

## Criação de um guia para selecionar métodos de avaliação para tecnologias de software *M-learning*

**Title: *Creating a guide for selecting evaluation methods for M-learning software technologies***

Ana Iza de Sousa Alencar  
Universidade Federal do Ceará  
anaizaalencar8@gmail.com

Anna Beatriz dos Santos Marques  
Universidade Federal do Ceará  
beatriz.marques@ufc.br

### Resumo

*M-learning é uma das principais ferramentas de aprendizagem no contexto educacional e se refere à mobilidade na aprendizagem, em qualquer lugar, a qualquer hora e por qualquer pessoa. Considerando que o M-learning é relativamente novo, é importante investigar métodos adequados de avaliação da qualidade. Assim, esta pesquisa, objetiva criar um guia para selecionar um método adequado para avaliar a qualidade de tecnologias M-learning. Neste artigo, apresenta-se a criação do guia proposto e um estudo de viabilidade. Ao realizar uma revisão bibliográfica, identificou-se métodos de avaliação, características, critérios e subcritérios de qualidade a serem considerados em uma aplicação M-learning. Cada método encontrado foi relacionado aos critérios de qualidade que avalia. O guia foi criado como uma tabela dinâmica em formato de planilha. Para avaliar a viabilidade do guia, um estudo de caso em duas turmas de Qualidade de Software foi conduzido. Os resultados mostram que mais de 80% dos participantes concordam que o guia é útil, fácil de usar e que pretendem utilizar o guia no futuro. Entretanto, foram observadas algumas dificuldades na utilização do guia, devido à falta ou pouca experiência com a ferramenta utilizada para a criação do guia. Como trabalho futuro, pretende-se desenvolver um software a fim de melhorar a facilidade de uso do guia e sua eficácia na sugestão de métodos de avaliação.*

**Palavras-Chave:** *M-learning, métodos de avaliação, qualidade de software de tecnologias educacionais*

### Abstract

*M-learning is one of the main learning tools in the educational context that refers to mobility in learning, anywhere, anytime and by anyone. Considering that M-learning is, relatively, new, it is important to search for suitable quality evaluation methods. Thus, in this research, we aim to create a guide to select a suitable method for evaluating the quality of M-learning technologies. In this paper, we present the development of an initial version of a guide developed and a feasibility study. By conducting a literature review, we identified evaluation methods, quality characteristics, criteria and subcriteria to be considered in a M-learning application. Thus, we characterized each method found and quality criteria evaluated by each method. We created a guide as a dynamic table in excel. In order to evaluate the feasibility of the guide, we conducted a case study in a Software Quality class. The results show that more than 80% of evaluators agree that the guide is useful, easy to use and that they intend to use the guide in the future. However, some difficulties in using the guide were observed, due to lack or little experience with the tool used for creating the guide, excel. As future work, we intend to develop a software tool in order to improve the guide's ease of use and its efficacy in suggesting evaluation methods.*

**Keywords:** *M-learning, Evaluation methods, Software quality in educational technologies*

## 1 Introdução

A aprendizagem móvel ou *M-learning* é considerada como uma das principais tendências atuais de aplicações das novas tecnologias no contexto educacional. *M-learning* refere-se à mobilidade na aprendizagem, em qualquer lugar, a qualquer momento e por qualquer pessoa. Entretanto, é importante ressaltar que para estabelecer uma definição de *M-learning* não podem ser ignorados três conceitos-chave: as tecnologias móveis, a ubiquidade ligada à mobilidade e os usos educacionais em contextos variados (Andrade, Araújo Junior & Silveira, 2015). Jogos, quizzes, tradutores, e-books e dicionários são exemplos de aplicações *M-learning*.

O processo de avaliação da qualidade de um software educativo consiste na utilização de uma ferramenta computacional, a partir da caracterização do software educativo e, também, de suas implicações para o uso educacional, com o intuito de garantir que um conjunto de requisitos necessários estão sendo atendidos (Pereira et al., 2016). A avaliação de uma estratégia instrucional visa determinar o nível de seu sucesso, avaliando se o público-alvo atingiu os objetivos definidos (Petri & von Wangenheim, 2017).

Na literatura são encontrados diversos métodos de avaliação que podem se adaptar a softwares educativos móveis. É possível citar alguns desses métodos que avaliam software educativo: Método de Reeves (Campos, 1989); *Learning Object Review Instrument* (LORI) (Nesbit, Belfer & Leacock, 2003); MEEGA+ (Petri, von Wangenheim & Borgatto, 2020) e MoLEva (Soad & Barbosa, 2017). Com esses e outros métodos existentes na literatura é difícil identificar o público-alvo que deverá utilizar cada método, tendo em vista que, existem métodos com critérios que somente especialistas conseguem avaliar e métodos que avaliam um software com base na percepção de seus usuários.

A escolha de uma abordagem incorreta pode resultar na ausência da observação de critérios definidos como necessários pelo avaliador, uma vez que deve haver uma consistência entre o que se pretende avaliar e os critérios presentes na abordagem (Da Silva et al., 2016). Para isso, faz-se necessário um estudo abrangente e classificatório dos métodos consolidados de avaliação para *M-learning*, direcionados a professores, alunos, tutores e especialistas. Desse modo, os avaliadores conseguem ter uma visão mais ampla sobre os critérios ou atributos que cada método atende e, assim, selecionar uma avaliação adequada ao seu objetivo.

Embora existam diversos métodos propostos para a avaliação de tecnologias *M-learning* e pesquisas que visam caracterizar os métodos existentes (Cáceres & Pow-Sang, 2018), não há uma abordagem que apoie a escolha de um método adequado de acordo com critérios de avaliação de interesse do avaliador. Desta forma, a escolha de um método de avaliação depende do conhecimento e expertise do avaliador em relação aos métodos existentes e disponíveis, o que pode limitar o acesso ao método adequado. Diante disso, a presente pesquisa propõe a criação de um guia para interessados em avaliar software educativo para *M-learning*. O guia sugere métodos de avaliação existentes de acordo com os critérios de qualidade desejados. Os métodos de avaliação são sugeridos a partir de uma base inicial criada por meio de uma revisão da literatura. Com o uso do guia, os avaliadores com objetivos diferentes têm um direcionamento sobre qual método de avaliação de qualidade devem adotar em uma determinada situação. Este artigo apresenta o processo adotado para a criação do guia e um estudo de caso conduzido com o intuito de avaliar sua viabilidade.

Com os resultados apresentados, espera-se contribuir para a adoção de métodos de avaliação adequados para tecnologias *M-learning* e, assim promover a melhoria da qualidade das tecnologias *M-learning* existentes. Além disso, pesquisas futuras podem ser conduzidas para aumentar a abrangência da base de métodos de avaliação e fornecer maior suporte ferramental para o uso do guia proposto.

O artigo está organizado da seguinte forma: a Seção 2 apresenta o background que ajuda a compreender termos essenciais retratados durante o artigo. Os trabalhos relacionados são discutidos na Seção 3. A Seção 4 define a metodologia explicando cada etapa da pesquisa. Os resultados são mostrados na Seção 5. Por fim, a Seção 6 apresenta as considerações finais e vertentes para trabalhos futuros.

## 2 Background

Entende-se software educativo de duas formas: como sendo a classe de interfaces para o ensino-aprendizagem ou um conjunto de artefatos criados para funcionarem como mediadores em atividades educativas de formação em áreas distintas do conhecimento (Gomes & Wanderley, 2003). Podendo ainda, serem vistos como sistemas que são usados autonomamente pelos aprendizes como instrumento à aprendizagem de algo.

Para Pereira et al. (2016), os softwares educativos podem auxiliar no processo de ensino-aprendizagem a partir da contextualização de determinados conteúdos, diante da realidade de cada aluno e fazendo com que este construa determinado conhecimento relativo ao conteúdo didático abordado. São exemplos de software educativo: jogos, tutoriais, simulação, ebooks, aplicativos de resolução de problemas, entre outros.

A aprendizagem móvel ou *M-learning (mobile learning)*, é uma abordagem relativamente nova na área de aplicações educacionais das novas tecnologias. Pode ser definida quando a interação entre os integrantes se dá através de dispositivos móveis e quando estes não estão em um local pré-determinado a qualquer hora (Andrade et al., 2015). Por ser um processo de aprendizagem contínuo, independente de tempo, espaço e quem o utiliza, o uso de aplicativos para *M-learning* está ganhando cada vez mais espaço no meio tecnológico (Cota et al., 2014).

Segundo Gomes et al. (2002), um dos desafios anteriores ao uso de softwares educativos é a avaliação desses sistemas, visto que surge como essencial para que o potencial dessa tecnologia contribua de forma efetiva para o processo educacional. Com a execução da avaliação é possível avaliar informações necessárias para a escolha de um software, julgar se a aplicação se adequa aos objetivos de busca para uma melhor aprendizagem e avaliar o software em termos de qualidade. Por isso, faz-se necessária a escolha adequada de um método que avalie completamente e corretamente um dado sistema segundo seus objetivos de qualidade.

Rocha e Campos (1993) definem que os objetivos de qualidade de um software são atingidos por meio de fatores de qualidade, que podem ser compostos por outros fatores, que por sua vez são avaliados por meio de critérios. Os critérios definem atributos ou características de qualidade para os fatores. Como atributos de qualidade de software, é possível citar os atributos definidos pelo modelo de qualidade da ISO/IEC 25010 (2011): adequação funcional, desempenho, compatibilidade, usabilidade, confiabilidade, segurança, manutenibilidade e portabilidade.

Andrade et al. (2015) argumentam que para definir os critérios de qualidade de tecnologias *M-learning* é necessário considerar como a aprendizagem poderá acontecer no contexto móvel e como o estudante poderá construir seu conhecimento individualmente e coletivamente. Assim, em se tratando de software educativo, além das características técnicas definidas pela ISO/IEC 25010 (2011), surgem características pedagógicas, socioeconômicas e socioculturais inerentes aos objetivos de qualidade pretendidos com o uso das tecnologias educacionais. Segundo Soad, Duarte Filho e Barbosa (2015), as características pedagógicas abordam as características relacionadas ao ensino e aprendizado, as características socioeconômicas estão associadas à viabilidade da tecnologia em relação ao seu custo e benefício, enquanto as características

socioculturais consideram aspectos culturais dos estudantes, assim como a maneira e contexto no qual a tecnologia será adotada.

Pereira et al. (2016) argumentam que a tomada de decisão dos educadores sobre o que deve ser avaliado em um software educativo, sob qual perspectiva, considerando as especificidades relativas ao público-alvo, tipo do software educativo e área de conhecimento, é um desafio para a informática na educação. Outro aspecto importante que precisa ser destacado durante a avaliação é quem será avaliador.

