

MoLEva: Um Método de Avaliação de Qualidade para Aplicativos Educacionais Móveis

Gustavo W. Soad¹, Ellen F. Barbosa¹

¹Instituto de Ciências Matemáticas e Computação – Universidade de São Paulo (USP) – São Carlos – SP – Brasil

gustavo.soad@usp.br, francine@icmc.usp.br

Abstract. *The use of learning applications has been growing continuously, allowing students and teachers greater flexibility and convenience in the execution of educational activities and practices. Despite the benefits, mobile learning has problems and challenges, for example, to adequately evaluate the quality of the educational applications developed. In this context, the present work proposes a method of quality evaluation for mobile learning applications. In addition, a case study was also conducted. From the results obtained, it was possible to identify points of improvement in the applications, besides providing a score that can be used as comparative for the choice of this type of application.*

Resumo. *A utilização de aplicativos educacionais móveis vem crescendo continuamente, possibilitando a alunos e professores maior flexibilidade e comodidade na execução de atividades e práticas educacionais. Apesar dos benefícios, a aprendizagem móvel possui problemas e desafios, por exemplo, avaliar adequadamente a qualidade dos aplicativos educacionais desenvolvidos. Nesse contexto, o presente trabalho propõe um método de avaliação de qualidade para aplicativos educacionais móveis. Além disso, também foi conduzido um estudo de caso. A partir dos resultados obtidos, foi possível identificar pontos de melhoria nos aplicativos, além de fornecer uma pontuação que pode ser utilizada como comparativo para a escolha desse tipo de aplicativo.*

1. Introdução

Nos últimos anos, temas relacionados ao ensino e aprendizagem têm sido amplamente discutidos e estudados pela comunidade científica. Em especial, aplicações computacionais têm apresentado uma crescente importância, desempenhando um papel fundamental em atividades de ensino e aprendizagem, sendo relevantes não apenas no âmbito acadêmico como também no meio industrial [Barbosa 2004, Rubens et al. 2014].

Tais aplicações proporcionaram, entre várias práticas educacionais inovadoras, uma nova modalidade de ensino - a aprendizagem móvel ou *mobile learning (m-learning)* [Nah et al. 2008, Wexler et al. 2008, Kearney et al. 2012]. Em linhas gerais, o *m-learning* caracteriza-se pela capacidade de proporcionar uma forte interação entre os aprendizes e instrutores, possibilitando aos mesmos contribuir, participar e acessar o ambiente de ensino por meio de dispositivos móveis (celulares, tablets, laptops, rádio, tv, entre outros) a qualquer momento e em qualquer lugar.

Apesar dos benefícios proporcionados no contexto de ensino e aprendizagem, por se tratar de um conceito novo e ainda incipiente, o *m-learning* apresenta problemas e desafios organizacionais, culturais e tecnológicos em sua utilização [Economides 2008, Sarrab et al. 2013, Sharples 2013]. Um destes problemas é que os métodos existentes para avaliação da qualidade de software ainda são muito genéricos, não contemplando aspectos específicos dos aplicativos educacionais móveis.

De fato, aspectos de qualidade, em particular, representam uma questão importante a ser abordada, principalmente em virtude da crescente popularidade dos aplicativos móveis em diferentes setores da sociedade. Neste cenário emergente, a qualidade não é apenas relacionada a aspectos técnicos. Há também a necessidade de lidar com questões intrínsecas (por exemplo, educacionais, socioculturais e socioeconômicas) relacionadas com as atividades diárias de aprendizes, professores e tutores [Economides 2008, Kearney et al. 2012, Abdurrahman et al. 2015].

Apesar de sua relevância, poucos estudos sobre as diretrizes de qualidade para aplicativos móveis são encontrados na literatura. Como consequência, torna-se difícil identificar mecanismos bem definidos e amplamente utilizados para apoiar a avaliação da qualidade de aplicativos educacionais móveis.

Dentro do contexto apresentado, o presente trabalho visa a proposição de um método para avaliar a qualidade de aplicativos educacionais móveis denominado MoLEva (*Mobile Learning Evaluation*).

O artigo encontra-se organizado como se segue. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 descreve o método MoLEva e a suas etapas de desenvolvimento. A Seção 4 traz a avaliação do método e apresenta a análise dos resultados obtidos. Por fim, na Seção 5, são sumarizadas as conclusões do trabalho e as perspectivas para continuidade da pesquisa.

2. Trabalhos Relacionados

Para definir o nível de qualidade de um produto de software é necessária uma avaliação de qualidade, que pode ser conduzida sob os aspectos do processo de desenvolvimento, que se refere às atividades realizadas durante as etapas do seu ciclo de vida, e sob a qualidade de produto, que se refere às características do produto após o seu desenvolvimento.

A base para o desenvolvimento de um método de avaliação é o seu modelo de qualidade. Um modelo de qualidade corresponde a um conjunto de características que se relacionam entre si e que fornecem a base para a especificação de requisitos de qualidade e avaliação de qualidade [ISO/IEC 9126 2001].

Visando estabelecer formas de garantir a qualidade de um produto de software, diversos modelos e padrões vêm sendo estabelecidos. Nestes modelos, a especificação e avaliação da qualidade do produto de software são fatores fundamentais para a garantia da qualidade adequada.

Existem dois modelos de qualidade que são considerados seminais: o modelo de McCall [Cavano e McCall 1978] e o modelo de Boehm [Boehm et al. 1976]. Além dos modelos citados, a norma ISO/IEC 25010 [ISO/IEC 25010 2010] também possui um papel importante. Esta norma define um modelo de qualidade composto por

características internas (características do produto em seu projeto e construção) e externas (características do produto quando é avaliado em um ambiente simulado e com dados simulados) relacionadas ao produto de software. As características definidas, por sua vez, são subdivididas em subcaracterísticas.

Além desses modelos, também existem modelos específicos para o contexto educacional, por exemplo, o TUP (*Technology Usability Pedagogy*) [Bednarik et al. 2004], que propõe um modelo que permite a avaliação de softwares educacionais. Outro modelo proposto também é definido em Duarte Filho e Barbosa (2013a), e seu objetivo é avaliar ambientes educacionais móveis.

