

AVALIAÇÃO DE SOFTWARES EDUCACIONAIS COM ÊNFASE EM JOGOS: UM PANORAMA DA LITERATURA

Giani Petri 

Professor no Colégio Politécnico da Universidade Federal de Santa Maria (UFMS) – giani.petri@ufsm.br

Resumo: *Softwares* educacionais estão sendo utilizados em diferentes níveis de ensino com o objetivo de auxiliar professores e estudantes no processo de ensino e aprendizagem. Assim, escolher um *software* educacional para levar para a sala de aula não é uma tarefa fácil, devido a diversidade de sistemas existentes e a insuficiência de evidências científicas sobre seus benefícios. Desse modo, é importante avaliá-los para obter evidências sobre sua qualidade. No entanto, algumas abordagens genéricas para a sua avaliação não consideram aspectos específicos de determinados tipos de *softwares* como jogos, aplicativos de celular *etc.* Assim, há a necessidade de utilizar abordagens desenvolvidas com o objetivo de avaliar os critérios, aspectos e especificidades de cada tipo de sistema. Nesse contexto, o objetivo deste artigo é apresentar um panorama sobre a literatura de avaliação de *softwares* educacionais, com ênfase em jogos, por meio de um mapeamento sistemático da literatura. Os resultados indicam que existem poucas abordagens que fornecem um suporte sistemático para avaliações de jogos. A maioria delas são *frameworks* e não métodos abrangentes, indicando uma falta de apoio em como conduzir as avaliações. Uma abordagem de avaliação se destaca dentre as encontradas, por ter sido desenvolvida e avaliada de forma sistemática, adotando um rigor científico nas avaliações de jogos e apresentando evidências de sua validade e confiabilidade.

Palavras-chave: Software educacional; Avaliação; Jogo educacional; Revisão da literatura.

EVALUATION OF EDUCATIONAL SOFTWARE WITH EMPHASIS ON GAMES: A LITERATURE OVERVIEW

Abstract: Educational software is being used at different education levels to assist teachers and students in the teaching and learning process. Thus, choose an educational software to take to the classroom is not an easy task, due to the diversity of existing tools and the lack of scientific evidence about its benefits. Thus, it is important to evaluate such software to obtain evidence on its quality. However, some generic evaluation approaches do not evaluate specific aspects of certain types of software such as games, mobile applications, etc. Thus, there is a need to use approaches developed to evaluate the criteria, aspects, and specificities of each type of system. In this context, the objective of this article is to present a literature overview on the evaluation of educational software, with an emphasis on games, through a systematic mapping study. The results indicate that there are few approaches, which provide systematic support, for game evaluations. Most of them are frameworks and not comprehensive methods, indicating a lack of support on how to conduct evaluations. An evaluation approach stands out among the approaches found in the literature, as it has been developed and evaluated systematically, adopting scientific rigor in game evaluations, and presenting evidence of its validity and reliability.

Keywords: Educational software; Evaluation; Educational game; Literature review.

Introdução

As Tecnologias da Informação e Comunicação (TIC) já são uma realidade em ambientes educacionais nos diferentes níveis de ensino. Com o objetivo de contribuir no processo de ensino e aprendizagem e no desenvolvimento de competências de estudantes, diferentes TIC são exploradas, desde a aplicação de dispositivos computacionais como computadores, *tablets*, lousas digitais, *etc.*, até o uso de *softwares* como jogos educacionais, aplicativos de celular, dentre outros.

Softwares educacionais (SE) objetivam dar suporte ao processo de aprendizagem, oferecendo elementos para que o usuário desenvolva competências durante e a partir de sua utilização ([LYRAS et al., 2014](#)). Um SE é considerado um tipo de *software* especializado ([ACM; IEEE-CS, 2013](#)), pois além de ser considerado tradicional, ele deve possuir objetivos educacionais e/ou pedagógicos bem definidos. Assim, um diferencia-se de outro *software* de aplicativo pelo fato de que durante o seu desenvolvimento, a maneira como os alunos aprendem deve ser levada em consideração ([ARDITO et al., 2006](#); [LYRAS et al., 2014](#)).

A diversidade de SE existentes e, muitas vezes, a insuficiência de estudos científicos sobre sua qualidade, dificultam o processo de escolha para serem inseridos em ambientes educacionais ([LYRAS et al., 2014](#); [PEREIRA et al., 2016](#)). Diante disso, é importante avaliar esse tipo de *software* de modo a obter evidências sobre sua qualidade e, também, identificar os benefícios que seu uso traz para os estudantes.

Na literatura, alguns estudos apresentam abordagens para a avaliação de SE tais como: Quali-EDU ([LIMA et al., 2015](#)), Petese ([COOMANS et al., 2015](#)) e LORI ([NESBIT et al., 2003](#)). Em geral, essas abordagens adotam *checklists* com um conjunto de critérios que incluem desde aspectos pedagógicos até a aspectos sobre a qualidade, como usabilidade, facilidade de uso, entre outros. Os critérios são geralmente associados a escalas numéricas de pontos com variação e/ou pesos para então definir se a ferramenta está em conformidade.

Embora o uso dessas abordagens para uma avaliação inicial de um SE seja interessante, elas apresentam algumas limitações. Em geral, elas são validadas somente de forma qualitativa, com estudos pilotos, usando um número limitado de ferramentas e com poucos estudantes ([BRITO; ALMEIDA; CAVALCANTI, 2003](#)). Esse cenário deixa a validade e

confiabilidade, do ponto de vista estatístico, questionáveis. Além disso, por serem abordagens genéricas, elas podem ser insuficientes para a avaliação de critérios específicos de alguns tipos de SE, como os jogos educacionais, aplicativos de celular, mundos virtuais e aplicativos com realidade aumentada ([PEREIRA et al., 2016](#)). Diante disso, há a necessidade de utilizar abordagens desenvolvidas com o objetivo de avaliar um tipo específico de SE, que englobem seus critérios, aspectos e especificidades ([BRITO; ALMEIDA; CAVALCANTI, 2003](#); [PEREIRA et al., 2016](#)).