Atualmente, existem métodos de avaliação com o mesmo objetivo (avaliar software educativo), mas com perspectivas diferentes. Existem métodos direcionados a vários públicos-alvo: métodos que avaliam a qualidade de software e são destinados a um especialista da área de qualidade; métodos para a avaliação da experiência do usuário que são voltados para os alunos que irão utilizar o software; e métodos que avaliam os aspectos pedagógicos de uma determinada área de ensino que são indicados para professores e tutores.

Tendo em vista tais desafios, esta pesquisa visa contribuir na seleção de métodos de avaliação de softwares educativos *M-learning*, com base em critérios de qualidade pré-estabelecidos e no tipo do avaliador.

### 3 Trabalhos Relacionados

Nesta seção são resumidamente descritos alguns trabalhos que focaram na avaliação da qualidade de softwares educativos *M-learning*. No final da seção, é feita uma discussão sobre o diferencial da pesquisa descrita neste artigo em relação aos trabalhos relacionados.

Em Baloh *et al.* (2015) é proposto um catálogo de avaliação para qualquer tipo de aplicativo educacional que se concentra em determinar os critérios de qualidade para aplicativos móveis de autoestudo, ou seja, que podem ser usados sem a orientação ou supervisão de um mentor. A estrutura proposta para o catálogo é hierárquica, dividida em cinco níveis. Os níveis superiores são simplificados, destinados a serem usados por usuários do aplicativo, os níveis inferiores são especializados, destinados a serem usados por desenvolvedores e revisores de aplicativos. Após a elaboração do catálogo foram selecionadas 21 aplicações *M-learning*. Essas aplicações foram classificadas de acordo com as 39 propriedades propostas no catálogo. Os autores do catálogo mencionam que apesar de obterem uma maior objetividade na avaliação, mais estudos precisam ser conduzidos para chegar na estrutura de catálogo desejada.

Soad, Duarte Filho e Barbosa (2015) conduziram uma revisão sistemática da literatura visando estabelecer um base de critérios de qualidade para tecnologias *M-learning*. Os autores apresentam um conjunto de características de qualidade considerando aspectos técnicos, pedagógicos, socioeconômicos e socioculturais e concluem que existem poucos estudos relacionados à qualidade de tecnologias *M-learning*. Em Soad, Duarte Filho e Barbosa (2016), os mesmos autores decidem abordar a carência de métodos que avaliam aspectos de qualidade específicos para aplicativos *M-learning*. Os autores consideram o processo de avaliação de qualidade proposto em Duarte Filho e Barbosa (2013), sendo preciso adaptar esse processo para aplicação *M-learning*. Com o intuito de validar de forma preliminar o método de avaliação de qualidade proposto, foram selecionados os seguintes aplicativos móveis de ensino de inglês: Duolingo, Wlingua e Rosetta Stone. Os aplicativos foram avaliados por um especialista e teve como resultado geral uma classificação na gama de aplicativos de aprendizagem móvel de qualidade média para os três aplicativos. No entanto, o Rosetta Stone apresentou um nível inferior em relação aos demais. A análise dos resultados mostra que o método é adequado para avaliar a

qualidade de *softwares M-learning*, mas os autores destacam a necessidade de uma validação mais formal e completa, apesar dos resultados observados serem promissores.

Cáceres & Pow-Sang (2018) conduziram um mapeamento sistemático para investigar quais métodos são adotados para avaliar a usabilidade de tecnologias *M-learning*. Como resultado, os autores identificaram que testes com usuários, questionários e entrevistas são os métodos mais adotados. Não foram identificados questionários específicos para tecnologias *M-learning*, indicando que métodos genéricos são adotados para avaliar a qualidade deste tipo de tecnologia.

Com o objetivo de aliar os conceitos desenvolvidos e consolidados nas pesquisas da área da educação com os da ergonomia de software, o trabalho de Nunes e Santos (2018) apresenta uma revisão das principais abordagens de avaliação de softwares educativos. Em seguida, os autores propõem, de acordo com a perspectiva de um profissional de pedagogia, uma avaliação ergonômico-pedagógica de aplicativos relacionados às disciplinas abordadas no Exame Nacional do Ensino Médio (ENEM). Foram analisados os três aplicativos: *Study Enem*, Simulado Enem 2017 e Enem 2017 Simulado Gabaritando. O resultado da avaliação do profissional de pedagogia determinou que os aplicativos precisam ser melhorados quanto à exploração dos assuntos, a metodologia e dar controle e liberdade ao usuário para se recuperar de um erro. Segundo os autores, os resultados dos aspectos ergonômicos foram satisfatórios apontando que “o design estético e minimalista bem como diálogos simples e naturais implicam na satisfação do usuário”.

Todos os trabalhos relacionados descritos apresentam métodos de avaliação para tecnologias *M-learning* e descrevem os atributos de qualidade contemplados pelos métodos propostos. Somente Nunes e Santos (2018) consideram aspectos ergonômicos do software. Visto que um software não precisa cobrir todos os atributos de qualidade existentes, o presente trabalho se diferencia dos trabalhos relacionados por sugerir métodos de avaliação existentes para tecnologias *M-learning* com base em atributos de qualidade definidos pelo avaliador. Assim, todos os trabalhos aqui mencionados servem de contribuição para a elaboração de um guia que ajude professores, alunos, tutores e especialistas na escolha de um método que melhor avalie a tecnologia de software *M-learning* desejada.

## 4 Metodologia

Esta seção descreve as etapas que foram realizadas para o alcance dos objetivos desta pesquisa. Como é visto na Figura 1, o desenvolvimento do guia é composto pelas seguintes etapas: (i) busca e seleção dos métodos de avaliação para tecnologias de software *M-learning*; (ii) criação do guia; e (iii) avaliação do guia. Cada etapa possui um conjunto de atividades.

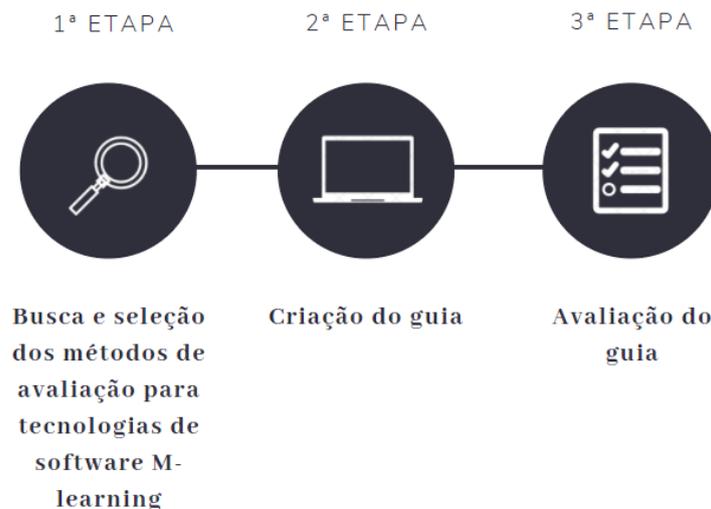


Figure 1: Etapas para a criação de um guia que auxilie na escolha de métodos de avaliação para tecnologias de software *M-learning*.

#### 4.1 Busca e seleção dos métodos de avaliação para tecnologias de software *M-learning*

Com base nas diretrizes para condução de mapeamento sistemático da literatura propostos por Kitchenham e Charters (2007), definimos uma estratégia de busca por métodos de avaliação para tecnologias *M-learning*. A estratégia consistiu em realizar busca automática nas bibliotecas digitais *Scopus*, *IEEE Xplore*, *ACM* e Google Acadêmico com a adoção da seguinte *string* de busca:

("usability" **OR** "user experience" **OR** "quality") **AND** ("mobile educational software" **OR** "m-learning" **OR** "mobile educational games" **OR** "mobile serious games") **AND** ("evaluation")

Analisando sinônimos para o termo “*quality*”, observou-se que, segundo a ISO 25010, existem diversas características e subcaracterísticas de qualidade de software que poderiam ser adotadas como sinônimo na *string* de busca. Para definir um escopo inicial para a etapa de busca e seleção de métodos, decidiu-se focar na característica de usabilidade, por ser uma característica com maior impacto no usuário final e sua experiência de uso com a tecnologia. Por isso, a *string* de busca contém como sinônimo de “*quality*”, os termos “usabilidade e “*user experience*”. O termo “*user experience*” foi incluído por estar relacionado à usabilidade e ser amplamente utilizado em pesquisas da área de qualidade de uso (Vermeeren et al., 2010). Existem outras características de qualidade inerentes a tecnologias *M-learning*, conforme discutido na Seção 2. Porém, para a definição de um escopo inicial para a criação da base de métodos de avaliação para o guia, decidiu-se adotar inicialmente os termos mencionados.

Os termos “*mobile educational games*” e “*mobile serious games*” foram incluídos como sinônimos de tecnologias *M-learning* porque analisando artigos sobre jogos educativos móveis, notou-se que termos genéricos como *M-learning* não costumam ser adotados. Então, decidiu-se adotar termos específicos para abranger métodos direcionados a jogos educativos móveis.

A busca automática retornou 60 artigos na *ACM*, 23 artigos na *IEEE Xplore*, 92 artigos na *Scopus* e mais de 10 mil resultados no Google Acadêmico. Foram considerados para análise, os 100 primeiros resultados do Google Acadêmico. Uma pesquisadora fez a leitura do título e resumo dos artigos retornados para selecionar os artigos que tivessem, de fato, relacionados ao tema da pesquisa. Nesta etapa, os seguintes critérios de inclusão foram considerados:

**CI1.** Artigo apresenta proposta de método para avaliação de tecnologias *M-learning*.

**CI2.** Artigo relata a adoção de método para avaliação de tecnologias *M-learning*.

Houve a necessidade de ler a conclusão dos artigos que não deixaram explícito o seu objetivo no resumo. Após essa análise inicial, 62 artigos foram selecionados para leitura completa. Os demais artigos foram excluídos, por não atenderem a nenhum dos critérios de inclusão. Uma segunda pesquisadora, com experiência na condução de mapeamento sistemático, validou os artigos incluídos e excluídos nesta etapa para reduzir um possível viés da primeira pesquisadora.

Após a seleção, os artigos foram armazenados na ferramenta para gerenciamento de referências bibliográficas *Mendeley*<sup>1</sup>. A ferramenta *Mendeley* foi adotada durante a leitura completa e extração de dados dos artigos. O texto completo dos artigos foi lido para uma segunda etapa de seleção com base nos seguintes critérios de exclusão:

**CE1.** Artigos duplicados.

**CE2.** Ano de publicação anterior a 2014.

**CE3.** Artigos com processo de avaliação durante o desenvolvimento de software.

**CE4.** Publicações que precisavam ser pagas para ter acesso.

Os artigos que não atendiam a nenhum dos critérios de exclusão, eram submetidos ao processo de extração dos dados. Foram selecionados 28 artigos para a etapa de extração. Os dados extraídos de cada artigo foram: título do artigo, ano de publicação, problema de pesquisa, solução proposta, características de qualidade avaliadas, público-alvo do método, recursos utilizados para adoção do método, tipo de pesquisa (teórica, experimental ou ambas). A completude e clareza dos dados extraídos foram validados por uma segunda pesquisadora.

