

## APLICAÇÃO COMPARATIVA DE DIFERENTES ABORDAGENS DE AVALIAÇÃO PARA O *SOFTWARE* EDUCATIVO DUOLINGO: A COMPLEXIDADE DE ESCOLHER UMA ABORDAGEM ADEQUADA

Raphael Salviano T. da Silva<sup>1</sup>  
Williane Rodrigues de A. Silva<sup>2</sup>  
Raimundo J. Cardoso Filho<sup>3</sup>  
Wendell S. Pereira<sup>4</sup>  
Yuska P. C. Aguiar<sup>5</sup>  
Vanessa F. Dantas<sup>6</sup>

### RESUMO

A pluralidade de abordagens para avaliação de *Software* Educativo (SE) exige a tomada de decisão sobre qual abordagem utilizar na avaliação do SE que se pretende adotar. As abordagens possuem critérios para verificação da conformidade do SE a partir de escalas numéricas, subjetivas ou de valores percentuais. Diante das diferenças o educador se enfrenta a complexidade da escolha de uma abordagem cujos critérios e instrumentos de avaliação sejam claros, abrangentes e não ambíguos, e os resultados sejam fáceis de analisar. Neste trabalho descreve-se a avaliação do Duolingo a partir de diferentes abordagens, para identificar a complexidade, a abrangência e a qualidade dos resultados da aplicação de cada abordagem.

**Palavras-chave:** avaliação de *software* educativo; instrumentos de avaliação; duolingo.

### 1. Introdução

É cada vez mais comum o uso de recursos digitais no processo de ensino-aprendizagem, uma vez que é crescente a influência da Era da Informação em todos os setores da sociedade, sendo a educação um dos mais afetados pela inovação tecnológica (COSTA *et al.*, 2003). Segundo Frescki (2008), a função da educação é auxiliar as pessoas na produção da sua própria realidade material e de sua consciência sobre ela. Desta maneira, os *Softwares* Educativos (SE) podem auxiliar no processo de ensino-aprendizagem a partir da contextualização de determinados conteúdos, diante da realidade de cada aluno.

---

1 Estudante da Licenciatura em Ciência da Computação do Departamento de Ciências Exatas (DCX), da Universidade Federal da Paraíba (UFPB), Campus IV.

2 Estudante do Mestrado em Sistemas de Informação da Universidade de São Paulo (USP).

3 Estudante da Licenciatura em Ciência da Computação do DCX, da UFPB, Campus IV.

4 Estudante da Licenciatura em Ciência da Computação do DCX, da UFPB, Campus IV.

5 Doutora em Engenharia Elétrica. Docente da UFPB, atuante no DCX.

6 Mestre em Ciência da Computação. Docente da UFPB, atuante no DCX.

O uso dos *softwares* educativos tem, em potencial, uma boa aceitação, pois estes acrescentam elementos lúdicos no processo de ensino-aprendizagem. No entanto, algumas dificuldades podem ser enfrentadas pelos educadores na implantação de SEs no ambiente escolar. A escolha por um SE que corresponde às expectativas do professor e às necessidades pedagógicas do conteúdo abordado, é uma tarefa que requer atenção. Portanto, a adoção de um SE, assim como de qualquer outro recurso pedagógico, deve ser conduzida com cautela, sendo necessário garantir que o SE atenda a um conjunto de requisitos para que seu uso seja positivo no processo de ensino-aprendizagem.

O SE precisa funcionar corretamente, seus elementos pedagógicos precisam estar de acordo com os propósitos dos conteúdos abordados, e a interface deve ser condizente com as características, necessidades e limitações dos usuários. Problemas em qualquer um destes níveis pode prejudicar o processo de construção do conhecimento pelos aprendizes. Portanto, é necessário, antes de decidir adotar um SE, certificar-se de que este será um elemento agregador no aprendizado. Portanto, é necessário, antes de decidir sobre a adoção de um SE, certificar-se de que este será um elemento agregador no aprendizado, avaliando-o antes de seu uso.

Para Brito e Cavalcante (2003) avaliar um SE consiste em identificar como este pode ser utilizado no âmbito educacional de modo que ajude o aprendiz a construir seu conhecimento. A diversidade de abordagens para avaliação de SE disponíveis na literatura apresenta desafios para o educador. É difícil escolher a abordagem a adotar, considerando a natureza do SE (*desktop, mobile, web*) e o *background* e *expertise* do avaliador; é possível, ainda, que a avaliação do mesmo SE por instrumentos diferentes resulte em diagnósticos distintos. Nesse contexto, o presente trabalho tem por objetivo identificar a influência da escolha da abordagem de avaliação de SE no diagnóstico resultante da sua aplicação, e apresentar as dificuldades encontradas durante a realização das avaliações.

O artigo está organizado em cinco seções, incluindo esta. Na seção 2 apresentam-se as abordagens consideradas no estudo. Na seção 3, descreve-se a

metodologia adotada na execução das avaliações. A seção 4 contém os resultados obtidos com a aplicação das abordagens na avaliação do Duolingo, amplamente utilizado para o estudo de idiomas. Por fim, na seção 5, estão as considerações finais, contemplando as limitações e contribuições desta pesquisa, e trabalhos futuros.

## 2. Métodos de Avaliação de Softwares Educacionais

A avaliação de um SE deve considerar as características voltadas à sua qualidade didático-pedagógica, de tal modo que os objetivos dos estudiosos da ergonomia de *software* e dos educadores possam convergir para um mesmo ponto (SILVA, 1998) *apud* (ANDRES; CYBIS, 2000). Vários autores (ROCHA; CAMPOS, 1992); (SILVA; VARGAS, 1999); (OLIVEIRA *et al.*, 2001) argumentam sobre a necessidade de considerar paradigmas pedagógicos na avaliação de SEs a fim de orientar os professores na aquisição e no emprego desses produtos. Além disso, os SEs devem ser funcionalmente corretos, de acordo com os preceitos de Engenharia de *Software* (SOMMERVILLE, 2007), sendo importante ainda que a interatividade e a interface oferecidas pelo SE sejam adequadas ao público ao qual se destina (PREECE *et al.*, 2013).