Nesse contexto, considerando que o desenvolvimento de jogos educacionais digitais vem crescendo nos últimos anos e representa cerca de 43,8% do mercado brasileiro ([FLEURY et al., 2014](#)), o objetivo deste artigo é apresentar um panorama sobre a literatura de avaliação de *softwares* educacionais, com ênfase em jogos. Para isso, é realizado um mapeamento sistemático da literatura de modo a identificar e analisar o rigor científico das abordagens encontradas. Os resultados desse artigo contribuem para pesquisadores da área de tecnologias educacionais e instrutores a obterem uma visão geral sobre área de pesquisa e maiores detalhes sobre abordagens sistemáticas que podem ser usadas na avaliação de *softwares* educacionais, com ênfase em jogos.

Revisão da Literatura

Com o objetivo de identificar abordagens de avaliação de *softwares* educacionais com ênfase em jogos, foi realizado um mapeamento sistemático da literatura seguindo o processo proposto por [Petersen et al. \(2015\)](#).

O processo do mapeamento é dividido em três fases principais: definição, execução e análise. Na fase de definição, os objetivos da pesquisa são identificados e um protocolo é definido. O protocolo especifica as questões centrais de pesquisa e os procedimentos usados para realizar a revisão, incluindo a definição de critérios de inclusão e exclusão, fontes de dados e a *string* de pesquisa. A fase de execução consiste na busca e identificação de estudos relevantes, e sua seleção de acordo com os critérios estabelecidos no protocolo. Uma vez identificado os estudos, os dados relacionados às questões de pesquisa são extraídos dos estudos selecionados, analisados e sintetizados durante a fase de análise ([PETERSEN et al., 2015](#)). As seções a seguir apresentam em detalhes as três fases do estudo de mapeamento.

Definição

O mapeamento sistemático da literatura objetiva identificar as abordagens utilizadas para avaliar jogos educacionais, analisando como elas são definidas, operacionalizadas, desenvolvidas e avaliadas, sendo uma atualização do processo de revisão da literatura apresentada em [Petri e Gresse von Wangenheim \(2016\)](#). Para atingir esse objetivo foram definidas as seguintes questões de pesquisa:

QP1: Quais métodos, modelos, escalas ou *frameworks* (abordagens) existem para avaliar jogos educacionais?

QP2: Quais critérios ou fatores de qualidade são avaliados?

QP3: Como a coleta e análise de dados é operacionalizada?

QP4: Como as abordagens foram desenvolvidas e avaliadas?

Critérios de Inclusão e Exclusão. De acordo com o objetivo da pesquisa e as questões definidas, critérios para a seleção de estudos relevantes foram definidos. A seleção inclui somente artigos que apresentam uma abordagem bem definida para a avaliação de jogos educacionais. Foram selecionados somente artigos de estudos empíricos/avaliações, escritos no idioma inglês (ou em português com um *abstract* em inglês), disponíveis em bibliotecas digitais, publicados entre janeiro de 1995 e dezembro de 2019).

Por outro lado, foram excluídos:

- Qualquer estudo não relacionado a uma abordagem bem definida para a avaliação de jogos educacionais;
- Qualquer artigo não relacionado a estudo empírico/avaliação;
- Artigos que apresentam a avaliação de um jogo educacional, mas não utilizam uma abordagem definida para a avaliação.

Fontes de dados e *string* de busca. As fontes de dados foram escolhidas com base na sua relevância, incluindo *ACM Digital Library* (<http://dl.acm.org/>), *IEEEExplore* (<http://ieeexplore.ieee.org/>), *SpringerLink* (<http://www.springerlink.com/>), *ScienceDirect* (<http://www.sciencedirect.com/>) e *Wiley Online Library* (<http://onlinelibrary.wiley.com/>). Além disso, a busca também foi realizada no *Google Scholar* (<http://scholar.google.com/>), que

indexa um grande conjunto de dados em várias fontes de pesquisa ([HADDAWAY et al., 2015](#)), mas que também possuam contribuições relevantes.

De acordo com o objetivo de pesquisa, foi definida a *string* de busca com base na identificação de conceitos principais, tais como modelo, jogo educacional e avaliação, incluindo também seus sinônimos, como apresentado no Quadro 1.

Quadro 1 – Termos de busca

Conceito	Sinônimos
<i>Model</i>	<i>method, framework, scale</i>
<i>educational game</i>	<i>serious game, game-based learning</i>
<i>evaluation</i>	<i>assessment</i>

Fonte: Do autor (2020)

Usando esses termos de busca e operadores lógicos, a consulta foi calibrada e adaptada em conformidade com a sintaxe específica de cada fonte de dados, como apresentado no Quadro 2.

Quadro 2 – Strings de busca

Fonte de Dados	String
ACM Digital Library	<i>(model OR method OR framework OR scale) AND ("educational game" OR "serious game" OR "game-based learning") AND (evaluation OR assessment) for: ((model OR method OR framework OR scale) AND ("educational game" OR "serious game" OR "game-based learning")) AND (evaluation OR assessment)) Published since January 1995</i>
IEEEExplore	<i>((model OR method OR framework OR scale) AND ("educational game" OR "serious game" OR "game-based learning") AND (evaluation OR assessment) IN metadata) AND (pyr >= 1995 AND pyr <= 2019)</i>
Springer Link	<i>'(model OR method OR framework OR scale) AND ("educational game" OR "serious game" OR "game-based learning") AND (evaluation OR assessment)' published between 1995 - 2019 within Article</i>
ScienceDirect	<i>pub-date > 1994 and ((model OR method OR framework OR scale) AND ("educational game" OR "serious game" OR "game-based learning") AND (evaluation OR assessment))</i>

Wiley Online Library	(model OR method OR framework OR scale) AND ("educational game" OR "serious game" OR "game-based learning") AND (evaluation OR assessment) in All Fields between years 1995 and 2019 Publication type: Journals
Google Scholar	(model OR method OR framework OR scale) AND ("educational game" OR "serious game" OR "game-based learning") AND (evaluation OR assessment) Custom range: 1995-2019

Fonte: Do autor (2020)

Execução

O mapeamento foi realizado em janeiro e fevereiro de 2020 pelo autor deste trabalho. O Quadro 3 apresenta os resultados retornados de cada fonte de dados. Do *Google Scholar*, *ACM Digital Library* e *ScienceDirect* foram selecionados somente os 1.000 resultados mais relevantes, observando uma baixa relevância após essas quantidades. Das fontes de dados *IEEEExplore*, *SpringerLink* e *Wiley Online Library* todos os artigos retornados foram analisados. Como resultado, um total de 5.131 artigos foram analisados na primeira etapa.

