G. & Moody, D. (2003). Evaluating the Effectiveness of Learning Interventions: an Information Systems Case Study. *Proc. of the 11th European Conf. on Information Systems*, Paper 80. Naples, Italy. [[GS Search](#)]
- Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure? *Academy of Management Learning & Education*, 9(2), 169-191. [[GS Search](#)]
- Sweetser, P. & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment*, 3(3), 1-24. [[GS Search](#)]
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston: Allyn and Bacon. [[GS Search](#)]
- 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). London: Springer. [[GS Search](#)]
- Thomas, G., Martin, D., & Pleasants, K. (2011). Using self- and peer-assessment to enhance students' future-learning in higher education. *Journal of University Teaching & Learning Practice*, 8(1), 1-17. [[GS Search](#)]
- Tonussi, L. P. & Hauck, J. C. R. (2017). Um Módulo de Jogo de Perguntas e Respostas para apoio ao Ensino de Gerência de Projetos Integrado ao Moodle. *Proc. of the Computer on the Beach*. Florianópolis, Brazil (in Portuguese). [[GS Search](#)]
- Trochim, W. M., & Donnelly, J. P. (2008). *Research methods knowledge base* (3rd ed.). Mason: Atomic Dog Publishing. [[GS Search](#)]
- Tullis, T. & Albert, W. (2008). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Burlington: Morgan Kaufmann. [[GS Search](#)]
- Wiebe, E. N., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior*, 32, 123-132. doi: [10.1016/j.chb.2013.12.001](https://doi.org/10.1016/j.chb.2013.12.001) [[GS Search](#)]
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in Software Engineering*. New York: Springer-Verlag Berlin Heidelberg. [[GS Search](#)]
- Yin, R. K. (2017). *Case study research: design and methods* (6th ed.). Thousand Oaks: Sage Publications, Inc. [[GS Search](#)]
- Zaibon, S. B. & Shiratuddin, N. (2010). Heuristics Evaluation Strategy for Mobile Game-Based Learning. *Proc of the 6th IEEE Int. Conf. on Wireless, Mobile and Ubiquitous Technologies in Education*, (pp. 127-131). Kaohsiung, Taiwan. doi: [10.1109/WMUTE.2010.27](https://doi.org/10.1109/WMUTE.2010.27) [[GS Search](#)]
- Zaibon, S. B. (2015). User Testing on Game Usability, Mobility, Playability, and Learning Content of Mobile Game-Based Learning. *Journal Teknologi*, 77(29), 131-139. [[GS Search](#)]