Os trabalhos citados anteriormente contribuíram para o desenvolvimento do método MoLEva. Além disso, as características definidas pela norma ISO/IEC 25010 foram fundamentais para a proposição do método MoLEva, já que esta norma define características de qualidade gerais para os produtos de software. Dessa maneira, é possível utilizar as características definidas incorporando-as ao modelo MoLEva, além de auxiliar no desenvolvimento da hierarquia do modelo de qualidade MoLEva.

3. MoLEva: Método de Avaliação de Qualidade

Com o objetivo de desenvolver o método de avaliação de qualidade para aplicativos educacionais móveis, foi necessário realizar estudos sobre as características de qualidade presentes nesse tipo de aplicativo. Para isso, foi conduzido um mapeamento sistemático a fim de obter um conjunto de características de qualidade para aplicativos educacionais móveis [Soad et al. 2015].

A partir dos resultados obtidos, foi definido um conjunto de características de qualidade para o contexto de aplicativos educacionais móveis, estabelecendo assim a base para a criação do modelo de qualidade para o método MoLEva. Outras pesquisas como Duarte Filho e Barbosa (2013a) e Acharya e Sinha (2013) também foram utilizadas para auxiliar na definição das etapas do processo de avaliação utilizado pelo método MoLEva.

Em linhas gerais, o processo de desenvolvimento do método MoLEva é composto por seis etapas, em conformidade com a ISO/IEC 14598 [ISO/IEC 14598 1998], as quais foram simplificadas e adaptadas para o contexto de aplicativos educacionais móveis. As etapas são:

- **Definir objetivos e contextualização da avaliação:** o objetivo definido para o método MoLEva é analisar e comparar os aplicativos educacionais móveis em relação aos aspectos técnicos, educacionais, socioeconômicos e socioculturais, identificando pontos fortes e fracos de cada aplicativo considerado.

- **Definir e configurar o modelo de qualidade:** o modelo de qualidade a ser adotado no processo estabelece um conjunto bem definido de critérios e requisitos para aplicativos educacionais móveis. É essencial que o modelo definido nessa etapa seja configurado de acordo com o contexto de aplicativos educacionais, pois dessa maneira, é possível fornecer uma base de características que esse tipo de aplicativo deve satisfazer para conseguir excelência em seu nível de qualidade.

- **Identificar métricas de avaliação:** nesta etapa, os requisitos de qualidade são mapeados em atributos que podem ser medidos. Em suma, os atributos mensuráveis são confi-

gurados com base no conhecimento de especialistas, por meio da criação de perguntas simples e diretas, que são avaliadas e pontuadas por meio de um checklist.

- **Definir níveis de pontuação e critérios de julgamento:** nesta etapa, são definidos os níveis de pontuação para as perguntas do checklist. Para isso, foi necessário primeiramente especificar os tipos de respostas nominais que seriam aplicadas às questões. Além disso, nessa etapa também são definidos os critérios de julgamento, que por sua vez definirão qual o nível de qualidade do aplicativo.

- **Projetar a avaliação:** nessa etapa deve ser definido o plano de avaliação que deverá ser seguido durante a execução da avaliação. Esse plano deve conter as diretrizes e procedimentos necessários para que as avaliações sejam feitas, além de definir a técnica de coleta de dados. No caso do MoLEva a técnica utilizada para avaliação é o checklist. Para facilitar na definição do plano de avaliação alguns pontos são sugeridos por Duarte Filho e Barbosa (2013a), como: (i) identificar o perfil dos avaliadores; (ii) mapear os recursos necessários para viabilizar a avaliação; (iii) tempo médio estipulado para a execução da avaliação; e (iv) como serão apresentados os resultados gerados pela avaliação.

- **Executar a avaliação:** na fase final, o avaliador deve executar três tarefas principais: (i) coletar as medidas de qualidade; (ii) comparar as medidas em relação aos critérios de qualidade pré-definidos; e (iii) avaliar os dados obtidos durante a avaliação. As etapas citadas devem ser executadas por desenvolvedores e especialistas em qualidade de software.

As etapas do processo citadas anteriormente foram executadas algumas vezes, com o objetivo de refinar o método MoLEva. Resultados sobre os primeiros testes com o método MoLEva foram publicados em Soad et al. (2016). Dessa forma, foi possível estabelecer o método de avaliação de qualidade, que consiste em três etapas, que são: (i) configuração da avaliação, que possibilita a configuração dos critérios de qualidade que devem ser considerados, além de possibilitar a customização dos critérios de julgamento; (ii) execução da avaliação, que é a etapa, na qual o checklist é aplicado; e (iii) análise dos resultados da avaliação, que apresenta os resultados obtidos com a execução da avaliação.

Para definir os principais componentes do método MoLEva, utilizou-se como base a norma ISO/IEC 14598 [ISO/IEC 14598 1998], estabelecendo assim um método formado por três componentes fundamentais:

- Modelo de qualidade: define um conjunto de características de qualidades para aplicativos educacionais móveis, fornecendo assim a base para a avaliação de qualidade;
- Métricas: definem a forma de coleta de dados para possibilitar a medição da qualidade dos aplicativos;
- Critérios de julgamento: definem como o aplicativo será julgado a partir dos dados coletados durante a avaliação.

3.1. Modelo de Qualidade

O principal componente do MoLEva é seu modelo de qualidade, responsável por definir o conjunto de características que fornece a base para a avaliação de qualidade de aplica-

tivos educacionais móveis. O modelo de qualidade do método MoLEva é dividido em três categorias, fundamentadas na divisão dos requisitos de qualidade para aplicativos *m-learning* definidos em Soad et al. (2015) e Duarte Filho e Barbosa (2013b). As categorias são:

- Pedagógica: são os aspectos educacionais presentes no aplicativo, ou seja, são as características que têm como foco o aprendizado do aluno;
- Social: referem-se aos aspectos sociais presentes no aplicativo, abordando características econômicas e culturais;
- Técnica: são os aspectos técnicos do aplicativo, ou seja, referem-se às características da aplicativo que contribuem para seu uso e funcionamento.