Após a etapa de extração dos dados, iniciou-se a etapa de seleção de métodos para a base inicial do guia proposto. Para tal, os métodos foram classificados em dois tipos: métodos concretos e métodos *ad-hoc*. Como métodos concretos foram considerados aqueles que possuem características de qualidade, onde os atributos são mensuráveis, permitindo assim que as características fossem medidas e pontuadas. Como método *ad-hoc* foram considerados aqueles métodos sem a definição de características de qualidade, cujo resultado da avaliação é algo subjetivo para cada avaliador. Segundo Silva et al. (2016), existem técnicas que não consideram critérios de qualidade do software, e assim induzem o avaliador a escolher um software educativo que, por exemplo, seja muito bom pedagogicamente, mas que apresentam erros funcionais, o que interfere no processo ensino-aprendizagem e vice-versa. Neste sentido, somente os métodos concretos foram selecionados para compor a base inicial de métodos do guia proposto. Ao final do processo de seleção e análise, 12 métodos concretos foram selecionados, conforme resume a Tabela 1.

Tabela 1: Métodos concretos selecionados para a base inicial do guia proposto.

Método de avaliação	Público-alvo	Citação
Checklist de Soad	Especialistas	Soad, Duarte Filho & Barbosa (2016)
Checklist de Porto	Especialistas	Porto, Barbosa & von Wangenheim (2018)
Questionário de Silva e Batista	Professores	Da Silva & Batista (2015)
PETESE (Pedagogical Ergonomic Tool for Educational Software Evaluation)	Especialistas	Coomans & Lacerda (2015)

<sup>1</sup> Mendeley Desktop: disponível para download em <https://www.mendeley.com/>

Método de avaliação	Público-alvo	Citação
Heuristics Evaluation Strategy	Alunos, professores e tutores	Zaibon & Shiratuddin (2010)
TUP Model	Professores	Bednarik, Gerdt, Miraftabi & Tukiainen (2004)
MAQSEI (Metodologia de Avaliação de Qualidade de Software Educacional Infantil)	Profissionais da área de educação, informática e afins	Ataide, Teixeira & da Silva (2003)
MEEGA+	Alunos, professores, tutores e especialistas	Petri, von Wangenheim & Borgatto (2016)
Avaliação ergonômico-pedagógica	Alunos, professores, tutores e especialistas	Nunes & Santos (2018)
MoLEva	Especialistas	Soad & Barbosa (2017)
Método de Lúlia e Simões	Professores	Vieira, Simões & Barreto (s.d)
Método de Barbosa	Especialistas e usuários	Barbosa, Oliveira & D'Carlo (2016)

Apenas 12 dos 20 métodos encontrados estabeleciam os critérios desejados: ser um método concreto, ou seja, método com características de qualidade consideradas claramente definidas, estar disponível gratuitamente para acesso dos usuários e avaliar uma tecnologia *M-learning*. Embora alguns dos métodos selecionados não sejam específicos para tecnologias *M-learning*, como MAQSEI e MEEGA+, estes métodos foram selecionados porque: (i) avaliam aspectos pedagógicos e técnicos relevantes para tecnologias *M-learning* ou (ii) existem relatos de sua adoção para avaliar a qualidade de tecnologias *M-learning*. O método MEEGA+, por exemplo, é direcionado à avaliação de jogos educacionais e pode ser adotado para avaliar jogos educacionais móveis, que são um tipo de tecnologia *M-learning*.

Os métodos apresentados na Tabela 1 foram considerados na criação de um guia para obter uma base de dados a partir de uma extração de dados desses métodos, como: características, critérios que eles avaliam e o público-alvo para o qual o método é destinado. Os métodos e suas informações são a base para a construção do guia.

## 4.2 Criação do guia

Para definir os critérios e subcritérios de qualidade a serem considerados no guia, foi considerada a base de critérios de qualidade fornecida por Soad, Duarte Filho & Barbosa (2015) por meio de uma revisão sistemática da literatura. A metodologia de revisão sistemática permite a sintetização de evidências e construção de base de conhecimento sobre determinado tópico de pesquisa com a adoção de um protocolo rigoroso que permite a replicação da pesquisa, fornecendo maior confiabilidade sobre os resultados (Kitchenham & Charters, 2007).

A partir da base de critérios de qualidade definida, foi possível relacionar os métodos de avaliação selecionados aos critérios de qualidade que avaliam. Nesta etapa, uma análise conceitual foi necessária, pois nos diferentes métodos, existiam critérios com nomenclaturas diferentes e mesma definição, assim como critérios que não foram definidos por Soad, Duarte Filho & Barbosa (2015) mas eram considerados pelos métodos de avaliação selecionados.

Para garantir a consistência e corretude no relacionamento entre métodos e critérios considerados, decidiu-se criar uma base de critérios e subcritérios de qualidade para *M-learning*. Uma pesquisadora conduziu a análise de nomenclaturas e definições de critérios e subcritérios

descritas nas publicações dos métodos de avaliação selecionados e em seguida, criou um mapa mental para cada critério de qualidade para tecnologias de software *M-learning*.

Uma segunda pesquisadora validou os relacionamentos entre os conceitos, visando reduzir inconsistências e redundâncias. Os mapas mentais foram organizados da seguinte forma: (i) os critérios de qualidade foram organizados em quatro categorias (características técnicas, características pedagógicas, características socioeconômicas e características socioculturais); (ii) criou-se um mapa mental para cada categoria de critérios de qualidade; (iii) criou-se um mapa mental para cada critério de qualidade, detalhando seus subcritérios. A Figura 2 exemplifica o mapa mental para a categoria de características socioeconômicas.

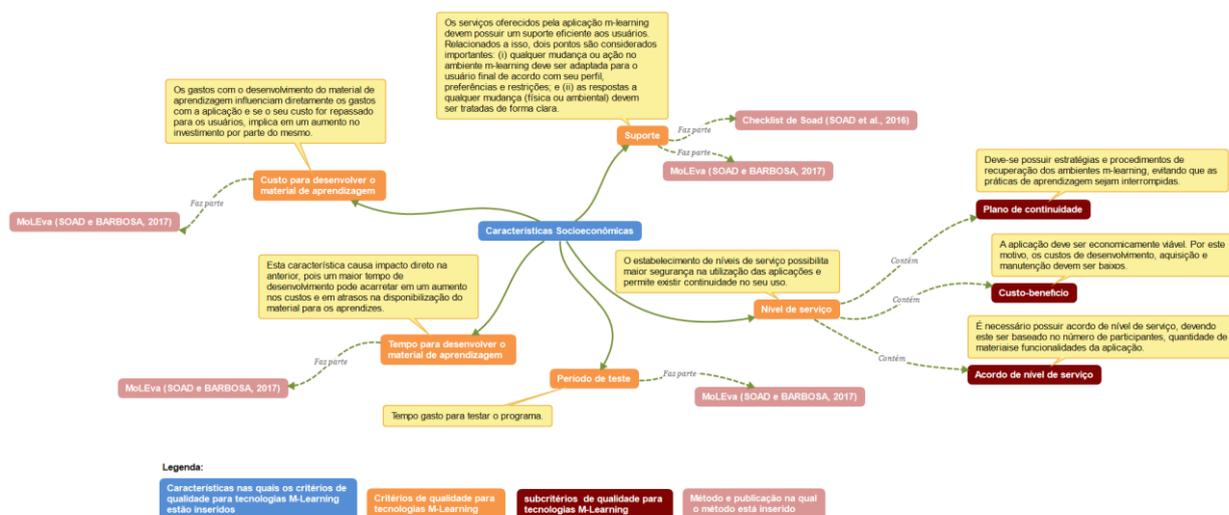


Figure 2: Exemplo de mapa mental da categoria "características socioeconômicas".

Os critérios de qualidade relacionados às características socioeconômicas são: suporte, custo para desenvolver o material de aprendizagem, tempo para desenvolver o material de aprendizagem, período de teste e nível de serviço. Na Figura 2, é possível visualizar por meio do relacionamento “faz parte” quais métodos de avaliação consideram cada critério de qualidade. O critério de qualidade nível de serviço possui como subcritérios: plano de continuidade, acordo de nível de serviço e custo-benefício. O relacionamento “contém” indica quais subcritérios estão relacionados a cada critério de qualidade. Para cada critério e subcritério, o mapa fornece uma definição para possibilitar um entendimento comum a respeito dos conceitos.

A Figura 3 ilustra um mapa mental que apresenta um maior nível de detalhamento sobre um critério de qualidade. Trata-se do critério de portabilidade, que pertence à categoria de características técnicas. A facilidade de instalação/instalabilidade, por exemplo, é um subcritério de portabilidade, contendo as respectivas definições e os métodos que os avaliam. Esse exemplo mostra que o critério de portabilidade contém cinco subcritérios que são: atualização e configuração, adaptação de dispositivos ou adaptabilidade (exemplo de subcritério que possui sinônimo), sucesso de instalação, facilidade de instalação ou instabilidade e substituíbilidade. O subcritério facilidade de instalação é avaliado pelo Checklist de Soad (Soad, Duarte Filho e Barbosa, 2016) e pelo método MoLEva (Soad & Barbosa, 2017). O segundo método adota um nome diferente para este subcritério (instalabilidade). Esta relação entre diferentes nomenclaturas para um mesmo critério ou subcritério de qualidade é representada no mapa pelo relacionamento “ou”.

Ao final desta etapa de análise conceitual e mapeamento entre métodos e critérios de qualidade, foram criados quatro mapas mentais sobre categorias e seus critérios de qualidade e 26 mapas mentais sobre critérios e subcritérios de qualidade. Ao todo, foram criados 30 mapas mentais e todos estes estão disponíveis e podem ser acessados pelo endereço eletrônico: < <https://doi.org/10.6084/m9.figshare.13138589.v2>>.