Quadro 3 – Resultados da busca por fonte de dados

	<i>Google Scholar</i>	<i>ACM</i>	<i>IEEE Xplore</i>	<i>Springer Link</i>	<i>Science Direct</i>	<i>Wiley</i>	Total
Analisados na primeira etapa	1.000	1.000	325	911	1.000	895	5131
Selecionados após primeira etapa	78	13	25	10	19	9	154
Selecionados após segunda etapa	7	0	5	0	3	1	16

Fonte: Do autor (2020)

Durante a primeira etapa, os resultados da pesquisa foram rapidamente analisados com base em seu título e no trecho do artigo apresentado no resultado da busca. O resumo foi lido apenas no caso de o título não fornecer evidência de nenhum critério de exclusão. Artigos irrelevantes e duplicados foram removidos. Nessa etapa, foram selecionados 154 artigos potencialmente relevantes. Então, foi realizada uma segunda etapa de seleção. Na segunda etapa, analisou-se o resumo completo dos artigos e, de forma rápida, o escopo dos mesmos para obter informações sobre a abordagem de avaliação. Como resultado, 16 artigos

(descrevendo um total de 11 abordagens) foram identificados como estudos primários (Quadro 4).

Quadro 4 – Trabalhos selecionados

Ano	Título	Referência
2006	How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?	(FREITAS; OLIVER, 2006)
2008	Development a general framework for evaluating games-based learning	(CONNOLLY; STANSFIELD; HAINEY, 2008)
2009	Towards the development of a games-based learning evaluation framework	(CONNOLLY; STANSFIELD; HAINEY, 2009)
2009	EGameFlow: a scale to measure learners' enjoyment of e-learning games	(FU; SU; YU, 2009)
2010	A refined evaluation framework for games-based learning	(HAINEY; CONNOLLY; BOYLE, 2010)
2011	A model for the evaluation of educational games for teaching software engineering	(SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011)
2012	Is game-based learning suitable for engineering education?	(CARVALHO, 2012)
2012	A game scale to evaluate educational computer games	(AK, 2012)
2012	Towards a comprehensive methodology for the research and evaluation of serious games	(MAYER, 2012)
2013	A brief methodology for researching and evaluating serious games and game-based learning	(MAYER et al., 2013)
2014	The research and evaluation of serious games: Toward a comprehensive methodology	(MAYER et al., 2014)
2015	Towards a construction and validation of a serious game product quality model	(GARCIA-MUNDO; GENERO; PIATTINI, 2015)
2017	An efficient framework for game-based learning activity	(CHEW, 2017)
2018	Serious games: Quality characteristics evaluation framework and case study	(ABDELLATIF; MCCOLLUM; MCMULLAN, 2018)
2018	MEEGA+, Systematic Model to Evaluate Educational Games	(PETRI et al., 2018)
2019	A Method for the Evaluation of the Quality of Games for Computing Education	(PETRI; GRESSE VON WANGENHEIM 2019)

Fonte: Do autor (2020)

Análise dos Dados

No total, foram identificados 16 artigos descrevendo 11 abordagens para avaliar jogos educacionais. Embora tenham sido considerados mais de 20 anos (1995-2019) na revisão,

apenas se encontrou publicações relevantes após 2006, com um aumento considerável de trabalhos publicados a partir de 2012. Isso mostra que o interesse em abordagens para avaliar *softwares* educacionais, com ênfase em jogos, vem crescendo nos últimos anos.

De modo a apresentar os resultados da revisão da literatura, cada questão de pesquisa é respondida separadamente.

QP1: Quais métodos, modelos, escalas ou frameworks (abordagens) existem para avaliar jogos educacionais?

Ao analisar os estudos selecionados, identificamos 11 abordagens diferentes para avaliar jogos educacionais. Cinco delas apresentam um *framework* ([FREITAS; OLIVER, 2006](#); [CONNOLLY; STANSFIELD; HAINEY, 2009](#); [CARVALHO, 2012](#); [CHEW, 2017](#); [ABDELLATIE; MCCOLLUM; MCMULLAN, 2018](#)); duas apresentam uma escala ([FU; SU; YU, 2009](#); [AK, 2012](#)), uma apresenta uma metodologia genérica ([MAYER, 2012](#)), duas mostram um modelo ([SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [GARCIA-MUNDO; GENERO; PIATTINI, 2015](#)), e outra expõe um método de avaliação ([PETRI; GRESSE VON WANGENHEIM, 2019](#)). Apresentamos uma breve descrição de cada abordagem selecionada.

O estudo de Connolly *et al.* descreve o *Evaluation Framework for Effective Games-based Learning* (GBL) ([CONNOLLY; STANSFIELD; HAINEY, 2008](#); [CONNOLLY; STANSFIELD; HAINEY, 2009](#); [HAINEY; CONNOLLY; BOYLE, 2010](#)), um *framework* para GBL que objetiva identificar o que pode ser potencialmente avaliado em um aplicativo GBL. A abordagem propõe a avaliação da GBL em relação ao desempenho, motivação, percepções e preferências do aluno, bem como do ambiente do próprio GBL e de como se dá a colaboração entre os participantes. O *framework* pode ser personalizado para requisitos específicos, dependendo da medição específica necessária.

Outra abordagem é o *framework* de Freitas e Oliver ([FREITAS; OLIVER, 2006](#)), que objetiva ajudar os tutores a avaliar o potencial do uso de jogos e aprendizado baseado em simulação em sua prática. Permite que os profissionais sejam mais críticos sobre como incorporam jogos e simulações em seus planos de aula, que pesquisadores e avaliadores desenvolvam métricas para apoiar a análise eficaz de jogos e simulações educacionais existentes e, que os projetistas educacionais considerem um conjunto mais especializado, e baseado no usuário, de fatores específicos. As quatro dimensões avaliadas pela abordagem

são: contexto, aluno (ou grupo de alunos), mundo da representação interna e processo de aprendizagem.

[Carvalho \(2012\)](#) indica um *framework* que avalia a eficiência do GBL com foco no ensino de engenharia. Cobrindo os dois primeiros níveis do modelo de avaliação de *Kirkpatrick* (reação e aprendizado), ele é dividido em três estágios: teste alfa, teste beta e teste gama, cada um com objetivos claros, protocolos e dados predefinidos e ferramentas de coleta. Objetiva avaliar a eficiência dos jogos em termos de jogabilidade, história, mecanismos, usabilidade, conhecimento, motivação e satisfação.