O modelo de qualidade possui uma estrutura composta por categorias, critérios e características de qualidade, dispostos de maneira hierárquica. Essa divisão foi fundamentada na norma ISO/IEC 25010 [ISO/IEC 25010 2010]. A Figura 1 apresenta uma visão geral do modelo de qualidade que compõe o MoLEva.

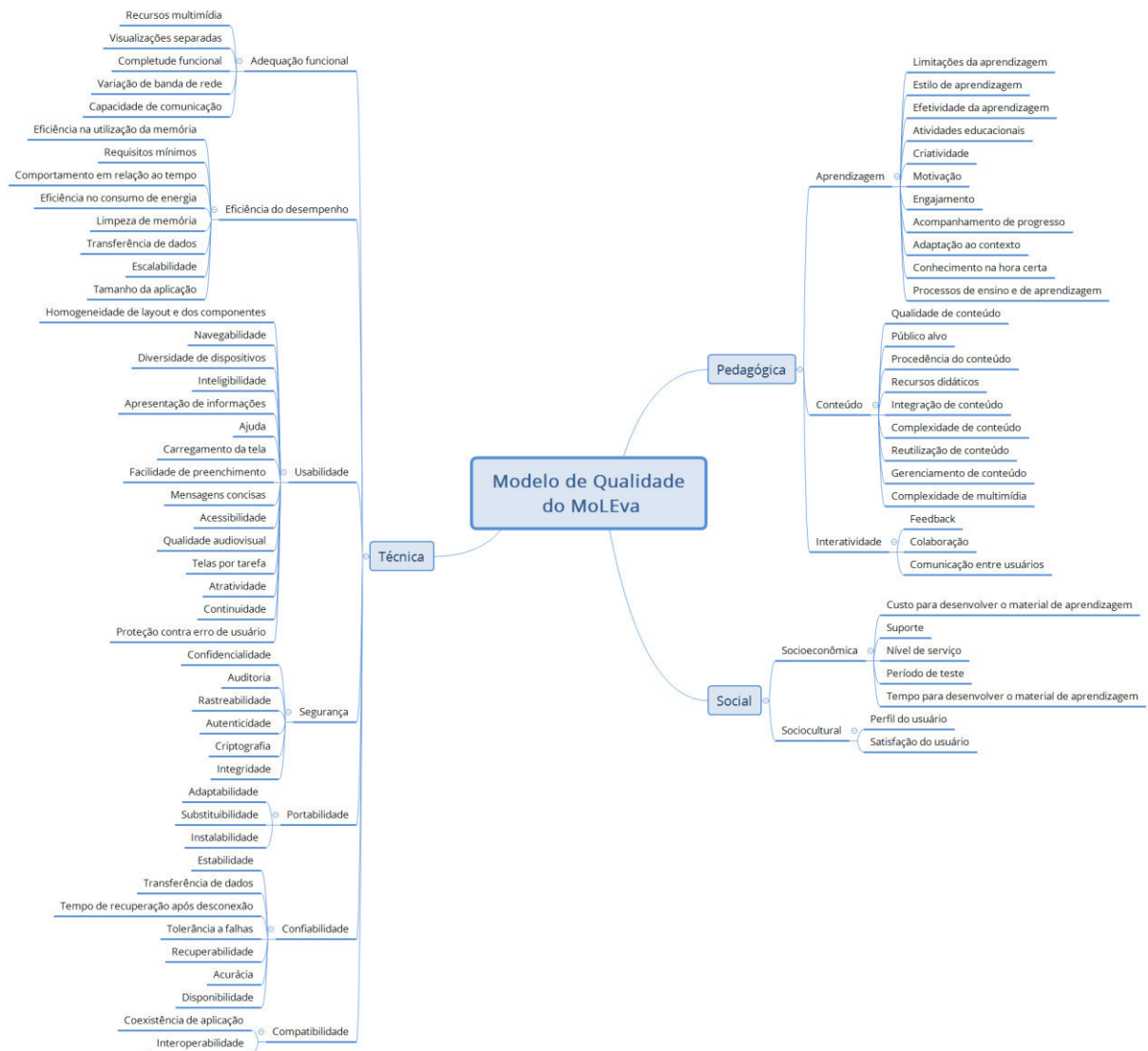


Figura 1. Modelo de Qualidade do MoLEva

As categorias definidas no modelo possuem uma subdivisão, totalizando 12 critérios de qualidade. Para a categoria Pedagógica, são definidos três critérios de qualidade. O primeiro é o critério **Aprendizagem** que é definido pela capacidade do aplicativo em disponibilizar funcionalidades que contribuam para o aprendizado do aluno. O segundo é **Conteúdo**, que é definido pela capacidade em fornecer conteúdo gerenciável e de qualidade. O terceiro (e último) critério pedagógico existente no modelo é a **Interatividade** que é definida como, a capacidade do aplicativo em disponibilizar funcionalidades que contribuam para que os usuários interajam entre si e com o aplicativo. Algumas das características relacionadas a categoria Pedagógica são apresentadas na Tabela 1.

Tabela 1. Características de Qualidade Relacionadas à Categoria Pedagógica

Característica	Descrição
Limitações na aprendizagem	Capacidade de auxiliar na identificação de lacunas de conhecimento.
Conhecimento na hora certa	Conteúdo oferecido no estágio correto de aprendizagem, evitando oferecer um conteúdo muito avançado que dependa de conhecimento ainda não obtido.
Adaptação ao contexto	Capacidade de adaptação ao contexto do usuário considerando idade, gênero, escolaridade, condições físicas do ambiente, entre outros.
Procedência do conteúdo	Todo material disponibilizado deve ser de fonte confiável.
Colaboração	Colaboração nas atividades, permitindo que os aprendizes, tutores e professores se relacionem por meio de wiki, jogos, microblogs, fóruns, grupos, entre outros.