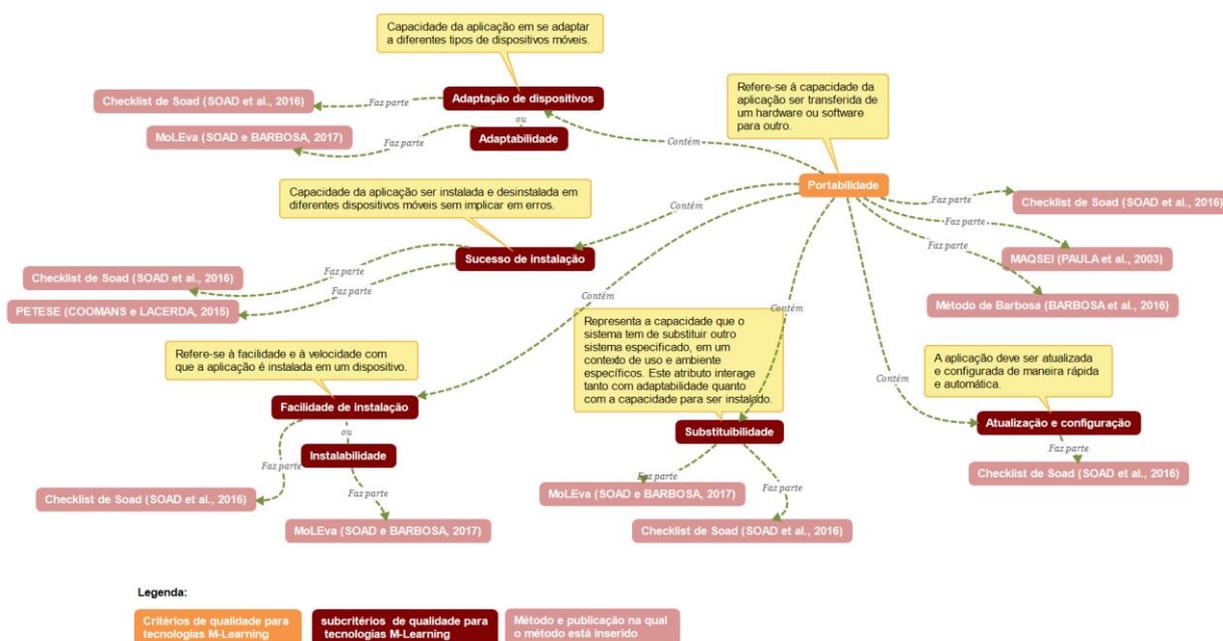


Figure 3: Exemplo do critério de qualidade portabilidade e seus subcritérios de qualidade.

A Tabela 2 apresenta uma visão sintetizada dos critérios e subcritérios identificados e considerados para a criação do guia. As definições dos critérios e subcritérios foram resultantes da análise conceitual conduzida. Devido à extensão do conteúdo da tabela, as definições de cada critério estão disponíveis em < <https://doi.org/10.6084/m9.figshare.13138589.v2>>.

Tabela 2: Resumo dos critérios e subcritérios de qualidade identificados e considerados no guia de avaliação.

Características Técnicas	
Critério	Subcritérios
<b>Adequação funcional:</b> disponibilidade de funções que atendam às necessidades implícitas e explícitas em relação à aplicação	Capacidade de comunicação; Recursos multimídia; Completude funcional; Compactação de mensagem; Precisão; Concisão das mensagens; Facilidade de acesso às funcionalidades; Eficácia de execução de tarefas; Mudança de contexto; Variação da rede; Acessibilidade; Interface amigável; Carregamento da tela; Telas por tarefa; Facilidade de preenchimento; Gestão de dados
<b>Eficiência do Desempenho:</b> desempenho em relação à quantidade de recursos utilizados sob condições estabelecidas, ou seja, refere-se à otimização do uso dos recursos disponíveis.	Gestão de dados; Tempo de carregamento; Tempo médio de resposta; Tempo de resposta para obter informação da memória; Comportamento em relação ao tempo; Tamanho da aplicação; Eficiência na utilização de memória; Eficiência no consumo de energia; Limpeza de memória; Escalabilidade

<u>Compatibilidade</u> : capacidade da aplicação em trocar informações ou realizar operações com outras aplicações, compartilhando o mesmo ambiente de hardware e/ou software.	Coexistência de aplicação; Interoperabilidade; Expansibilidade
<u>Usabilidade</u> : capacidade da aplicação em ser utilizada por usuários específicos em um contexto específico, oferecendo maneiras que permitam que a aplicação seja entendida, aprendida, utilizada e atraente ao usuário.	Tempo de aprendizado; Ajuda e/ou Documentação; Tempo de compreensão; Homogeneidade do layout; Homogeneidade dos componentes; Entrada de dados; Navegação; Inteligibilidade; Continuidade; Apresentação de informações; Atratividade; Proteção de erro do usuário; Prevenção de erros; Atenção focada; Diversão; Desafio; Confiança; Aprendizabilidade; Diversidade de dispositivos
<u>Heurísticas Pedagógicas</u> : auxiliar os alunos a encontrarem as respostas sobre determinado assunto por conta própria, através do encaminhamento feito por meio de algumas perguntas, por exemplo.	Pertinência em relação ao programa curricular; Carga de trabalho; Significado de códigos e denominações; Consistência e Padrões; Correspondência entre o software e o mundo real
<u>Design de interface</u> : maximizar a usabilidade e a experiência do usuário para tornar a interação do usuário efetiva, eficiente e satisfatória.	Não possui
<u>Confiabilidade</u> : capacidade da aplicação em proporcionar um comportamento consistente com o esperado durante um longo período de tempo.	Disponibilidade; Transferência de dados; Estabilidade; Recuperabilidade; Tolerância a falhas; Tempo de recuperação após desconexão
<u>Segurança</u> : capacidade da aplicação em proteger as informações e dados, protegendo-os de acesso não autorizado e garantido seu acesso de acordo com os diferentes níveis de autorização.	Acesso; Criptografia; Mensagens seguras e informações nos dispositivos; Permitir salvar pontos de backup em caso de bateria baixa; Rastreabilidade; Autenticidade; Integridade; Auditoria; Utilização de perfil de usuário
<u>Portabilidade</u> : capacidade da aplicação ser transferida de um hardware ou software para outro.	Adaptabilidade; Sucesso de instalação; Instalabilidade; Substituibilidade; Atualização e configuração
<u>Manutenção</u> : esforço para localizar e corrigir um erro ou inconsistência.	Não possui
<b>Características Pedagógicas</b>	
<u>Interatividade</u> : boa condução do usuário durante a interação com suas interfaces, que facilite o aprendizado e utilização do programa, e, conseqüentemente, melhore o desempenho e diminua o número de erros.	Feedback; Interação social
<u>Aprendizagem</u> : características relacionadas ao aprendizado.	Identificação de deficiências de aprendizagem; Identificação de perfil; Processos de ensino e aprendizagem; Ambiente de aprendizagem; Ganho de conhecimento; Conhecimento na hora certa; Atividades educacionais; Adaptação ao contexto; Iniciativa de aquisição de conhecimento; Estilo de aprendizagem; Atividades de instrução
<u>Conteúdo</u> : materiais utilizados nas aplicações.	Qualidade de conteúdo; Público Alvo; Complexidade multimídia; Fonte de conteúdo; Gerenciamento de conteúdo; Integração de conteúdo; Recursos didáticos; Reutilização de conteúdo de aprendizagem; Complexidade de conteúdo;
<u>Aplicação</u> : características técnicas da aplicação que podem interferir nas questões pedagógicas.	Visualizações separadas; Esforço cognitivo das atividades; Qualidade audiovisual

<u>Papel do professor:</u> capacidade de o software pode ser usado sem a intervenção de um instrutor. Os professores atuam no papel de guias, monitores, treinadores, tutores ou facilitadores.	Não possui
<b>Características Socioeconômicas</b>	
<u>Custo para desenvolver o material de aprendizagem:</u> Os gastos com o desenvolvimento do material de aprendizagem influenciam diretamente os gastos com a aplicação e se o seu custo for repassado para os usuários, implica em um aumento no investimento por parte do mesmo.	Não possui
<u>Período de teste:</u> Tempo gasto para testar o programa.	Não possui
<u>Tempo para desenvolver o material de aprendizagem:</u> um maior tempo de desenvolvimento pode acarretar em um aumento nos custos e em atrasos na disponibilização do material para os aprendizes.	Não possui
<u>Nível de serviço:</u> o estabelecimento de níveis de serviço possibilita maior segurança na utilização das aplicações e permite existir continuidade no seu uso.	Plano de continuidade; Acordo de nível de serviço; Custo-benefício.
<u>Suporte:</u> possuir um suporte eficiente aos usuários	Não possui
<b>Características Socioculturais</b>	
<u>Curva de aprendizado:</u> Refere-se ao tempo em que o usuário consegue manusear a tecnologia móvel, ou seja, está diretamente relacionada à tecnologia dos dispositivos móveis.	Não possui
<u>Perfil do usuário:</u> De acordo com a idade dos usuários, diferentes limitações podem existir. Por este motivo, é necessário que a aplicação seja adaptável.	Não possui
<u>Satisfação do usuário:</u> A fim de suprir as expectativas do usuário, a aplicação deve satisfazê-lo, evitando assim abandono na utilização da aplicação.	Não possui
<u>Colaboração:</u> A colaboração deve estar presente nas aplicações <i>m-learning</i> , permitindo que os aprendizes e tutores se relacionem.	Não possui
<u>Atividades individuais e em grupos:</u> As atividades proporcionam a possibilidade de se aprender individualmente, mas também motivam o desenvolvimento de tarefas em conjunto.	Não possui

Com as informações sobre os métodos de avaliação selecionados e com o conjunto definido de critérios e subcritérios de qualidade, foi possível iniciar a criação do guia. O guia foi desenvolvido no Excel por meio de filtros em uma tabela dinâmica e está disponível em: < <https://doi.org/10.6084/m9.figshare.13138589.v2>>. Como ilustra a Figura 4, a tabela dinâmica contém quatro colunas, sendo elas: categoria, critério/subcritério, método que avalia e público-alvo do método. Os critérios e subcritérios ficaram na mesma coluna, porque existem critérios de um método que são considerados como subcritérios de outro método e vice versa.