[Chew \(2017\)](#) apresenta um *framework* que define fatores a serem considerados no *design* e análise de atividades de aprendizagem baseadas em jogos. Os fatores considerados na abordagem são o envolvimento cognitivo, o comportamento, o envolvimento emocional, a imersão e o desafio. Na fase de *design* da atividade de aprendizado baseada em jogos, os autores adaptaram os fatores e aplicaram o processo de *Design Thinking* para permitir considerações mais profundas sobre os fatores de jogos identificados. Um questionário é desenvolvido com base nos fatores para analisar a eficácia das atividades de aprendizagem.

Outro *framework* foi proposto por [Abdellatif, Mccollum e Mcmullan \(2018\)](#), com o objetivo de avaliar várias dimensões de jogos sérios, combinando características de qualidade. Essa proposta foi projetada para medir características que não exigem a realização de um experimento e pode ser aplicado em um curto período. A abordagem inclui características que, quando ausentes, impedem que jogos sérios entreguem seu conteúdo educacional para um público designado de forma eficaz, incluindo usabilidade, compreensão, motivação, engajamento e experiência do usuário.

O artigo de [Fu, Su e Yu \(2009\)](#) discute a *EGameFlow*, uma escala que avalia o prazer do usuário de jogos de *e-learning* para ajudar os desenvolvedores a entender pontos fortes e fracos da percepção dos alunos de acordo com o nível 1 de avaliação (reação). Ele avalia a qualidade do jogo em relação a oito fatores: imersão, interação social, desafio, clareza de objetivos, *feedback*, concentração, controle e melhoria do conhecimento.

Outra escala foi proposta por [Ak \(2012\)](#), a qual visa a seleção de bons jogos educacionais de computador. Tem como objetivo medir a qualidade dos jogos antes de aplicá-lo nas aulas. A qualidade do jogo é medida em termos de diversão e aprendizado.

Uma metodologia abrangente para a pesquisa e avaliação de jogos sérios foi proposta por [Mayer \(2012\)](#). Ela contém uma estrutura, modelos conceituais, projetos de pesquisa, construções e escalas de avaliação e técnicas de coleta de dados. A metodologia avalia jogos sérios em três momentos diferentes (pré-jogo, no jogo e pós-jogo) em termos de experiências/habilidades anteriores, desempenho, jogabilidade, experiência, satisfação do jogador e aprendizado.

Além desses estudos, o modelo proposto por [Savi, Gresse von Wangenheim e Borgatto \(2011\)](#), nomeado como MEEGA (Modelo para Avaliação de Jogos Educacionais), foi desenvolvido especificamente para a avaliação de jogos educacionais para o ensino de engenharia de *software*. Ele concentra-se no nível de avaliação 1 (reação), capturando a reação dos alunos após o jogo, aplicando um questionário padronizado. O MEEGA mede três fatores de qualidade dos jogos educacionais: motivação, experiência do usuário e aprendizado.

O modelo *QSGame-Model*, proposto por [Garcia-Mundo, Genero e Piattini \(2015\)](#), é um modelo de qualidade de produto específico para jogos sérios, sendo uma customização da norma ISO/IEC 25010 ([ISO, 2011](#)), alterando e adaptando suas definições e sub-características para o contexto de jogos sérios. A personalização é principalmente em termos das sub-características de usabilidade e adequação funcional.

Em [Petri e Gresse von Wangenheim \(2019\)](#) é descrito o Método MEEGA+ para avaliação de jogos educacionais, como uma evolução do modelo MEEGA de [Savi et al. \(2011\)](#). O método MEEGA+ é composto por um modelo de avaliação, que sistematicamente avalia jogos em termos de usabilidade e experiência do jogador por meio de um instrumento de medição padronizado e fornece uma escala, que classifica o jogo avaliado em níveis de qualidade. Além disso, o método MEEGA+ também define um processo, que descreve em detalhes as fases, atividades e produtos de trabalho para orientar instrutores e pesquisadores na condução de avaliações de jogos para o ensino de computação.

QP2: Quais critérios ou fatores de qualidade são avaliados?

Para responder a essa pergunta, foram analisados os fatores e/ou critérios de qualidade avaliados pelas abordagens identificadas. Em resumo, identificamos 64 fatores

diferentes que foram utilizados para avaliar jogos educacionais. Todas as abordagens usam mais de um fator de qualidade para avaliar os jogos.

Os fatores mais frequentemente usados pelas abordagens selecionados são: aprendizagem (7), usabilidade (5), interação social (5), desafio (5) e imersão (4). Normalmente, a avaliação da aprendizagem refere-se à melhoria de competências. [Connolly et al. \(2009\)](#) define aprendizado como uma melhoria no desempenho do aluno como resultado da intervenção. Outras abordagens avaliam a melhoria da aprendizagem/conhecimento com base nas percepções dos alunos ([FU; SU; YU, 2009](#); [SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)).

A usabilidade é definida em termos de consciência do progresso, consistência da interface (cores, fontes), controles e *feedback* visual ([CARVALHO, 2012](#)). Dois estudos definem usabilidade com base na norma ISO/IEC 25010 ([GARCIA-MUNDO; GENERO; PIATTINI, 2015](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)).

Interação social refere-se à criação de um sentimento de ambiente compartilhado e estar conectado com outras pessoas em atividades de cooperação ou competição ([FU; SU; YU, 2009](#); [SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)). Desafio significa que um jogo precisa ser suficientemente desafiador em relação ao nível de competência do jogador. O aumento da dificuldade deve ocorrer em um ritmo apropriado que acompanha a curva de aprendizado. Novos obstáculos e situações devem ser apresentados ao longo do jogo para minimizar a fadiga e manter os alunos interessados ([SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [CHEW, 2017](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)). A imersão permite que o jogador tenha uma experiência de profundo envolvimento no jogo, criando um desafio com foco no mundo real, para que ele se esqueça do mundo exterior durante o jogo ([FU; SU; YU, 2009](#); [SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [CHEW, 2017](#)).