A próxima categoria estabelecida no modelo de qualidade do MoLEva é a categoria Social. Para essa categoria são definidos dois critérios de qualidade. O primeiro é o critério **Socioeconômico**, que é caracterizado pela capacidade em proporcionar viabilidade na utilização do aplicativo, fazendo uma relação de custo e benefício. O segundo critério relacionado à categoria Social é o **Sociocultural**. Ele é caracterizado pela capacidade do aplicativo em se adaptar ao contexto sociocultural do usuário, considerando características como idade, nível de ensino, determinado grupo cultural, entre outros. Algumas características relacionadas à categoria Social são apresentadas na Tabela 2.

Tabela 2. Características de Qualidade Relacionadas à Categoria Social

Característica	Descrição
Custo para desenvolver o material de aprendizagem	Diminuição dos custos de criação do material de aprendizagem, pois estes influenciam diretamente os gastos com o aplicativo e, conseqüentemente, no investimento feito pelo usuário.
Suporte	Suporte aos usuários em relação aos serviços oferecidos pelo aplicativo.
Perfil do usuário	Configurabilidade do aplicativo para tratar diferentes possíveis limitações de acordo com a idade, sexo, religião, raça, entre outras características dos usuários.

A última categoria a ser apresentada é a Técnica, que é composta por sete critérios de qualidade. O primeiro critério apresentado é a **Adequação Funcional**, que é definido como a capacidade do aplicativo em disponibilizar funções que atendam às necessidades implícitas e explícitas em relação ao aplicativo. O segundo critério é a **Eficiência do Desempenho**. Ele é caracterizado pela capacidade do aplicativo em fornecer um bom desempenho em relação à quantidade de recursos utilizados sob condições estabelecidas, ou seja, refere-se à otimização do uso dos recursos disponíveis.

O terceiro critério é definido como **Compatibilidade**. Ele se caracteriza como a capacidade do aplicativo em trocar informações ou realizar operações com outros aplicativos, compartilhando o mesmo ambiente de hardware e software. O quarto critério é a **Usabilidade**. Esse critério é definido como a capacidade do aplicativo em ser utilizado

por usuários específicos em um contexto específico, oferecendo maneiras que possibilitam o aplicativo ser entendido, aprendido, utilizado e atraente ao usuário.

O quinto critério é definido como **Confiabilidade**. Ele é caracterizado pela capacidade do aplicativo em proporcionar um comportamento consistente com o esperado durante um longo período de tempo. O sexto critério é definido como **Segurança**. Esse critério é caracterizado pela capacidade do aplicativo em proteger as informações e dados, protegendo-os de acesso não autorizado e garantindo seu acesso de acordo com os diferentes níveis de autorização.

Por fim, o último critério de qualidade é definido como **Portabilidade**. Este critério é caracterizado pela capacidade do aplicativo ser transferido de um hardware ou software para outro. Algumas das características relacionadas à categoria Técnica são apresentadas na Tabela 3.

Tabela 3. Características de Qualidade Relacionadas à Categoria Técnica

Característica	Descrição
Recursos multimídia	Capacidade de suportar diferentes tipos de mídias, como vídeo, texto, áudio, entre outros.
Comportamento em relação ao tempo	Tempo de espera do usuário até o carregamento completo das informações solicitadas.
Coexistência de aplicativo	Capacidade de vários aplicativos serem executados no mesmo dispositivo sem ocasionar impactos negativos de umas nas outras.
Facilidade de preenchimento	Capacidade de facilitar o preenchimento de dados pelo usuário. A quantidade de campos para entrada de dados no aplicativo é considerada e, já que existem limitações de tela e teclado, não se deve exigir a entrada de muitos dados.
Disponibilidade	Disponibilidade e acessibilidade a qualquer hora e em qualquer lugar.
Criptografia	Utilização de criptografia nos dados sensíveis que são trafegados e armazenados pelo aplicativo (exemplo: senhas ou documentos do usuário).
Adaptabilidade	Capacidade do aplicativo em se adaptar a diferentes tipos de dispositivos móveis.

3.2. Métricas

Para determinar o valor associado às características de qualidade definidas pelo método MoLEva, é necessária uma medição. Dessa maneira, é possível medir e interpretar a qualidade do critério que está sendo avaliado.

O termo “métrica” utilizado não possui o sentido matemático usual, mas se refere à uma escala e um método que pode ser utilizado para medição. Essa definição é fundamentada nas diretrizes da ISO/IEC 14598 [ISO/IEC 14598 1998], na fase de especificação da avaliação.

As métricas que compõem o método de qualidade MoLEva foram definidas mapeando as características de qualidade em atributos mensuráveis, permitindo assim que elas fossem medidas e pontuadas. Para isso, optou-se em utilizar a técnica de checklist. A aplicação dessa técnica consiste em um conjunto de perguntas objetivas, que permitem a coleta de informações por meio de respostas simples e objetivas.

O desenvolvimento do checklist do MoLEva utilizou como base o proposto em Duarte Filho (2016). As perguntas contidas nesse checklist foram adaptadas para o contexto de aplicativos educacionais móveis. Além disso, o modelo de qualidade descrito anteriormente e ilustrado na Figura 1, foi essencial para a criação de novas perguntas, já que ele é a base para o método de avaliação MoLEva.

Ao final do desenvolvimento, o checklist (<https://goo.gl/xHbaEn>) foi composto por 85 perguntas, abordando os 12 critérios (sub-categorias) de qualidade definidos no

modelo da Figura 1. Esse conjunto de perguntas é fundamental para a avaliação de qualidade, já que a partir dele são coletadas as informações necessárias para definição do nível de qualidade do aplicativo.

3.3. Critérios de Julgamento

A definição de critérios de julgamento permite determinar o nível de qualidade do aplicativo. Por meio dessa definição, é possível fazer comparações entre os aplicativos, identificando-se, também, possibilidades de melhorias.

Para definir o nível de qualidade do aplicativo a ser avaliado e identificar os possíveis pontos de atenção, é necessário atribuir valores quantitativos para as perguntas contidas no checklist. Por esse motivo, as respostas obtidas na aplicação da avaliação de qualidade devem ser convertidas em dados quantitativos. Para as respostas binárias, os valores atribuídos são 10 pontos para respostas “Sim” e 0 pontos para respostas “Não”.