1	Característica	Critério/subcritério	Método que avalia	Público alvo do método
2	Técnicas	Adequação funcional	MAQSEI	Professor de informática ou áreas afins
3	Técnicas	Adequação funcional	MAQSEI	Professor
4	Técnicas	Eficiência do Desempenho	MAQSEI	Professor de informática ou áreas afins
5	Técnicas	Eficiência do Desempenho	MAQSEI	Professor
6	Técnicas	Eficiência do Desempenho	TUP Model	Professor
7	Técnicas	Equipamentos de software e hardware	MAQSEI	Professor de informática ou áreas afins
8	Técnicas	Equipamentos de software e hardware	MAQSEI	Professor
9	Técnicas	Equipamentos de software e hardware	TUP Model	Professor
10	Técnicas	Utilização de recursos computacionais	MAQSEI	Professor de informática ou áreas afins
11	Técnicas	Utilização de recursos computacionais	MAQSEI	Professor
12	Técnicas	Utilização de recursos computacionais	TUP Model	Professor
13	Técnicas	Compatibilidade	PETESE	Especialista
14	Técnicas	Compatibilidade	MAQSEI	Professor de informática ou áreas afins
15	Técnicas	Compatibilidade	MAQSEI	Professor
16	Técnicas	Usabilidade	Checklist de Soad	Especialista
17	Técnicas	Usabilidade	Checklist de Porto	Professor
18	Técnicas	Usabilidade	Questionário de Silva e Batista	Professor
19	Técnicas	Usabilidade	TUP Model	Professor
20	Técnicas	Usabilidade	MAQSEI	Professor de informática ou áreas afins
21	Técnicas	Usabilidade	MAQSEI	Professor
22	Técnicas	Usabilidade de Jogo	Heuristics Evaluation Strategy	Professor
23	Técnicas	Usabilidade de Jogo	Heuristics Evaluation Strategy	Tutor
24	Técnicas	Usabilidade de Jogo	Heuristics Evaluation Strategy	Aluno
25	Técnicas	Heurísticas Pedagógicas	MAQSEI	Professor de informática ou áreas afins
26	Técnicas	Heurísticas Pedagógicas	MAQSEI	Professor
27	Técnicas	Design de interface	Checklist de Porto	Professor
28	Técnicas	Confiabilidade	MolEva	Especialista
29	Técnicas	Confiabilidade	MAQSEI	Professor de informática ou áreas afins
30	Técnicas	Confiabilidade	MAQSEI	Professor
31	Técnicas	Segurança	Checklist de Soad	Especialista
32	Técnicas	Segurança	TUP Model	Professor
33	Técnicas	Segurança	MAQSEI	Professor de informática ou áreas afins

Figure 4: Tabela dinâmica utilizada na criação do guia.

A Figura 5 ilustra os filtros que permitem que o usuário realize a busca por métodos de avaliação. A busca pode ser realizada por: conjunto de critérios, público-alvo, categorias de critérios de qualidade.

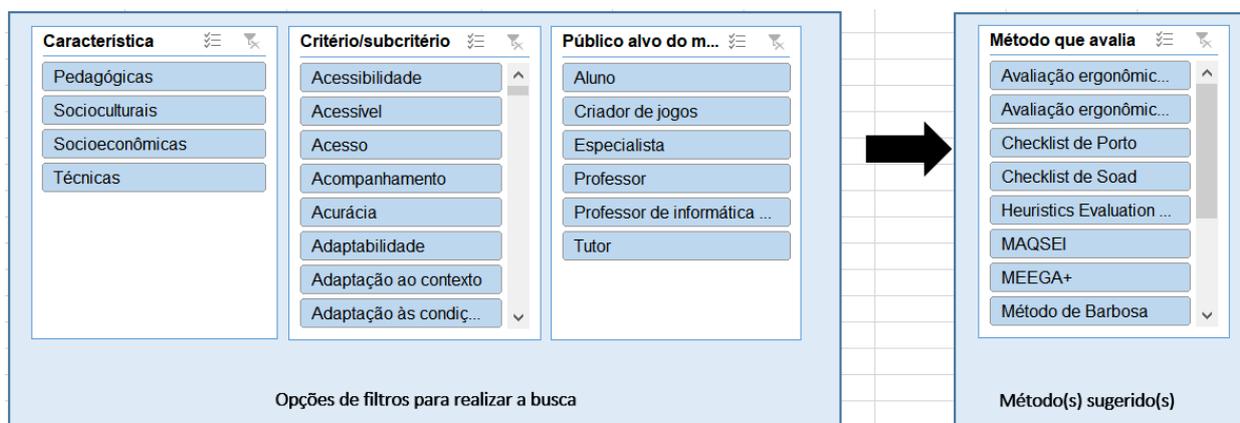


Figure 5: Filtros para sugerir os métodos.

Para o guia sugerir um ou mais métodos adequados aos requisitos de busca do avaliador, é necessário que seja selecionado pelo menos uma opção de um dos filtros, ou seja, o avaliador decide se quer buscar o método adequado por meio de todos os filtros ou somente um. Após a seleção das opções nos filtros, o componente de filtro “método que avalia” mostra um ou mais métodos que atendem parcialmente ou totalmente aos requisitos do avaliador. A Figura 6 ilustra um exemplo de busca, em que foi selecionada a categoria de características pedagógicas, os critérios de qualidade “acompanhamento” e “aprendizagem” e o público-alvo é o “aluno”. Como resultado da busca, o guia destaca no filtro “método que avalia” dois métodos que são adequados para os requisitos desejados: Avaliação ergonômico-pedagógica e MEEGA+.

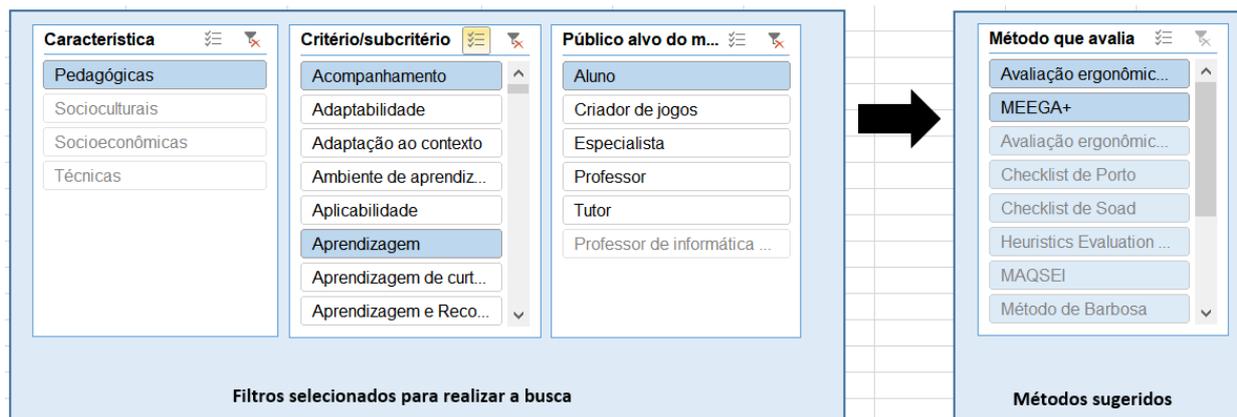


Figure 6: Exemplo de busca para indicação de um método de avaliação pelo guia proposto.

Caso o avaliador selecione um dos métodos retornados, o guia destaca quais requisitos de busca são atendidos pelo método. A Figura 7 ilustra que o método Avaliação ergonômico-pedagógica não avalia o critério de aprendizagem, pois ele não é mais exibido nos critérios em destaque no filtro. Desta maneira, quando houver mais de um método como resultado, o avaliador deve analisar quais requisitos de busca cada método atende, colocando-os em uma escala de prioridades. Somente após essa análise, o avaliador será capaz selecionar o método mais relevante aos seus interesses.

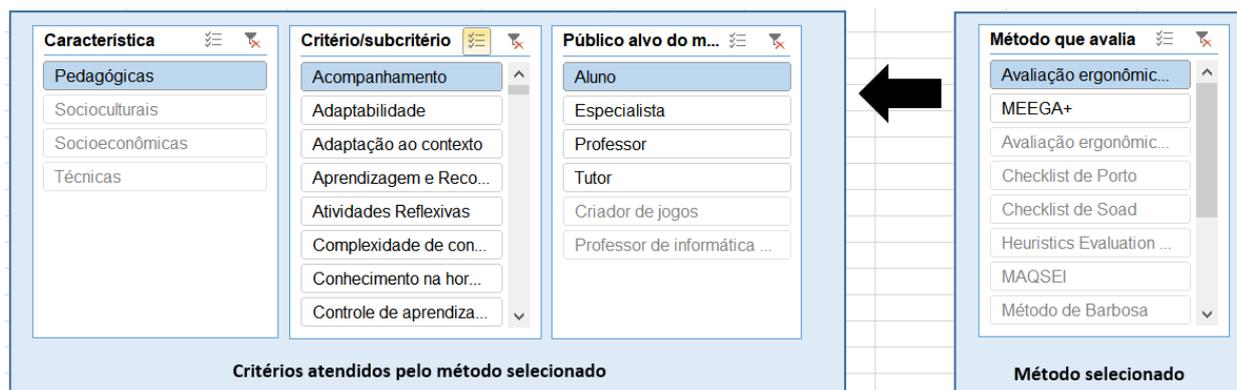


Figure 7: Seleção de um método sugerido pelo guia.

Atualmente, o guia é disponibilizado para uso por meio de um pacote contendo: o guia em formato de planilha do Excel, a base contendo os mapas mentais sobre critérios de qualidade e suas definições e o conjunto de publicações sobre cada método considerado na base inicial do guia. Por meio de sua adoção, os avaliadores conseguem identificar: (i) qual método deve ser utilizado para avaliar um conjunto determinado de critérios, (ii) quais critérios de qualidade cada método contempla, (iii) público-alvo dos métodos e (iv) em qual publicação o método está descrito. Com a criação do guia, decidiu-se avaliar sua viabilidade como apoio ao processo de avaliação de tecnologias *M-learning* por meio de um estudo de caso.

### 4.3 Avaliação do guia

Para avaliar o guia foi realizado um estudo de caso em ambiente acadêmico com o objetivo de investigar os pontos positivos e negativos do guia com base na percepção de responsáveis por uma avaliação de tecnologia *M-learning*. O estudo de caso será descrito segundo as diretrizes propostas por Runeson et al. (2012) para o relato de estudos de caso em Engenharia de Software.

**Contexto e participantes:** O estudo de caso foi conduzido em duas turmas da disciplina de Qualidade de Software da Universidade Federal do Ceará no contexto de um trabalho prático da disciplina no qual os estudantes deveriam avaliar uma tecnologia *M-learning*. O trabalho prático estava relacionado ao conteúdo de Qualidade do Produto, que faz parte da ementa da disciplina e aborda diferentes métodos para avaliar a qualidade de um software. A escolha dessas turmas foi estratégica, uma vez que esses estudantes se encaixam na maioria dos perfis do público-alvo apontados pelos métodos, como aluno, especialista em qualidade e tutor. Os participantes foram estudantes da disciplina de Qualidade de Software, ofertada aos cursos de graduação em Engenharia de Software e Ciência da Computação. A docente da disciplina era uma pesquisadora envolvida nesta pesquisa. No estudo de caso realizado na Turma 1, participaram 48 estudantes e na Turma 2 participaram 39 estudantes, totalizando 87 participantes. O estudo de caso em ambas as turmas seguiu os mesmos procedimentos.

**Preparação:** O guia foi utilizado como base para selecionar um método adequado para avaliar uma tecnologia *M-learning*. Para a realização do estudo de caso foram criados cenários com o intuito de obter o máximo de situações diferentes, ou seja, critérios de qualidade diferentes para serem avaliados, diferentes perfis de públicos-alvo e tecnologias *M-learning* diferentes. Porém, todos os cenários possuíam um escopo delimitado e padronizado para que cada equipe tivesse o mesmo grau de dificuldade durante o estudo de caso: uma tecnologia *M-learning* a ser avaliada e seis diferentes critérios e/ou subcritérios de qualidade. As pesquisadoras sugeriram tecnologias a serem avaliadas, mas cada equipe poderia selecionar uma tecnologia de seu interesse para ser avaliada. A Figura 8 ilustra um exemplo de cenário fornecido às equipes. Foram criados 10 diferentes cenários.