Apenas um artigo selecionado apresentou resultados de uma análise exploratória dos constructos teóricos (fatores de qualidade) que seu instrumento medição (questionário) estava medindo ([PETRI; GRESSE VON WANGENHEIM, 2019](#)). Os autores destacam que a qualidade de jogos educacionais é medida em termos de experiência do jogador e usabilidade.

QP3: Como a coleta e análise de dados é operacionalizada?

Para responder a essa pergunta, analisamos como as abordagens operacionalizam a avaliação, incluindo estratégias de pesquisa, instrumentos de coleta de dados e métodos de análise de dados.

Analisando as estratégias de pesquisa, classificamos os artigos de acordo com os tipos de estudo comuns: experimental, quase-experimental e não experimental. Quatro deles ([FU; SU; YU, 2009](#); [CARVALHO, 2012](#); [CHEW, 2017](#); [ABDELLATIF; MCCOLLUM; MCMULLAN, 2018](#)) fornecem uma abordagem de avaliação a ser conduzida de maneira *ad-hoc*, não indicando claramente a estratégia da pesquisa adotada. Outros dois ([SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)) propõem a realização da avaliação na forma de um estudo de caso (não experimental). Eles definem explicitamente o objetivo da avaliação e fornecem um questionário padronizado com base no modelo definido a ser aplicado após o tratamento (jogo educacional) para coletar dados sobre a percepção dos alunos.

O método MEEGA+ ainda descreve um processo, que apresenta detalhes de todas as atividades que precisam ser conduzidas para avaliar um jogo educacional, sendo bastante útil para avaliadores com pouco experiência na condução desse tipo de estudo ([PETRI; GRESSE VON WANGENHEIM, 2019](#)). A abordagem proposta por ([MAYER, 2012](#)) define uma estratégia quase-experimental, semelhante ao desenho experimental, mas sem uma alocação aleatória de alunos para o grupo experimental ou controle.

Nenhuma informação sobre a operacionalização da avaliação foi fornecida por ([AK, 2012](#); [CONNOLLY; STANSFIELD; HAINEY, 2009](#); [FREITAS; OLIVER, 2006](#); [GARCIA-MUNDO; GENERO; PIATTINI, 2015](#)). Ao analisar o tipo de instrumentos de coleta de dados, identificamos que a maioria das abordagens coleta dados por meio de questionários (8), mas apenas três foram sistematicamente desenvolvidos e avaliados estatisticamente ([FU; SU; YU, 2009](#); [SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)). Carvalho (2012) também utilizou um questionário como instrumento de coleta de dados, mas não forneceu informações sobre a sua validade. Ao examinar o

formato de resposta dessas escalas, identificamos que a escala *Likert* é a mais utilizada (5 estudos), geralmente representando o menor e o maior grau em que os entrevistados concordam com os itens. Além disso, uma escala ordinal também é usada (3 estudos) para medir características específicas. Outros métodos de coleta de dados aplicados incluem entrevistas semiestruturadas ([CARVALHO, 2012](#)) e testes para avaliar o conhecimento dos alunos ([CARVALHO, 2012](#)).

Ao investigar os métodos de análise de dados dos estudos selecionados, apenas três abordagens ([FU; SU; YU, 2009](#); [SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)) fornecem informações sobre quais métodos são utilizados. [Savi, Gresse von Wangenheim e Borgatto \(2011\)](#) usam métodos estatísticos descritivos (mediana/moda) e técnicas de visualização gráfica, como histograma e diagramas de frequência. [Fu, Su e Yu \(2009\)](#) também usam métodos estatísticos descritivos, como média, desvio padrão e coeficiente de correlação de *Pearson* para examinar a dependência entre variáveis. Além disso, os autores também incluem testes de hipóteses para rejeitar (ou aceitar) uma hipótese com relação a um fator de qualidade do jogo. O teste *t* é usado para comparar duas médias amostrais, em um projeto de um fator e dois tratamentos, e a ANOVA é usada para avaliar a discrepância no nível de prazer psicológico entre os sujeitos ([FU; SU; YU, 2009](#)). Em [Petri e Gresse von Wangenheim \(2019\)](#) é oferecida uma planilha para auxiliar na análise dos dados coletados pelo questionário, gerando de forma automatizada gráficos de frequência, e dados como mediana e média.

QP4: Como estas abordagens foram desenvolvidas e avaliadas?

Ao avaliar os estudos selecionados, identificamos que a maioria deles (7) não reporta uma metodologia para desenvolver a abordagem de avaliação. Em geral, elas parecem ser desenvolvidas de maneira *ad-hoc* ([CARVALHO, 2012](#)) ou apenas com base em construtos teóricos ([AK, 2012](#); [MAYER, 2012](#); [CONNOLLY; STANSFIELD; HAINEY, 2008](#)), mas não fornecendo uma definição explícita do objetivo, medidas ou instrumentos de coleta de dados. Uma abordagem (*QSGame-Model*) ([GARCIA-MUNDO; GENERO; PIATTINI, 2015](#)) foi desenvolvida adotando uma metodologia *top-down* ([FRACH; CARVALLO, 2003](#)), usada para adaptar modelos de qualidade a um domínio específico, personalizando características gerais para um contexto específico.

Por outro lado, três abordagens relatam uma metodologia sistemática para seu desenvolvimento ([FU; SU; YU, 2009](#); [SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)). *EGameFlow*, MEEGA e sua evolução, o método MEEGA+, seguem o Guia de Desenvolvimento de Escalas ([DEVELLIS, 2016](#)) para desenvolver sistematicamente um instrumento de medição. Além disso, o MEEGA e o MEEGA+ foram desenvolvidos usando a abordagem GQM (*Goal/Question/Metric*) ([BASILI; CALDIERA; ROMBACH, 1994](#)) para definir explicitamente um programa de medição para avaliar os fatores de qualidade definidos. O método MEEGA+ ainda define um processo de avaliação com base em uma metodologia para modelagem de processos ([ACUÑA; FERRÉ, 2001](#)) e utiliza o modelo matemático de Teoria de Resposta ao Item ([PASQUALI; PRIMI, 2003](#)) para definir a escala de medição.