Com relação às respostas por grau de concordância, a pontuação atribuída é mostrada na Tabela 4. Para cada grau de concordância, foi atribuído um significado possível, objetivando traduzir o dado quantitativo em qualitativo, facilitando assim o significado numérico da resposta. Os valores atribuídos para as respostas foram baseados nos critérios de pontuação definidos em Duarte Filho e Barbosa (2013a).

Tabela 4. Pontuações Definidas para as Respostas por Grau de Concordância

Grau de concordância	Significado da resposta	Pontuação
5	Eficiente	10
4	Aceitável	7.5
3	Regular	5
2	Deficiente	2.5
1	Ineficiente	0

As pontuações são atribuídas a cada pergunta do checklist. Essas perguntas contemplam os 12 critérios de qualidade definidos no método MoLEva. O cálculo final do nível de qualidade é definido pela média aritmética da pontuação de cada critério.

Para efeito de normalização dos resultados, os critérios de qualidade recebem a porcentagem de pontos equivalentes à pontuação recebida. Por exemplo, para um critério que possua a pontuação máxima igual a 80 pontos, esses pontos serão equivalentes a 100% de qualidade.

Outro fator que interfere no cálculo do nível de qualidade são as respostas que não possuem pontuação e as perguntas que não devem ser consideradas no cálculo da qualidade, que são: (i) “não se aplica”, é utilizada para proposições que fazem referência a um aspecto que não se enquadra ao domínio do aplicativo móvel sendo avaliado; e (ii) “avaliação prejudicada”, que se refere a proposições que o avaliador não está em condições de avaliar. Isso pode ocorrer por falta de recursos, informações ou conhecimento específico.

Nesse caso o resultado final deve ser normalizado inferindo o valor 100% para o total possível das questões de cada critério avaliado. O cálculo para obter o nível de qualidade por critério é dado pela fórmula a seguir:

$$NQCR = \frac{(p_1 + p_2 + p_3 + \dots + p_n) * 10}{TPC}$$

onde, $NQCR$ é o Nível de Qualidade por Critério; TPC é o Total de Perguntas consideradas por Critério; p é a pontuação atribuída para cada pergunta; e n é o total de perguntas do critério avaliado.

A partir dos resultados obtidos para cada critério, é possível calcular o nível de qualidade para cada uma das três categorias definidas no modelo (Pedagógica, Social e Técnica). O cálculo para obter o nível de qualidade por categoria é dado pela fórmula a seguir:

$$NQCA = \frac{NQCR_1 + NQCR_2 + NQCR_3 + \dots + NQCR_n}{TCQ}$$

onde, $NQCA$ é o Nível de Qualidade por Categoria; TCQ é o Total de Critérios de Qualidade relacionados à categoria avaliada; e n é o total de critérios relacionados à categoria avaliada.

Por fim, o resultado final que representa o nível de qualidade geral do aplicativo, considerando-se as três categorias é calculado. O cálculo para obter o nível de qualidade geral do aplicativo é dado pela fórmula a seguir:

$$NQA = \frac{NQP + NQS + NQT}{TC}$$

onde, NQA é o Nível de Qualidade do Aplicativo; NQP é o Nível de Qualidade da categoria Pedagógica; NQS é o Nível de Qualidade da categoria Social; NQT é o Nível de Qualidade da categoria Técnica; e TC é o Total de Categorias avaliadas, nesse caso será igual.

Para definir o nível de qualidade, o resultado final é atribuído de acordo com os critérios de julgamento definidos a partir do trabalho de Martinez et al. (1999) e também utilizado em Duarte Filho e Barbosa (2013a). Esta definição foi feita com o objetivo de definir um valor padrão. Dessa maneira, o nível de qualidade do aplicativo é dividido em três com seus respectivos valores padrões: (i) Superior, pontuação igual ou maior que 80%; (ii) Médio, pontuação igual ou maior que 50% e menor que 80%; e (iii) Baixo, pontuação menor que 50%.

De acordo com os possíveis níveis de qualidade do aplicativo, considera-se que um aplicativo com o nível superior possui um bom nível de qualidade, já que para obter essa pontuação o aplicativo conseguiu satisfazer os critérios de qualidade avaliados. Por outro lado, um nível de qualidade médio revela que o aplicativo necessita de melhorias.

Por fim, um nível de qualidade baixo sugere que o aplicativo não conseguiu alcançar níveis satisfatórios de qualidade. Dessa maneira, o resultado da avaliação sugere que esse aplicativo deva passar por uma reestruturação, com base nos critérios de qualidade não satisfeitos a partir da avaliação conduzida.

Embora o método MoLEva defina pontuações para suas questões, é importante ressaltar que os valores definidos podem ser alterados de acordo com os objetivos da avaliação. Dessa maneira, é possível inclusive atribuir pesos diferentes para cada critério de qualidade. Essa possibilidade permite que o método possa ser adaptado para cada necessidade. Por este motivo, a definição da pontuação exige conhecimento e bom senso do especialista que deseja customizá-las e atribuí-las ao método.

4. Avaliação do Método MoLEva

Para avaliar o método MoLEva, optou-se pela condução de um estudo de caso. O estudo de caso envolvendo a aplicação do método MoLEva foi conduzido no contexto de aplicativos para o ensino de idiomas. Além de disponibilizarem várias opções de aplicativos em diferentes sistemas operacionais, esse tipo de aplicativo possui um objetivo bem definido, que é o ensino de idiomas. Para este estudo de caso foram selecionados dois aplicativos educacionais móveis: (i) Duolingo; e (ii) Wlingua.

Tais aplicativos foram escolhidos pois, além de possuírem uma boa avaliação nas lojas de aplicativos, eles também estão disponíveis para os três principais sistemas operacionais existentes atualmente (Android, iOS e Windows).