Equipe responsável pela avaliação/inspeção: Equipe 1  
Equipe participante da avaliação: Equipe 2

**Cenário 1**

Suponha que você é um(a) especialista em critérios de qualidade de softwares e deseja avaliar a aplicação móvel <nome da aplicação> - com foco na aprendizagem, na portabilidade da aplicação para que usuários não tenham problema quando acessarem o aplicativo em outro dispositivo móvel, e também na colaboração, permitindo um maior relacionamento entre os usuários. Os critérios/subcritérios que serão avaliados são:

- Ganhar conhecimento
- Processos educativos e aprendizagem definidos
- Atividades instrucionais
- Interação entre usuários
- Colaboração
- Adaptação do dispositivo

Método encontrado por meio do guia: \_\_\_\_\_

Figure 8: Exemplo de um dos cenários utilizados pelos alunos avaliadores.

**Procedimentos de coleta de dados:** Para investigar como foi a experiência dos participantes na utilização do guia, elaborou-se um questionário baseado no Modelo de Aceitação de Tecnologia (*Technology Acceptance Model 3 – TAM3*) proposto por Venkatesh & Bala (2008). Esse modelo procura prognosticar a adoção e a utilização de novas tecnologias de informação por potenciais usuários. Selecionou-se três indicadores do TAM3 para serem adotados na pesquisa: percepção sobre utilidade, percepção sobre facilidade de uso e intenção de uso. O questionário criado possui 10 afirmações e uma questão aberta para sugestões. O respondente avalia cada afirmação por meio de uma escala de Likert de 7 pontos variando de “concordo totalmente” à “discordo totalmente” com uma opção neutra “nem concordo, nem discordo”. As afirmativas adotadas para avaliar cada indicador são descritas na Tabela 3. O questionário continha uma questão aberta para explorar sugestões de melhoria sobre o guia. O questionário foi elaborado de

tal forma que sua introdução descrevia resumidamente os preceitos éticos do estudo de caso, indicando que o seu preenchimento era voluntário, seus dados seriam anonimizados para a análise dos dados da pesquisa e suas respostas não impactariam na nota obtida no trabalho prático.

Tabela 3: Afirmativas adotadas para avaliar os indicadores do Modelo TAM3.

Indicador	Afirmativa
Percepção sobre utilidade	O uso do Guia iria melhorar meu desempenho na busca por um método que avalie critérios de qualidade de tecnologias <i>M-learning</i> (acredito ter encontrado um método adequado aos meus critérios de avaliação em um tempo menor do que levaria sem usar esta abordagem).
	O uso do Guia na busca por um método que avalie critérios de qualidade de tecnologias <i>M-learning</i> iria aumentar minha produtividade.
	O uso do Guia iria melhorar minha eficácia na busca por um método que avalie critérios de qualidade de tecnologias <i>M-learning</i> (acredito ter encontrado um método mais completo do que encontraria sem usar esta abordagem).
	Considero o Guia útil para buscar um método que avalie critérios de qualidade de tecnologias <i>M-learning</i> .
Percepção sobre facilidade de uso	Minha interação com o Guia é clara e compreensível.
	Utilizar o Guia não requer muito esforço mental.
	O Guia é fácil de usar.
	Considero fácil utilizar o Guia para buscar um método que avalie critérios de qualidade de tecnologias <i>M-learning</i> .
Intenção de uso	Dado que eu tenha acesso ao Guia, eu prevejo que iria utilizá-lo.
	Eu preferiria usar o Guia do que pesquisar na literatura para encontrar por um método que se adeque aos meus critérios de avaliação.

**Execução:** Todos os estudantes foram organizados em 18 equipes para um trabalho prático da disciplina: 10 equipes da Turma 1 e 8 equipes da Turma 2. O objetivo do trabalho consistia em avaliar a qualidade de um aplicativo *M-learning*.

Na primeira etapa do trabalho prático, as equipes deveriam realizar o planejamento da avaliação da tecnologia *M-learning*. O planejamento aconteceu em um laboratório de informática e seguiu os seguintes passos: (1) recebimento de um cenário; (2) uso do guia proposto sob supervisão das pesquisadoras; (3) seleção de um método de avaliação adequado. O guia foi disponibilizado por meio do Drive, em uma pasta que continha: o guia em formato de planilha, artigos científicos sobre cada método de avaliação considerado pelo guia e mapas mentais com informações sobre os critérios e subcritérios considerados pelo guia. As equipes utilizaram o guia como apoio para selecionar um método adequado de acordo com o cenário recebido. Em alguns casos, o guia retornou mais de um método para os requisitos desejados pela equipe. Para resolver esse problema, as pesquisadoras orientaram as equipes a selecionar o método que atendia o maior número de critérios e/ou subcritérios desejáveis. Em caso de ter mais de um método com a mesma quantidade de critérios e/ou subcritérios, a escolha era realizada pela equipe, de acordo com sua preferência. A Tabela 4 resume os diferentes cenários recebidos por cada equipe e o método selecionado após o uso do guia.

Tabela 4: Resumo dos cenários recebidos por cada equipe e o método selecionado após o uso do guia proposto.

Equipe	Tecnologia avaliada	Critérios de qualidade considerados	Método de avaliação selecionado
1	<i>WattPad</i>	Ganhar conhecimento, processos educativos e aprendizagem definidos, atividades instrucionais, interação entre usuários, colaboração, adaptação do dispositivo	Checklist de Soad
2	<i>Semper</i>	Acesso, integridade, confiabilidade, atividades educacionais, adaptabilidade, e estilo de aprendizagem	MoLEva

Equipe	Tecnologia avaliada	Crítérios de qualidade considerados	Método de avaliação selecionado
3	<i>Lingualeo</i>	Aprendizagem, adaptação ao contexto, homogeneidade do layout, navegação, continuidade, atenção focada	MoLEva
4	<i>Duolingo</i>	Usabilidade, requisitos mínimos, atração/motivação, ajuda e/ou documentação, layout, design de interface	Método de Barbosa
5	<i>Peak</i> Treinamento Cerebral	Conteúdo, controle e autonomia do usuário, elementos multimídia, aprendizagem, conhecimento na hora certa, avaliação	MEEGA+
6	<i>Run-Aedes</i>	Usabilidade, oportunidade de retornar à atividade em caso de erro, mobilidade, conteúdo, qualidade de conteúdo, diálogos simples e naturais	MoLEva
7	Perguntados	Diversão, atração/motivação, desafio, acessibilidade, aprendizibilidade, efetividade da aprendizagem	MEEGA+
8	Turma da Galinha Pintadinha	Controle do Usuário, mensagens de erro, atratividade, ajuda e/ou documentação, acompanhamento, criatividade	Avaliação ergonômico-pedagógica
9	Física Interativa	Mobilidade, facilidade de acesso às funcionalidades, proteção de erro do usuário, compartilhamento, ajuda e/ou documentação, suporte	Método de Barbosa
10	Estuda.com	Diversão, atração/motivação, desafio, acessibilidade, aprendizibilidade, efetividade da aprendizagem	Checklist de Soad
11	<i>Duolingo</i>	Ganhar conhecimento, processos educativos e aprendizagem definidos, atividades instrucionais, interação entre usuários, colaboração, adaptação do dispositivo	MoLEva
12	<i>GeekieGames</i>	Acesso, integridade, confiabilidade, atividades educacionais, adaptabilidade, estilo de aprendizagem	MoLEva
13	<i>HandTalk</i>	Aprendizagem, adaptação ao contexto, homogeneidade do layout, navegação, continuidade, atenção focada	MoLEva
14	Aprender 50 Línguas	Usabilidade, requisitos mínimos, atração/motivação, ajuda e/ou documentação, layout, design de interface	MAQSEI
15	Mimo	Conteúdo, controle e autonomia do usuário, elementos multimídia, aprendizagem, conhecimento na hora certa, avaliação	MEEGA+
16	<i>Drops</i>	Usabilidade, oportunidade de retornar à atividade em caso de erro, mobilidade, conteúdo, qualidade de conteúdo, diálogos simples e naturais	Heuristic Evaluation Strategy
17	<i>TinyCards</i>	Diversão, atração/motivação, desafio, acessibilidade, aprendizibilidade, efetividade da aprendizagem	MEEGA+
18	Perguntados	Controle do usuário, mensagens de erro, atratividade, ajuda e/ou documentação, acompanhamento, criatividade	MEEGA+

Na segunda etapa do trabalho prático, as equipes deveriam executar a avaliação planejada. Duas aulas da disciplina foram alocadas para a execução da avaliação sob supervisão das pesquisadoras. A avaliação aconteceu de modo que uma determinada equipe era responsável pela avaliação/inspeção, onde seu papel era aplicar o método de avaliação, enquanto outra equipe participante utilizava a tecnologia, exercendo o papel de usuário ou inspetor, de acordo com o método selecionado. Desta forma, na 1ª aula de execução da avaliação, metade das equipes atuou como responsáveis pela avaliação e a outra metade das equipes atuou como participantes das avaliações. As equipes foram organizadas em duas salas de aula e supervisionados pela docente e por dois alunos voluntários, ambos estudantes do curso de Engenharia de Software. A mesma organização foi realizada na 2ª aula de execução da avaliação, sendo que as equipes inverteram os papéis.

Por fim, na terceira etapa do trabalho prático, durante duas aulas da disciplina, as equipes apresentaram os resultados do trabalho prático, abrangendo o planejamento, execução e resultados da avaliação. Após concluídas as apresentações, cada participante avaliou o guia por meio do

questionário previamente descrito nesta seção. O questionário foi respondido por 43 estudantes da Turma 1 e 21 estudantes da Turma 2.

Análise dos dados: Os dados obtidos por meio dos questionários foram tabulados para a geração de gráficos. As respostas às questões abertas foram transcritas para identificar possibilidades de melhorias no guia. A próxima seção apresenta os resultados obtidos a partir desta análise.

## 5 Resultados

Os resultados obtidos como feedback da utilização do guia estão ilustrados na Figura 9. É possível observar que o guia teve uma ampla aceitação pelos participantes, com a maioria de porcentagem de concordância em relação a todas as afirmativas avaliadas.

Os resultados das afirmativas 1 a 4 tratam da percepção sobre a utilidade do guia proposto. Os participantes concordam que o guia é útil, que seu uso poderá melhorar seu desempenho, aumentar sua produtividade e eficácia na avaliação de tecnologias *M-learning*. Não houve discordância em relação à afirmativa 4 “Considero o guia útil”. Sobre as outras 3 afirmativas, observa-se um índice baixo de discordância, não passando de 3,2%. Ou seja, mais de 95% da turma reconhecem a utilidade do guia.