Investigamos também os fatores utilizados para avaliar as abordagens. Identificamos que a maioria delas (8) não define explicitamente critérios. Normalmente, elas são propostas e parcialmente avaliadas por meio de alguns estudos piloto, aplicando-a para avaliar um jogo em sala de aula ([CARVALHO, 2012](#); [MAYER, 2012](#); [CONNOLLY; STANSFIELD; HAINEY, 2009](#); [FREITAS; OLIVER, 2006](#)). Nenhuma informação sobre a avaliação foi encontrada para a abordagem proposta por [Ak \(2012\)](#) e [Garcia-Mundo, Genero e Piattini \(2015\)](#).

Por outro lado, três abordagens descrevem uma avaliação sistemática ([FU; SU; YU, 2009](#); [SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)). O MEEGA ([SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#)) foi avaliado em termos de aplicabilidade, utilidade, validade e confiabilidade por meio de três estudos de caso em dois cursos diferentes, usando três jogos ([GRESSE VON WANGENHEIM; BORGATTO, 2011](#)). Um total de 79 pontos de dados foi coletado e analisado em relação a correlação de itens, correlação item-total, variância, média e coeficiente alfa de *Cronbach*.

O *EGameFlow* ([FU; SU; YU, 2009](#)) foi avaliado em termos de análise de itens, confiabilidade e validade em quatro sessões de jogo no mesmo curso, usando diferentes jogos de *e-learning* ([FU; SU; YU, 2009](#)). Um total de 166 pontos de dados foi coletado e analisado usando os seguintes testes: média, desvio padrão, comparação extrema de grupos, teste de homogeneidade, teste *t*, ANOVA, correlação de *Pearson* e correlação alfa de *Cronbach*.

O Método MEEGA+ ([PETRI et al., 2018](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)) foi avaliado de forma quantitativa e qualitativa. A análise quantitativa envolveu a análise da validade e confiabilidade do modelo, com base em dados de 1048 estudantes, que jogaram 24 diferentes jogos educacionais. A análise quantitativa envolveu uma análise fatorial, análise de correlação de itens usando o coeficiente de correlação de *Spearman* e, coeficiente alfa de *Cronbach*. A análise qualitativa envolveu um painel com 19 especialistas em jogos educacionais que avaliaram o método quanto a sua autenticidade, validade, usabilidade, correteude, completude, consistência, compreensibilidade, não ambiguidade e flexibilidade.

Discussão dos Resultados

Ao verificar as abordagens selecionadas (QP1), identificamos que a maioria delas são *frameworks* (5) para avaliar jogos educacionais ([FREITAS; OLIVER, 2006](#); [CONNOLLY; STANSFIELD; HAINEY, 2009](#); [CARVALHO, 2012](#); [CHEW, 2017](#); [ABDELLATIF; MCCOLLUM; MCMULLAN, 2018](#)). Normalmente, os *frameworks* definem um conjunto de critérios que vão da perspectiva pedagógica à perspectiva do jogo, incluindo contexto, ambiente, especificações do aluno, preferências, jogabilidade, experiência do usuário *etc.* ([FREITAS; OLIVER, 2006](#); [CONNOLLY; STANSFIELD; HAINEY, 2009](#); [CARVALHO, 2012](#)). Esses critérios são usados para orientar e ajudar os instrutores a avaliar jogos educacionais em um contexto de aprendizagem e área de conhecimento específicos ([FREITAS; OLIVER, 2006](#)). Assim, eles são considerados uma abordagem flexível e fácil de usar, com a capacidade de ajudar os profissionais a refletir sobre os processos e abordagens de aprendizagem ([FREITAS; OLIVER, 2006](#)). No entanto, os *frameworks* em si não fornecem orientações sobre como conduzir a avaliação, coleta e análise de dados, principalmente, para pesquisadores sem experiência na condução de estudos de avaliação.

Nesse sentido, os trabalhos apresentados por ([FU; SU; YU, 2009](#); [AK, 2012](#)) propõem escalas que fornecem instrumentos para mensurar a qualidade dos jogos ([FU; SU; YU, 2009](#)). No entanto, apenas a escala *EGameFlow* ([FU; SU; YU, 2009](#)) foi avaliada analisando sua validade e confiabilidade como um instrumento para avaliar o nível de diversão proporcionado por jogos de *e-learning* para seus usuários ([FU; SU; YU, 2009](#)). Por outro lado,

nenhuma avaliação da escala proposta por [Ak \(2012\)](#) foi encontrada, deixando questionável sua validade e confiabilidade ([KITCHENHAM; PFLEEGER; FENTON, 1995](#); [KIMBERLIN; WINTERSTEIN, 2008](#)).

Com suporte mais abrangente, [Mayer \(2012\)](#) propõe uma metodologia genérica de avaliação para jogos sérios. Porém, embora inclua um *framework*, modelos conceituais, escalas de avaliação e técnicas de coleta de dados, nenhuma informação sobre a aplicabilidade e validade desse método foi encontrada. Por outro lado, uma abordagem que se destaca na revisão da literatura foi o método MEEGA+ ([PETRI; GRESSE VON WANGENHEIM, 2019](#)). Desenvolvido com base no modelo MEEGA ([SAVI, GRESSE VON WANGENHEIM; BORGATTO, 2011](#)), que é amplamente utilizado na prática, o método MEEGA+ se destacou frente às outras abordagens por ter sido desenvolvido e avaliado de forma sistemática, adotando uma metodologia rigorosa.

Ao investigar os fatores de qualidade utilizados para avaliar os jogos educacionais (QP2), observamos que existe uma grande diversidade. No entanto, a melhoria da aprendizagem é o fator mais avaliado. A aprendizagem é frequentemente avaliada comparando-se o nível de competência após o jogo com o prévio, normalmente com base em uma pontuação pré/pós-teste (MAYER, 2012) ou através de uma autoavaliação depois a partida ([FU; SU; YU, 2009](#); [SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)). Além da aprendizagem, a maioria das abordagens também considera vários outros fatores de qualidade, como desafio, competência, interação social, diversão, usabilidade *etc.*, confirmando também os resultados de [Calderón e Ruiz \(2015\)](#) no que diz respeito a jogos educacionais em diversas áreas do conhecimento. Esses fatores são avaliados por serem considerados importantes para promover um aprendizado mais profundo e ativo.

Em geral, também foi observado uma falta de suporte metodológico fornecido para operacionalizar a coleta e análise de dados (QP3). Apenas três abordagens (*EGameFlow*, MEEGA e o método MEEGA+) fornecem uma definição explícita dos instrumentos de coleta de dados e métodos de análise de dados.