4.1. Planejamento do Estudo de Caso

Antes da execução do estudo de caso foi necessário fazer o seu planejamento. Para isso foi necessária a definição dos seus objetivos. Além disso, também são definidos os perfis dos participantes do estudo. Por fim, são definidos os instrumentos utilizados durante o estudo de caso.

Objetivo

O objetivo do estudo de caso é validar a utilização do método MoLEva na avaliação de aplicativos educacionais móveis. Para isso é feita a análise do resultado obtido após a execução do método.

Participantes

Para o presente estudo de caso, foram selecionados oito alunos de pós-graduação do ICMC-USP na área Engenharia de Software, sendo um doutor, seis doutorandos e um mestre. Dentre os alunos participantes, dois já atuavam como professor.

Instrumentação

No que se refere à instrumentação utilizada no estudo de caso, os seguintes documentos foram apresentados aos participantes: (i) termo de adesão ao estudo experimental; (ii) formulário de caracterização do participante; (iii) documento com a orientação das atividades que o participante deveria realizar; (iv) ferramenta web utilizada para a aplicação do método MoLEva (<http://moleva-gustavosoad.rhcloud.com>).

4.2. Execução do Estudo de Caso

Para a execução do estudo de caso os participantes selecionados foram divididos em dois grupos. Essa divisão foi feita a partir das respostas de cada um deles no formulário de caracterização do participante. Esse documento foi submetido aos participantes objetivando coletar as informações necessárias para a divisão dos grupos.

Para que esta divisão fosse feita de forma balanceada, as seguintes características foram consideradas por ordem de prioridade: (i) conhecimento sobre qualidade de software; (ii) experiência em avaliação de qualidade de aplicativos; (iii) utilização dos aplicativos educacionais móveis contidos no estudo de caso; e (iv) nível de conhecimento no idioma inglês.

4.3. Análise dos Resultados

O estudo de caso foi executado por oito participantes, sendo que os dispositivos utilizados durante o estudo de caso foram exclusivamente smartphones. Já os sistemas operacionais utilizados foram variados. Embora a maioria tenha sido Android (5 avaliações), também foram utilizados o iOS (2 avaliações) e Windows (1 avaliação).

A análise dos resultados da avaliação de qualidade dos aplicativos é importante para a avaliação do método MoLEva, já que por meio dos resultados é possível identificar se a partir da aplicação do método conseguiu-se definir um nível de qualidade para o aplicativo, além de revelar suas deficiências. Para este estudo de caso, os valores dos critérios de julgamento mantiveram-se com o padrão definido pelo método MoLEva.

Com o objetivo de analisar duas avaliações de qualidade individualmente (uma do Duolingo e outra do Wlingua), foram selecionadas as avaliações que foram realizadas pelos avaliadores mais experientes de cada aplicação (o nível de experiência dos avaliadores foi baseado no seu conhecimento sobre avaliação de qualidade de software e aplicativos educacionais móveis). A Figura 2, apresenta a avaliação de qualidade dos aplicativos Duolingo e Wlingua sob a perspectiva dos critérios de qualidade estabelecidos pelo método.

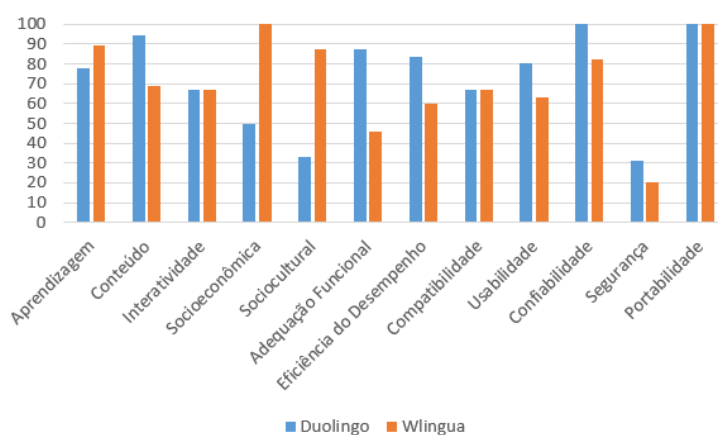


Figura 2. Resultado da Avaliação de Qualidade dos Aplicativos por Critérios de Qualidade: Avaliadores Experientes

O resultado por critérios de qualidade indica que o método MoLEva conseguiu identificar alguns critérios que chamam a atenção para melhoria nos aplicativos. Um desses critérios é “Segurança”, identificado nos dois aplicativos como sendo o critério com menor pontuação. Este resultado mostra que uma análise relacionada à segurança desses aplicativos deve ser realizada, pois aspectos de segurança podem acarretar riscos para os aplicativos e seus usuários.

Além disso, o aplicativo Duolingo possui outros dois critérios com o nível de qualidade mais baixo. Um deles é o critério “Sociocultural”, que sugere que o aplicativo pode ter problemas ao fornecer funcionalidades e se adaptar ao contexto cultural de seus usuários. Por exemplo, o aplicativo não possui mecanismos para denunciar mensagens ou conteúdos considerados ofensivos. O outro critério avaliado com o nível baixo é o “Socioeconômico”, indicando que o aplicativo deve dar maior atenção aos fatores soci-

ais e econômicos que permeiam o aplicativo. Por exemplo, foi identificado que não é possível reportar um problema pelo aplicativo, caso um problema ocorra.

Já o aplicativo Wlingua obteve melhores pontuações relacionadas a estes critérios. Por outro lado, o critério “Adequação Funcional” não foi bem pontuado, o que sugere que o aplicativo pode ter deficiências em suas funcionalidades, além de possíveis funcionalidades que deveriam existir no aplicativo, mas que não estão presentes.

Embora alguns critérios não tenham sido bem avaliados, é importante ressaltar que para ambos aplicativos o critério “Portabilidade” resultou em 100% da pontuação, indicando que os aplicativos podem ser utilizados em diversos tipos de dispositivos sem que ocorra problemas.