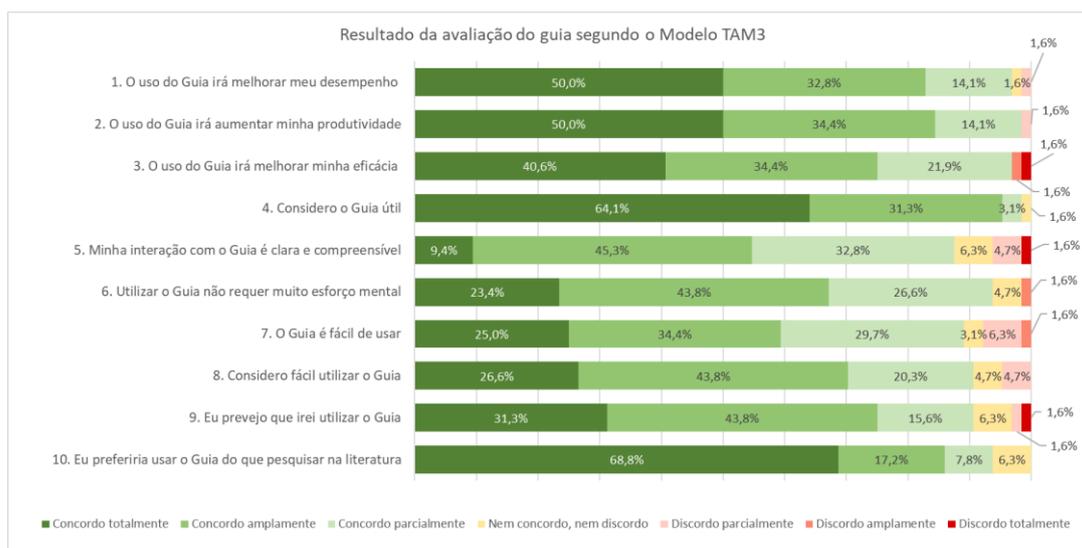


Figure 9: Apresentação do resultado das afirmações do questionário.

Sobre a percepção da facilidade de uso, deve-se analisar os resultados das afirmativas 5 a 8. A Figura 9 mostra que, de modo geral, os participantes experimentaram uma boa interação de uso com o guia, pois a maioria concordou (totalmente ou não) com as quatro afirmativas. Contudo, é necessária analisar os índices de discordância e neutralidade que se destacaram. Com relação à afirmativa 5 “Minha interação com o guia é clara e compreensível”, 6,3% dos participantes discordaram (totalmente ou não). Sobre a afirmativa 6 “Utilizar o Guia não requer muito esforço mental”, 4,7% dos participantes foram neutros e 1,6% discordaram. Observa-se que 7,9% dos participantes discordaram da afirmativa 7 “O guia é fácil de usar”. Em relação à afirmativa 8 “Considero fácil utilizar o guia”, 4,7% dos participantes discordaram. Tais resultados indicam a necessidades de melhorias no guia para promover uma maior facilidade de uso.

Em relação à intenção de usar o guia, também se observa que a maioria dos participantes (mais de 90%) indicaram concordância em relação às duas afirmativas que avaliam este indicador.

93,7% dos participantes indicaram que preferem utilizar o guia do que buscar na literatura por um método que se adeque aos seus requisitos de busca. Outros 6,3% dos participantes nem concordaram e nem discordaram com as afirmações, sendo o indicador com mais respostas neutras.

As respostas dos participantes fornecidas na questão aberta foram analisadas e estão sintetizadas na Tabela 5. Para melhor representar as respostas dos participantes, dividiu-se a tabela em quatro cores. A cor amarela representa as sugestões de melhoria para o guia, a cor verde indica as experiências positivas, a cor vermelha indica as experiências negativas e por fim, a cor azul descreve afirmações sobre o uso do guia.

Por meio dos comentários dos participantes, é possível perceber a necessidade de transformar o guia em uma aplicação móvel ou web. Com isso, espera-se que várias lacunas que existem no guia atual sejam resolvidas como interface mais atrativa, com cores mais dinâmicas, obter um resultado para a busca de forma mais precisa e explicar um eventual caso de ter mais de um resultado (acurácia). Em casos de o guia sugerir mais de um método é interessante mostrar o grau de cobertura de cada método sobre os critérios pretendidos, ou seja, a razão entre a quantidade de critérios selecionados que cada método avalia e a quantidade total de critérios selecionados.

Tabela 5: Comentários fornecidos pelos participantes do estudo de caso.

“Seria interessante fornecer descrições para deixar mais claro como utilizar o guia. Sugiro um manual de uso que fala desde quais programas usar até como escolher os critérios.” - P3, P8, P9, P12, P18, P20, P24, P26, P46, P49
“Transformar o Guia em um aplicativo, explicando cada método (sem precisar procurar pelo artigo), e que já dê ferramentas (como perguntas e checklist).” - P4, P6, P7
“Tirar do excel e fazer um aplicativo para melhorar a usabilidade e desempenho” – P44, P45, P47, P64
“O guia serviu muito bem ao que se propôs.” - P2, P21, P23
“O guia é muito fácil e claro de utilizar” - P8, P13
“Utilizar os filtros do guia são difíceis, como por exemplo: no entendimento do método de avaliação, como funciona os filtros, a metodologia é bastante complexa” - P6, P10, P11, P15, P21
“No início foi um pouco difícil de saber como selecionar os critérios.” - P3, P9, P24
“Em relação ao preenchimento e visualização dos resultados fica um pouco confuso” – P57
“A limitação é que o guia está no excel, poucas pessoas possuem o domínio dessa ferramenta” - P7, P26

Sobre as experiências com o uso do guia, alguns participantes afirmaram que tiveram dificuldade na utilização do guia e que perceberam que o guia não mostrou ao avaliador o melhor método a se utilizar. Acredita-se que os critérios definidos no cenário podem ter influenciado para isso, pois como os critérios foram definidos de forma aleatória nos cenários, não se observou previamente se haveria um método que cobrisse a maioria dos critérios. Assim, algumas equipes não obtiveram um método de avaliação com cobertura razoável para os critérios definidos no cenário recebido. Entretanto, outros participantes indicaram que o guia cumpriu o que se propôs, que o guia é claro, eficiente e fácil de ser utilizado.

Os participantes apontaram que muitos participantes não dominam a ferramenta excel, o que causou adversidades no uso do guia e na interação com os mecanismos de filtro. É possível deduzir que os participantes que tiveram uma experiência positiva, tinham familiaridade e um bom conhecimento sobre o Excel. Desta forma, reforça-se a ideia de que transformar o guia em

um aplicativo pode promover uma experiência com maior facilidade de uso e desempenho para os avaliadores.

## 6 Discussão

Os resultados obtidos nesta pesquisa demonstram a viabilidade de criação de um guia para apoiar profissionais interessados em avaliar a qualidade de tecnologias *M-learning*. A pesquisa possibilitou a criação de um guia com uma base inicial composta por métodos que permitem avaliar tecnologias *M-learning*. O guia facilita o processo de seleção de métodos adequados, pois sugere um ou mais métodos de avaliação com base em critérios de avaliação selecionados pelo avaliador.

A avaliação do guia permitiu obter dados sobre sua utilidade, facilidade de uso e potencial adoção futura. Embora a avaliação tenha sido conduzida em ambiente acadêmico com estudantes de graduação, considera-se que a percepção dos estudantes sobre o guia permitiu analisar sua viabilidade e identificar necessidades de melhoria. Ressalta-se que durante a avaliação, os estudantes exerceram o papel de avaliadores interessados em selecionar um método adequado para avaliação, simulando o papel de um profissional da área. Os dados obtidos na avaliação demonstram que o guia foi percebido como útil para avaliação de tecnologias *M-learning* pela maioria dos participantes. Embora a maioria dos participantes tenha experimentado uma boa interação com o guia, o uso de filtros na planilha ocasionou algumas dificuldades de uso. Ainda assim, mais de 90% dos participantes indicaram intenção em adotar o guia futuramente.

Os resultados indicam o potencial do guia em apoiar profissionais na avaliação da qualidade e seleção de tecnologias *M-learning* adequadas à sua prática de ensino. Assegurar que uma tecnologia *M-learning* possui boa qualidade e promoverá uma boa experiência de uso aos estudantes pode potencializar os benefícios de sua adoção no processo de ensino e aprendizado. O guia torna informações sobre métodos de avaliação de qualidade mais acessível aos educadores que podem ter pouco ou nenhum conhecimento sobre o que e como avaliar a qualidade de tecnologias *M-learning*.

Esta pesquisa adotou a base de critérios de qualidade definida por Soad, Duarte Filho e Barbosa (2015) para tecnologias *M-learning*. Neste artigo, os critérios de qualidade são adotados para definir o que um avaliador pretende avaliar em uma tecnologia *M-learning* e sugerir métodos que contemplem os critérios de interesse. Desta forma, os resultados da pesquisa de Soad, Duarte Filho e Barbosa (2015) são empregados na proposta de uma nova solução. Tendo em vista que um software *M-learning* não precisa cobrir todo o conjunto de características de qualidade para esse tipo de aplicação, o guia colabora com a adoção de métodos existentes na literatura, tais como Nunes & Santos (2018), Da Silva & Batista (2015) e Petri, von Wangenheim & Borgatto (2016). Cada método avalia um subconjunto de critérios que podem ser de interesse do avaliador, o guia então sugere a adoção dos métodos existentes, contribuindo para a qualidade de tecnologias *M-learning*.

Atualmente, o guia pode ser adotado e possui uma versão funcional em planilha, porém com uma base estática de métodos para avaliação. Esta base pode ser atualizada por meio de uma extensão da revisão da literatura conduzida no âmbito desta pesquisa, que considerou inicialmente critérios de qualidade relacionados à usabilidade. Vale ressaltar que apesar da adoção de termos relacionados à usabilidade, a busca permitiu a identificação de métodos de avaliação que consideram características pedagógicas, socioeconômicas e socioculturais. O mapeamento sistemático conduzido por Cáceres & Pow-Sang (2018) sobre métodos de avaliação de usabilidade para *M-learning* pode ser considerado para enriquecer a base de métodos de avaliação do guia.

Para isso, os critérios de qualidade considerados nos métodos identificados por Cáceres & Pow-Sang (2018) devem ser identificados e relacionados às características de qualidade para tecnologias *M-learning*. Assim é possível realizar o mapeamento entre os novos métodos de avaliação e os critérios de qualidade considerados no guia.