Como resultado da revisão da literatura, foi identificado uma fraqueza significativa na maneira como os instrumentos (normalmente questionários) são desenvolvidos de forma *ad-*

hoc (QP4). No entanto, para obter resultados válidos, é imprescindível definir e operacionalizar sistematicamente as medidas e os instrumentos de coleta de dados ([WOHLIN et al., 2012](#)). Apenas três abordagens (*EGameFlow*, MEEGA, e MEEGA+) propõem questionários sistematicamente desenvolvidos e avaliados, sendo desenvolvidos adotando o guia de desenvolvimento de escala proposto por [DeVellis \(2016\)](#). E ainda, são os únicos estudos ([FU; SU; YU, 2009](#); [SAVI; GRESSE VON WANGENHEIM; BORGATTO, 2011](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)) que relatam explicitamente uma avaliação sistemática (QP4). Os critérios utilizados para a validação são definidos com base na teoria de desenvolvimento de escalas ([DEVELLIS, 2016](#)), incluindo aplicabilidade, utilidade, validade e confiabilidade. As outras abordagens selecionadas em nossa revisão parecem ter sido avaliadas por meio de estudos piloto, não validando os modelos/instrumentos de coleta de dados em si ([FREITAS; OLIVER, 2006](#); [CONNOLLY; STANSFIELD; HAINEY, 2009](#); [CARVALHO, 2012](#); [MAYER, 2012](#)).

Em resumo, analisando os resultados, foi observado um pequeno número de abordagens sistemáticas, válidas e confiáveis usadas para a avaliação de jogos que abrangem tanto a avaliação da aprendizagem quanto de aspectos importantes para fornecer uma experiência positiva e envolvente ao jogador. No entanto, uma abordagem se destaca das outras, o Método MEEGA+ ([PETRI; GRESSE VON WANGENHEIM, 2019](#)), por ter sido sistematicamente desenvolvido e avaliado de forma quantitativa e qualitativa, por diversos especialistas em jogos educacionais. O Método MEEGA+ ainda se diferencia por oferecer um suporte metodológico abrangente, incluindo um modelo de avaliação, um questionário padronizado, uma escala de medição e um processo que guia professores na condução de avaliação de jogos educacionais.

Ameaças à Validade

Como em qualquer revisão da literatura, existem algumas ameaças à validade. Portanto, foram identificadas ameaças potenciais e aplicadas estratégias de mitigação para minimizar seu impacto no estudo.

Identificação de estudos. Uma ameaça é a omissão de estudos relevantes. Para minimizar esse risco, foi definida cuidadosamente a *string* de busca para ser o mais inclusivo possível, considerando não apenas os conceitos principais, mas também sinônimos. O risco de excluir estudos relevantes foi ainda mais mitigado pelo uso de várias fontes de dados que cobrem uma grande quantidade das publicações científicas.

Seleção de estudos e extração de dados. As ameaças na seleção e extração de dados foram minimizadas fornecendo uma definição detalhada dos critérios de inclusão/exclusão. Foi definido e documentado um protocolo rígido para a seleção dos estudos. A extração de dados foi prejudicada em vários casos, pois muitos estudos não foram relatados em alinhamento com as estratégias de pesquisa comuns. Nesses casos, as informações foram inferidas com base nas informações contidas no artigo.

Considerações Finais

Os resultados da revisão da literatura indicam que existem poucas abordagens que fornecem um suporte sistemático e válido para avaliação de jogos. A maioria deles são *frameworks* e não métodos abrangentes, indicando falta de apoio sobre como conduzir as avaliações. Além disso, a maioria das abordagens também parece ser desenvolvida de maneira *ad-hoc*, não fornecendo uma definição explícita do objetivo, medidas ou instrumentos de coleta de dados confiáveis e válidos.

Uma abordagem de avaliação, o método MEEGA+ ([PETRI et al., 2018](#); [PETRI; GRESSE VON WANGENHEIM, 2019](#)), se destacou dentre as encontradas na literatura, por ter sido desenvolvido e avaliado de forma sistemática, adotando um rigor científico nas avaliações de jogos e apresentando evidências de sua validade e confiabilidade. O método MEEGA+ fornece um suporte abrangente para a avaliação da qualidade de jogos usados como estratégia instrucional sendo composto por um modelo que define fatores de qualidade para serem avaliados por meio de um instrumento de medição padronizado e validado; uma escala, que classifica o jogo em níveis de qualidade; e um processo de avaliação de jogos. O método MEEGA+ pode ser usado por desenvolvedores de jogos, instrutores e pesquisadores de modo a avaliar a qualidade de jogos como base para identificar melhorias e/ou adotá-los na prática.

Todo o material do método MEEGA+ está disponível gratuitamente no seguinte endereço:
<http://www.gqs.ufsc.br/quality-evaluation/meega-plus/>.

Referências

- ABDELLATIF, Abdelbaset J.; MCCOLLUM, Barry; MCMULLAN, Paul. Serious Games: Quality Characteristics Evaluation Framework and Case Study. **IEEE Integrated STEM Education Conference (ISEC)**, Princeton, p. 112-119, 2018.
- ACM; IEEE-CS. **Computer Science Curricula 2013**: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science. 2013. Disponível em:
https://www.acm.org/binaries/content/assets/education/cs2013_web_final.pdf. Acesso em: 10 mar. 2015.
- ACUÑA, Silvia T.; FERRÉ, Xavier. Software Process Modelling. **ISAS-SCI**, Orlando, p. 1-6, 2001.
- AK, Oguz. A Game Scale to Evaluate Educational Computer Games. **Procedia - Social and Behavioral Sciences**, v. 46, p.2 477-2481, 2012.
- ARDITO, Carmelo *et al.* An approach to usability evaluation of e-learning applications. **Universal Access in the Information Society**, v. 4, n. 3, p. 270–283, 2006.
- BASIL, Victor R.; CALDIEIRA, Gianluigi; ROMBACH, Dieter H. Goal Question Metric Paradigm. **Encyclopedia of Software Engineering**, v. 1, p. 528–532, 1994.
- BRITO, Cristina L.; ALMEIDA, Iolanda A.C.; CAVALCANTI, Lialda B. O que se avalia e o que é preciso avaliar em um Software Educativo? **Anais do Workshop de Informática na Escola**, Porto Alegre, v. 1, n. 1, p. 334-344, 2003.
- CALDERÓN Alejandro; RUIZ Mercedes. A Systematic Literature Review on Serious Games Evaluation: an Application to Software Project Management. **Computers & Education**, v. 87, p. 396-422, 2015.
- CARVALHO, Carlos V. Is game-based learning suitable for engineering education? **Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)**, Marrakech, p. 1-8, 2012.
- CHEW, Boon-Seng. An efficient framework for game-based learning activity. **2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE)**, Hong Kong, p. 147-150, 2017.
- CONNOLLY, Thomas M.; STANSFIELD, Mark H.; HAINEY, Thomas. Development a General Framework for Evaluating Games-based learning. **Proceedings of the 2nd European conference on games-based learning**. Universitat Oberta de Catalunya Barcelona, p. 105-114, 2008.
- CONNOLLY, Thomas M.; STANSFIELD, Mark. H.; HAINEY, Thomas. Towards the development of a games-based learning evaluation framework. **Games-based learning advancements for multi-sensory human computer interfaces: Techniques and effective practices**, Hershey, p. 251-273, 2009.