Apêndice A. Informações dos estudos de caso conduzidos

Jogo	Área de conhecimento (ACM/IEEE, 2013)	Plataforma	Contexto	Disciplina/Semestre	Instituição	Tamanho da amostra
Casos de uso	Engenharia de Software	Não-digital	Pós-graduação em Ciência da Computação	Estratégias de Ensino na Engenharia de Software/2016-2	Universidade Federal de Santa Catarina	6
Castelo Lógico	Fundamentos de Desenvolvimento de Software	Não-digital	Pós-graduação em Ciência da Computação	Estratégias de Ensino na Engenharia de Software/2016-2	Universidade Federal de Santa Catarina	6
Comunica	Engenharia de Software	Não-digital	Pós-graduação em Ciência da Computação	Estratégias de Ensino na Engenharia de Software/2016-2	Universidade Federal de Santa Catarina	6
EAReqGame (Petri & Chiavegatti, 2015)	Engenharia de Software	Digital	Graduação em Computação	Engenharia de Software/2016-2	Instituto Federal Catarinense	6
				Engenharia de Software/2017-1	Instituto Federal Catarinense	41
SORTIA - Heapsort (Battistella et al., 2017)	Algoritmos e Complexidade	Não-digital	Graduação em Computação	Estrutura de Dados/2016-2	Universidade Federal de Santa Catarina	17
				Estrutura de Dados/2016-2		7
				Estrutura de Dados/2017-1		23
				Estrutura de Dados/2017-1		6
				Estrutura de Dados/2017-2		17
				Estrutura de Dados/2017-2		11
Jogo do Processo Unificado	Engenharia de Software	Não-digital	Pós-Graduação em Ciência da Computação	Estratégias de Ensino na Engenharia de Software/2016-2	Universidade Federal de Santa Catarina	6
Kahoot! PMQuiz (Petri et al., 2016)	Engenharia de Software	Digital	Graduação em Computação	Planejamento e Gestão de Projetos/2016-2	Universidade Federal de Santa Catarina	29
				Gestão de Projetos/2016-2		20
				Gestão de Projetos/2016-2	Universidade Federal de Santa Maria	11
				Planejamento e Gestão de Projetos/2017-1		17
				Gestão de Projetos/2017-1	Universidade Federal de Santa Catarina	24
				Planejamento e Gestão de Projetos/2017-2		18
Memorizando Heurísticas	Interação Humano-Computador	Não-digital	Pós-Graduação em Ciência da Computação	Estratégias de Ensino na Engenharia de Software/2016-2	Universidade Federal de Santa Catarina	6
O Jogo das Classes	Engenharia de Software	Não-digital	Pós-Graduação em Ciência da Computação	Estratégias de Ensino na Engenharia de Software/2016-2	Universidade Federal de Santa Catarina	6
PMDome (PMDome, 2017)	Engenharia de Software	Não-digital	Pós-Graduação em Ciência da Computação	Gestão de Tecnologia da Informação/2016-2	Instituto Federal Farroupilha	20
			Graduação em Computação	Workshop Gestão de Projetos	Instituto Federal de Santa Catarina	27
PM Master	Engenharia de Software	Não-digital	Graduação em Computação	Planejamento e Gestão de Projetos/2016-2	Universidade Federal de Santa Catarina	24
				Gestão de Projetos/2016-2		21
				Planejamento e Gestão de Projetos/2017-1		17
				Gestão de Projetos/2017-1		18
				Planejamento e Gestão de Projetos/2017-2		16
Detective Game	Engenharia de Software	Não-digital	Graduação em Computação	Planejamento e Gestão de Projetos/2016-2	Universidade Federal de Santa Catarina	26
				Planejamento e Gestão de Projetos/2017-1		17
				Planejamento e Gestão de		17



				Projetos/2017-2		
SORTIA Quicksort (Battistella et al., 2017)	Algoritmos e Complexidade	Não-digital	Graduação em Computação	Estrutura de Dados/2016-2	Universidade Federal de Santa Catarina	21
				Estrutura de Dados/2016-2		6
				Estrutura de Dados/2017-1		25
				Estrutura de Dados/2017-1		6
				Estrutura de Dados/2017-2		18
				Estrutura de Dados/2017-2		11
QuizGame Moodle (Tonussi & Hauck, 2017)	Engenharia de Software	Digital	Graduação em Computação	Gestão de Projetos/2017-1	Universidade Federal de Santa Catarina	21
Risk Management Game	Engenharia de Software	Não-digital	Graduação em Computação	Planejamento e Gestão de Projetos/2016-2	Universidade Federal de Santa Catarina	31
				Gestão de Projetos/2016-2		23
				Planejamento e Gestão de Projetos/2017-1		21
				Gestão de Projetos/2017-1		21
				Engenharia de Software/2017-01	Instituto Federal de Santa Catarina	36
				Engenharia de Software/2017-01	Instituto Federal de Santa Catarina	31
Salve o Rei	Fundamentos de Desenvolvimento de Software	Não-digital	Pós-Graduação em Ciência da Computação	Estratégias de Ensino na Engenharia de Software/2016-2	Universidade Federal de Santa Catarina	6
SCRUM'ed (Schneider, 2015)	Engenharia de Software	Digital	Graduação em Computação	Engenharia de Software/2016-2	Universidade Luterana do Brasil	18
SCRUMIA	Engenharia de Software	Não-digital	Graduação em Computação	Planejamento e Gestão de Projetos/2016-2	Universidade Federal de Santa Catarina	26
				Planejamento e Gestão de Projetos/2017-1		19
				Planejamento e Gestão de Projetos/2017-2		17
Total						843