Para viabilizar a adoção e manutenção eficaz do guia pela comunidade, uma aplicação está sendo desenvolvida para permitir a atualização constante da base de métodos para avaliação e fornecer uma interação com maior usabilidade aos usuários na seleção de critérios de avaliação e reconhecimento dos métodos sugeridos pelo guia. Pretende-se avaliar a aplicação com profissionais da área de educação para consolidar os dados sobre a utilidade e a facilidade de uso do guia, dado que este configura o público-alvo do guia proposto.

Esta pesquisa configura um esforço inicial no sentido de contribuir para que profissionais da Educação possam considerar critérios de qualidade na seleção de tecnologias *M-learning* e profissionais da área de Computação que desenvolvem tecnologias *M-learning* possam avaliá-las e melhorar sua qualidade com base em critérios estabelecidos. Além disso, esta pesquisa visa encorajar a avaliação de tecnologias *M-learning* para obter mais evidências sobre sua qualidade e experiências de uso promovidas, contribuindo assim para a evolução da área, consolidação de tecnologias e construção de uma base de conhecimento acessível a toda a comunidade.

## 7 Considerações Finais e Trabalhos Futuros

Esta pesquisa objetivou relatar o processo de criação e avaliação de um guia proposto para pessoas interessadas na avaliação de tecnologias *M-learning* possam selecionar o método mais adequado conforme os critérios de qualidade de seu interesse. A elaboração da pesquisa ocorreu a partir de uma revisão da literatura para identificação dos métodos existentes, das características, critérios e subcritérios de qualidade que cada método avalia e da definição de critérios e subcritérios de qualidade que uma aplicação *M-learning* deve apresentar. Em seguida, foi possível criar e avaliar o guia de acordo com a experiência de uso de cada avaliador.

Visto que os métodos não precisam atender a todos os critérios e subcritérios de qualidade existentes, o guia promove a escolha de um método existente na literatura que mais se adequa aos critérios de qualidade de interesse do avaliador. A avaliação do guia por meio de um estudo de caso com estudantes de Qualidade de Software indicou que o guia foi percebido como útil e fácil de usar pela maioria dos estudantes. Assim como houve indicação de intenção de usar o guia futuramente em outras atividades de avaliação de tecnologias *M-learning*. Contudo, o fato de o guia estar em formato de planilha, ocasionou dificuldades no seu uso e manipulação, assim como no entendimento dos resultados fornecidos.

Neste sentido, como trabalho futuro, pretende-se transformar o guia em uma aplicação web ou móvel que permita uma melhor interação com o usuário, possibilitando maior usabilidade, desempenho, acessibilidade e suporte ao usuário. Além disso, sugere-se realizar um mapeamento sistemático que cubra outras características de qualidade e métodos de avaliação para tecnologias de software *M-learning*, enriquecendo a base inicial de métodos do guia, que requer uma constante atualização.

## Referências

Andrade, M. V. M., Araújo Jr, C. F., & Silveira, I. F. (2015). Critérios de qualidade para aplicativos educacionais no contexto dos dispositivos móveis (M-Learning). In *Nuevas Ideas*

- em Informática Educativa. Congresso Internacional sobre Informática na Educação (TISE)*. [[GS Search](#)]
- Ataíde, A. P. R., Teixeira, A. B. M., & da Silva, C. I. P. (2003). Maqsei-uma metodologia de avaliação de qualidade de software educacional infantil. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, (Vol. 1, No. 1, pp. 356-365). DOI: [10.5753/cbie.sbie.2003.356-365](#) [[GS Search](#)]
- Baloh, M., Zupanc, K., Košir, D., Bosnić, Z., & Scepanovic, S. (2015). A quality evaluation framework for mobile learning applications. In *4th Mediterranean Conference on Embedded Computing (MECO)* (pp. 280-283). IEEE. DOI: [10.1109/MECO.2015.7181923](#) [[GS Search](#)]
- Barbosa, G., Oliveira, E., & D'Carlo, D. (2016). Usabilidade em aplicativos móveis educacionais: Um conjunto de heurísticas para avaliação. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 27, No. 1, p. 777). DOI: [10.5753/cbie.sbie.2016.777](#) [[GS Search](#)]
- Bednarik, R., Gerdt, P., Miraftabi, R., & Tukiainen, M. (2004). Development of the TUP model- Evaluating educational software. In *IEEE International Conference on Advanced Learning Technologies* (pp. 699-701). IEEE. DOI: [10.1109/ICALT.2004.1357627](#) [[GS Search](#)]
- Cáceres, S. V. & Pow-Sang, J. A. (2018). A systematic mapping review of usability evaluation methods for educational applications on mobile devices. In *7th International Conference On Software Process Improvement (CIMPS)*, Guadalajara, Jalisco, Mexico, pp. 59-68. IEEE. DOI: [10.1109/CIMPS.2018.8625629](#) [[GS Search](#)]
- CAMPOS, G. H. B. D. (1989). *Construção e validação de ficha de avaliação de produtos educacionais para microcomputadores*. Rio de Janeiro. [[GS Search](#)]
- Coomans, S., & Lacerda, G. S. (2015). Petese, a pedagogical ergonomic tool for educational software evaluation. *Procedia Manufacturing*, 3, 5881-5888. DOI: [10.1016/j.promfg.2015.07.895](#) [[GS Search](#)]
- Cota, C. X. N., Díaz, A. I. M., & Duque, M. Á. R. (2014). Developing a framework to evaluate usability in m-learning systems: Mapping study and proposal. In *Second international conference on technological ecosystems for enhancing multiculturalty* (pp. 357-364). DOI: [10.1145/2669711.2669924](#) [[GS Search](#)]
- Da Silva, M. G., & Batista, S. C. F. (2015). Metodologia de avaliação: análise da qualidade de aplicativos educacionais para matemática do ensino médio. *RENOTE-Revista Novas Tecnologias na Educação*, 13(1). DOI: [10.22456/1679-1916.57641](#) [[GS Search](#)]
- Da Silva, R. S., Silva, W. R., Cardoso Filho, R., Pereira, W., Aguiar, Y., & Dantas, V. (2016). Avaliação de Software Educativo: a complexidade de escolher uma abordagem adequada. In *Congresso sobre Tecnologias na Educação (Ctrl+E 2016)*, (pp. 116-126). [[GS Search](#)]
- Gomes, A. S., & Wanderley, E. G. (2003). Elicitando requisitos em projetos de Software Educativo. In *Anais do Workshop de Informática na Escola* (Vol. 1, No. 1, pp. 119-130). DOI: [10.5753/cbie.wie.2003.119-130](#) [[GS Search](#)]
- Gomes, A. S., Castro Filho, J. A., Gitirana, V., Spinillo, A., Alves, M., Melo, M., & Ximenes, J. (2002). Avaliação de software educativo para o ensino de matemática. In *WIE 2002 Workshop Brasileiro de Informática Educativa*. Florianópolis: SBC. [[GS Search](#)]
- ISO/IEC 25010 (2011). ISO/IEC 25010:2011, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuARE) — System and software quality models. [[GS Search](#)]

- Kitchenham, B. & Charters, S. (2007) Guidelines for performing Systematic Literature Reviews in Software Engineering, Version 2.3, EBSE Technical Report EBSE-2007-01, Keele University and University of Durham. [[GS Search](#)]
- Nunes, R., & Santos, I. (2018). A importância da avaliação ergonômico-pedagógica de aplicativos educacionais e os desafios encontrados no campo da aprendizagem móvel. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 29, No. 1, p. 41). DOI: [10.5753/cbie.sbie.2018.41](#) [[GS Search](#)]
- Pereira, W. S., Cardoso Filho, R. J., Silva, W. R. D. A., Da Silva, R. S. T., Dantas, V. F., & Aguiar, Y. P. (2016). Validação de uma abordagem combinada para avaliação de software educativo: Avanços e Desafios. *Revista Tecnologias na Educação*, 16(8). [[GS Search](#)]
- Petri, G., & von Wangenheim, C. G. (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](#) [[GS Search](#)]
- Petri, G., von Wangenheim, C. G., & Borgatto, A. F. (2016). MEEGA+: an evolution of a model for the evaluation of educational games. *INCoD/GQS*, 3. [[GS Search](#)]
- Petri, G., von Wangenheim, C. G., & Borgatto, A. F. (2020). MEEGA+: Um Modelo para a Avaliação de Jogos Educacionais para o ensino de Computação. *Revista Brasileira de Informática na Educação*, 27(3), 52. DOI: [10.5753/rbie.2019.27.03.52](#) [[GS Search](#)]
- Porto, J. V., Barbosa, H., & von Wangenheim, C. G. (2018). Proposta de um Checklist de Avaliação de Usabilidade de Aplicativos Android no Contexto Educacional. In *Computer on the Beach*, (pp. 870-879). [[GS Search](#)]
- Runeson, P., Host, M., Rainer, A., & Regnell, B. (2012). Case study research in software engineering: Guidelines and examples. John Wiley & Sons. [[GS Search](#)]
- Soad, G. W., & Barbosa, E. F. (2017). MoLEva: um método de avaliação de qualidade para aplicativos educacionais móveis. In *Simpósio Brasileiro de Qualidade de Software*, (pp. 91-105). [[GS Search](#)]
- Soad, G. W., Duarte Filho, N. F., & Barbosa, E. F. (2015). Uma contribuição ao estabelecimento de características de qualidade para aplicações educacionais móveis. In *Simpósio Brasileiro de Qualidade de Software*, (pp. 165–179). [[GS Search](#)]
- Soad, G. W., Duarte Filho, N. F., & Barbosa, E. F. (2016). Quality evaluation of mobile learning applications. In *2016 IEEE Frontiers in Education Conference (FIE)* (pp. 1-8). IEEE. DOI: [10.1109/FIE.2016.7757540](#) [[GS Search](#)]
- Vargo, J., Nesbit, J. C., Belfer, K., & Archambault, A. (2003). Learning object evaluation: computer-mediated collaboration and inter-rater reliability. *International Journal of Computers and Applications*, 25(3), 198-205. DOI: [10.1080/1206212X.2003.11441703](#) [[GS Search](#)]
- Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2), 273-315. DOI: [10.1111/j.1540-5915.2008.00192.x](#) [[GS Search](#)]
- Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010, October). User experience evaluation methods: current state and development needs. In *6th Nordic conference on human-computer interaction: Extending boundaries* (pp. 521-530). DOI: [10.1145/1868914.1868973](#) [[GS Search](#)]

Vieira, M. M. S., Simões, L. L. F., & de Oliveira Barreto A. L. (s.d) Avaliação de software educativo: aspectos pedagógicos e técnicos. [[GS Search](#)]

Zaibon, S. B., & Shiratuddin, N. (2010). Heuristics evaluation strategy for mobile game-based learning. In *2010 6th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education* (pp. 127-131). IEEE. DOI: [10.1109/WMUTE.2010.27](https://doi.org/10.1109/WMUTE.2010.27) [[GS Search](#)]