Além disso, é importante ressaltar que o aplicativo Duolingo obteve 100% para o critério “Confiabilidade”, indicando que o aplicativo possui um comportamento consistente. Já o aplicativo Wlingua também obteve 100% de pontuação em outro critério que foi o “Socioeconômico”, indicando que o aplicativo é viável e possui uma boa relação custo-benefício.

Para os dois avaliadores citados, também foram analisadas as informações obtidas por categoria, apresentadas na Figura 3. Observa-se que o Duolingo apresenta um baixo nível de qualidade para as questões relacionadas à categoria “Social”, embora apresente uma pontuação consistente para os critérios “Pedagógicos” e “Técnicos”. Por este motivo, é possível concluir que o aplicativo deveria ter atenção à categoria “Social”. Já o aplicativo Wlingua possui todos os critérios entre o nível médio e superior, demonstrando assim que todos as categorias estão com um bom nível de qualidade. Embora o resultado seja satisfatório para o aplicativo Wlingua, é importante que as deficiências identificadas sejam melhoradas, removendo assim as lacunas existentes no aplicativo.

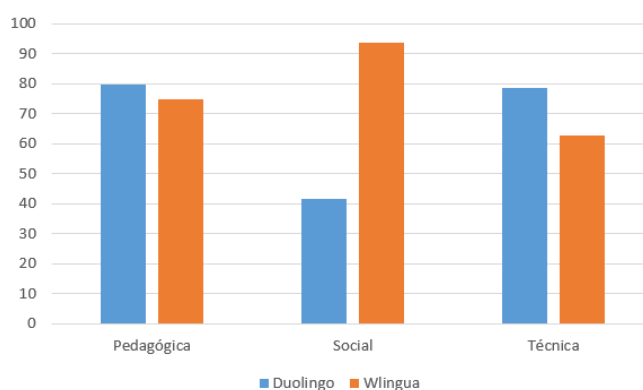


Figura 3. Resultado da Avaliação de Qualidade dos Aplicativos por Categoria: Avaliadores Experientes

Após a avaliação individual feita com os avaliadores mais experientes, também foram obtidos os dados relacionados à consolidação de todas as avaliações. A Figura 4, apresenta o resultado por critério de qualidade considerando a média de todas as avaliações. Observa-se que para o Duolingo houve uma melhora nas avaliações que estavam com pontuações menores na avaliação individual, porém a característica “Socioeconômica” continua com pontuação baixa. Este resultado é importante, pois

caso a pontuação seja um consenso entre os avaliadores, isso pode representar uma maior assertividade na análise do nível de qualidade.

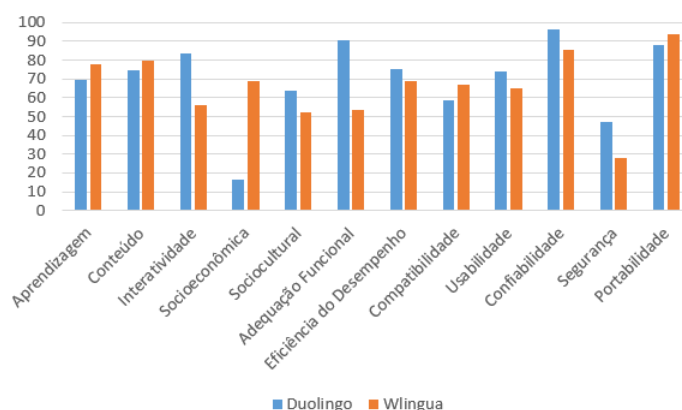


Figura 4. Resultado da Avaliação de Qualidade dos Aplicativos por Critério de Qualidade

Já as pontuações mais altas relacionadas ao Duolingo continuaram as mesmas, indicando que o resultado obtido com as avaliações está consistente com relação aos critérios mais bem avaliados. Com relação ao aplicativo Wlingua, suas pontuações também se mantiveram consistentes, porém nota-se que quase todos os critérios apresentaram quedas ao considerar todas as avaliações realizadas.

Com relação ao resultado por categoria, apresentado na Figura 5, observa-se que os níveis de qualidade obtiveram poucas alterações. Entretanto, para o aplicativo Wlingua, é possível identificar uma considerável alteração da avaliação individual para a geral. Na avaliação individual a categoria “Social” obteve 93,75 pontos, já na geral resultou em 60,42 pontos. Este resultado mostra uma diferença considerável entre as avaliações. Dessa maneira, é importante que seja considerado nos trabalhos futuros a investigação de divergências no entendimento das perguntas relacionadas à esta categoria.

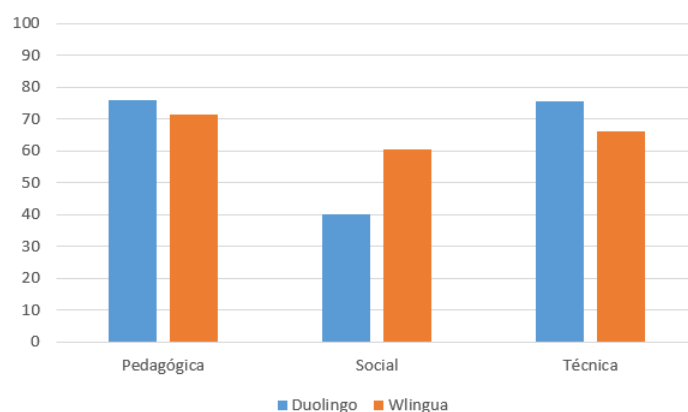


Figura 5. Resultado da Avaliação de Qualidade dos Aplicativos por Categoria

Por fim, o nível de qualidade definido para o aplicativo Duolingo foi de 63,91 pontos, já o Wlingua recebeu a pontuação final igual a 65,88 pontos. Este resultado mostra que ambos os aplicativos foram classificados com o nível médio de qualidade.

4.4. Ameaças à Validade

O primeiro aspecto a ser considerado é que a condução do estudo ocorreu de maneira remota, não sendo executada em um ambiente controlado. Dessa maneira, todos os avaliadores receberam as orientações por meio de documentos. Além disso, não houve nenhum treinamento relacionado ao método MoLEva, visto que as informações sobre o método estavam contidas apenas nos documentos.