COOMANS, Stéphanie; LACERDA, Gilberto S. Petese, a Pedagogical Ergonomic Tool for Educational Software Evaluation. **Procedia Manufacturing**, v. 3, p. 5881-5888, 2015.

DEVELLIS, Robert F. **Scale development: theory and applications**. 4ª ed. Los Angeles: SAGE, 2016.

FLEURY, Afonso; SAKUDA, Luiz O.; CORDEIRO, José H. D. O. **1º Censo da Indústria Brasileira de Jogos Digitais**. São Paulo: GEDIGames, NPGT, Escola Politécnica, 2014. Disponível em: http://www.bndes.gov.br/SiteBNDES/bndes/bndes_pt/Galerias/Arquivos/conhecimento/seminario/seminario_mapeamento_industria_games042014_RelApoioCensoIndustriaBrasileiradeJogos.pdf. Acesso em: 21 jan. 2020.

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

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

FU, Fong-Ling; SU, Rong-Chang; YU, Sheng-Chin. EGameFlow: A scale to measure learners' enjoyment of e-learning games. **Computers & Education**, v. 52, n. 1, p. 101-112, 2009.

GARCIA-MUNDO, Lilia; GENERO, Marcela; PIATTINI, Mario. Towards a Construction and Validation of a Serious Game Product Quality Model. **7ª International Conference on Games and Virtual Worlds for Serious Applications (VS-Games)**, Skovde, p. 1-8, 2015.

HADDAWAY, Neal R. *et al.* The role of Google Scholar in evidence reviews and its applicability to grey literature searching. **PloS one**, v. 10, n. 9, p. e0138237, 2015.

HAINES, Thomas; CONNOLLY, Thomas M.; BOYLE, Liz A. A Refined Evaluation Framework for Games-based Learning. **Proceedings of the 4ª European conference on games based learning**, Copenhagen, p 1-11, 2010.

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

KIMBERLIN, Carole L.; WINTERSTEIN, Almut G. Validity and reliability of measurement instruments used in research. **American Journal of Health-System Pharmacy**, v. 65, n. 23, p. 2276-84, 2008.

KITCHENHAM, Barbara; PFLEEGER, Shari L.; FENTON, Norman. Towards a Framework for Software Measurement Validation. **IEEE Transactions on Software Engineering**, v. 21, n. 12, p. 929-944, 1995.

LIMA, Jefferson F. *et al.* Quali-EDU: Um processo de avaliação da qualidade de software educacional. **Simpósio Brasileiro de Informática na Educação**, v. 26, n.1, p. 229-238, 2015.

LYRAS, Dimitrios P. *et al.* Educational Software Evaluation: A Study from an Educational Data Mining Perspective. **The International Journal of Multimedia & Its Applications**, v. 6, n.3, p. 1-20, 2014.

MAYER, Igor *et al.* **A Brief Methodology for Researching and Evaluating Serious Games and Game-Based Learning**. Psychology, Pedagogy, and Assessment in Serious Games, p.357-393, 2013.

MAYER, Igor *et al.* The research and evaluation of serious games: Toward a comprehensive methodology, **British Journal of Educational Technology**, v. 45, n. 3, p. 502-527, 2014.

MAYER, Igor. Towards a Comprehensive Methodology for the Research and Evaluation of Serious Games, **Procedia Computer Science**, v. 15, p. 233-247, 2012.

NESBIT, John C.; BELFER, Karen; LEACOCK, Tracey. Learning object review instrument (LORI). **E-learning research and assessment network**, 2003.

PASQUALI, Luiz; PRIMI, Primi. Fundamentos da Teoria da Resposta ao Item –TRI. **Avaliação Psicológica: Interamerican Journal of Psychological Assessment**, v. 2, n. 2, p. 99-110, 2003.

PEREIRA, Wendell S. *et al.* Avaliação de Software Educativo: Análise de Abordagens para Definição de Diretrizes. **Nuevas Ideas en Informática Educativa**. Santiago do Chile, p. 557-562, 2016.

PETERSEN, Kai; VAKKALANKA, Sairam; KUZNIARZ, Ludwik. Guidelines for conducting systematic mapping studies in software engineering: an update. **Information Software Technology**. V. 64, p. 1–18, 2015.

PETRI, Giani, WANGENHEIM, Christiane G. V. A Method for the Evaluation of the Quality of Games for Computing Education. **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**, v. 8, n. 1, p. 951, 2019.

PETRI, Giani; WANGENHEIM, Christiane G. V. How to evaluate educational games: a systematic literature review. **Journal of Universal Computers Science**, v. 22, n. 7, p. 992-1021, 2016.

PETRI, Giani; WANGENHEIM, Christiane G. V.; BORGATTO, Adriano F. MEEGA+, Systematic Model to Evaluate Educational Games. **Encyclopedia of Computer Graphics and Games**, v. 3, p. 1-7, 2018.

SAVI, Rafael; WANGENHEIM, Christiane G. V.; BORGATTO, Adriano F. A Model for the Evaluation of Educational Games for Teaching Software Engineering. **25º Brazilian Symposium on Software Engineering**, p. 194–203, 2011.

WOHLIN, Claes et al. **Experimentation in Software Engineering**. Hidelberg: Springer, 2012.