Outro ponto importante está relacionado com a interpretação dos itens avaliados. Apesar de existir um exemplo para cada questão do checklist, uma interpretação incorreta pode ocasionar em uma avaliação incorreta.

Além disso, os participantes selecionados são todos do meio acadêmico e da mesma universidade. Seria interessante a participação de uma quantidade maior de participantes e de uma maneira mais abrangente, considerando outras universidades. É importante observar, ainda, que dentre os participantes nenhum é da indústria.

Ressalta-se que novos experimentos vêm sendo planejados de modo a considerar as ameaças identificadas. A execução de novos experimentos faz parte das atividades futuras associadas à continuidade da pesquisa, devendo ser realizada em curto prazo.

5. Conclusão

A partir dos resultados obtidos com o estudo de caso realizado foi possível concluir que o método MoLEva é factível de ser aplicado na avaliação de qualidade de aplicativos educacionais móveis. Este resultado é favorável para o método MoLEva, pois ele foi capaz de indicar pontos relevantes de melhoria para os aplicativos avaliados. Ressalta-se, também, que sua avaliação possibilita a comparação entre aplicativos, permitindo que o usuário escolha o aplicativo que possui os pontos positivos mais relevantes para sua necessidade, auxiliando assim na tomada de decisão.

Apesar das análises apresentadas fornecerem indícios de que o método MoLEva pode ser utilizado nas avaliações de qualidade de aplicativos educacionais móveis, é fundamental que mais estudos e experimentos formais sejam planejados e conduzidos, em diferentes domínios de aplicação.

Por fim, pretende-se fazer novas pesquisas baseadas nos resultados obtidos com o MoLEva, buscando desenvolver uma versão mais ágil de ser executada e automatizando alguns pontos da avaliação.

Agradecimentos

Os autores agradecem o financiamento brasileiro Agências - FAPESP (Processo 2014 / 03389-9), CAPES (Procad 071/2013) e CNPq.

Referências

- Abdurrahman, J., Beer, M., and Crowther, P. (2015). Pedagogical requirements for mobile learning: a review on mobilelearn task model. *Journal of Interactive Media in Education*, 2015(1):1–17.
- Acharya, A. and Sinha, D. (2013). Assessing the quality of m-learning systems using ISO/IEC 25010. *International Journal of Advanced Computer Research*, 3(3).
- Barbosa, E. F. (2004). Uma contribuição ao processo de desenvolvimento e modelagem de módulos educacionais. PhD thesis, Universidade de São Paulo.

- Bednarik, R., Gerdt, P., Mirafitabi, R., and Tukiainen, M. (2004). Development of the tup model - evaluating educational software. In IEEE International Conference on Advanced Learning Technologies, 2004. Proceedings., pages 699–701.
- Boehm, B. W., Brown, J. R., and Lipow, M. (1976). Quantitative evaluation of software quality. In Proceedings of the 2nd international conference on Software engineering, pages 592–605. IEEE Computer Society Press.
- Cavano, J. P. and McCall, J. A. (1978). A framework for the measurement of software quality. ACM SIGSOFT Software Engineering Notes, 3(5):133–139.
- Duarte Filho, N. F. (2016). Uma contribuição ao estabelecimento de uma arquitetura de referência para ambientes de aprendizagem móvel. PhD thesis, Universidade de São Paulo.
- Duarte Filho, N. F. and Barbosa, E. F. (2013a). A contribution to the quality evaluation of mobile learning environments. In Frontiers in Education Conference, 2013 IEEE, pages 379–382.
- Duarte Filho, N. F. and Barbosa, E. F. (2013b). A requirements catalog for mobile learning environments. In Proceedings of the 28th Annual ACM Symposium on Applied Computing, SAC '13, pages 1266–1271, New York, NY, USA. ACM.
- Economides, A. A. (2008). Requirements of mobile learning applications. International Journal of Innovation and Learning, 5(5).
- ISO/IEC 14598 (1998). ISO/IEC 14598: Information Technology - Evaluation of Software Products. ISO/IEC.
- ISO/IEC 25010 (2010). ISO/IEC 25010 - Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. Technical report.
- ISO/IEC 9126 (2001). ISO/IEC 9126. Software engineering – Product quality. ISO/IEC.
- Kearney, M., Schuck, S., Burden, K., and Aubusson, P. (2012). Viewing mobile learning from a pedagogical perspective. Research in Learning Technology, 20(0).
- Martinez, M., Azevedo, G., Lopes, S., Pagliuso, P., Colombo, R., Rodrigues, M., and Jino, M. (1999). The software product evaluation database-supporting mede-pros. In Software Engineering Standards, 1999. Proceedings. Fourth IEEE International Symposium and Forum on, pages 182–191.
- Nah, K. C., White, P., and Sussex, R. (2008). The potential of using a mobile phone to access the internet for learning efl listening skills within a korean context. ReCALL, 20(03):331–347.
- Rubens, N., Kaplan, D., and Okamoto, T. (2014). E-learning 3.0: Anyone, anywhere, anytime, and ai. In New Horizons in Web Based Learning, pages 171–180. Springer.
- Sarrab, M., Al-Shihi, H., and Rehman, O. (2013). Exploring major challenges and benefits of m-learning adoption. British Journal of Applied Science & Technology, 3(4):826 – 839.
- Sharples, M. (2013). Mobile learning: research, practice and challenges. Distance Education in China, 3(5):5–11.
- Soad, G., Filho, N. F. D., and Barbosa, E. F. (2015). Uma contribuição ao estabelecimento de características de qualidade para aplicações educacionais móveis. In XIV Simpósio Brasileiro de Qualidade de Software (SBQS), pages 165–179.
- Soad, G., Filho, N. F. D., and Barbosa, E. F. (2016). Quality evaluation of mobile learning applications. In 2016 IEEE Frontiers in Education Conference (FIE), pages 1–8.
- Wexler, S., Brown, J., Metcalf, D., Rogers, D., and Wagner, E. (2008). Mobile learning: What it is, why it matters, and how to incorporate it into your learning strategy. Guild Research.