Segundo Chacon *et al.* (2012) as principais abordagens objetivas (uso de *checklist*) para a avaliação de *softwares* educativos são: (i) Método de Reeves (Campos, 1989); (ii) Técnica de Mucchielli (SILVA, 1998); (iii) *Learning Object Review Instrument* (LORI) (NESBIT *et al.*, 2002); (iv) Técnica de Inspeção de Conformidade Ergonômica de Software Educacional (TICESE) (GAMEZ, 1998).

A variedade de abordagens disponíveis para avaliar SEs reflete, de forma significativa, em diferenças entre estas, que se apresentam em relação (i) ao número de critérios considerados; (ii) à natureza destes critérios (aspectos pedagógicos, qualidade de *software*, qualidade de uso); (iii) à natureza e granularidade das escalas adotadas (instrumentos de mensuração) na avaliação dos critérios e, (iv) à qualidade dos resultados alcançados (diagnóstico). Neste sentido, o educador pode se questionar sobre

qual seria a abordagem mais adequada para avaliar um SE. A fim de familiarizar o educador sobre as características específicas de cada uma das abordagens, segue uma breve explicação sobre estas.

### **2.1. Método de Reeves**

Este método consiste em um conjunto de 24 critérios para avaliação do SE, sendo 10 relacionados à interface do usuário (usabilidade) e 14 voltados para aspectos pedagógicos. Como instrumento de mensuração cada critério está associado a uma escala bidirecional não numérica, que se utiliza de conceitos antagônicos. Para os critérios de usabilidade, os conceitos positivos estão à direita e os negativos à esquerda. Por exemplo: para o critério “facilidade de uso”, à esquerda da seta tem-se o conceito “difícil”, e à direita “fácil”. Cabe ao avaliador verificar, para cada critério, como o SE se apresenta, e realizar uma marca sobre a escala. Neste caso, havendo predominância das marcações à esquerda, percebe-se fragilidade do SE em relação aos aspectos de usabilidade. Para os critérios pedagógicos, os conceitos apresentados nas setas não são, necessariamente, positivos e negativos. Por exemplo, para “Sequenciamento Instrucional”, à esquerda tem-se “Reducionista” e à direita “Construtivista”. A marcação de pontos à direita ou à esquerda, não indica, explicitamente, a qualidade do SE para estes critérios, mas sim sua caracterização. Como resultado das avaliações, tem-se o conjunto de critérios e suas setas marcadas. O diagnóstico é realizado pelo educador mediante a análise da disposição dos pontos marcados nas setas. Sugere-se que os pontos sejam ligados a fim de facilitar a visualização da informação.

### **2.2. Técnica de Mucchielli**

Permite a avaliação global de *software* considerando o público para o qual o SE foi desenvolvido. A técnica propõe 10 (dez) critérios a serem avaliados a partir de um *checklist* cujos itens estão relacionados com uma escala numérica com pontuação variando entre 1 (um) e 5 (cinco). O resultado de sua aplicação resulta em um conjunto de barras para cada qual um critério foi associado à um valor da escala adotada contendo a marcação dos valores atribuídos pelo avaliador.

### **2.3. Avaliação de LORI**

Consiste em um *checklist* que compreende um total de 9 critérios, sendo 1 pedagógico, 4 relativos à interface, e 4 de qualidade de *software*. Assim como Mucchielli, os critérios estão associados a escalas numéricas de 5 pontos com variação de números inteiros de 1 (um) à 5 (cinco). O resultado obtido com sua aplicação é semelhante ao de Mucchielli.

### **2.4. Técnica de TICESE**

Esta técnica possui um foco maior sobre a ergonomia de *software* aplicada a produtos educativos informatizados, orientando o avaliador a realizar uma inspeção na ergonomia do SE. A técnica de TICESE é composta por três módulos: (i) módulo de classificação, um formulário a respeito de informações técnicas do *software*; (ii) módulo de avaliação, subdividido em dois tipos de avaliações, a avaliação da documentação que possui 2 (dois) critérios, e a avaliação do produto que contém 10 (dez) critérios; e (iii) o módulo de avaliação contextual, que é composto por apenas (1) um critério. Os critérios de cada módulo possuem um peso que varia de 0 (zero) à 1,5 (um e meio), onde 0 (zero) representa não se aplica, 1 (um) importante e 1,5 (um e meio) muito importante. O resultado obtido é um valor percentual de adequação do SE aos critérios analisados.

## **3. Metodologia Aplicada**

Para analisar, comparativamente, as abordagens para avaliação de SE apresentadas na seção anterior, adotou-se a seguinte estratégia: (i) Definição da equipe de avaliação: quatro discentes do curso de Licenciatura em Ciência da Computação da Universidade Federal da Paraíba (UFPB-Campus IV), sob a supervisão de duas professoras com experiência em Interação Humano-Computador (IHC), qualidade de *software* e engenharia de *software*; (ii) Experiência da equipe: dois discentes cursaram a disciplina de Softwares Educacionais oferecida pela IES, tendo experiência prévia com avaliação de SE a partir da aprovação do métodos de Reeves. Todos os docentes

passaram por um período de estudo teórico sobre as abordagens de avaliação, experienciando a aplicação destas com auxílio das professoras; um dos discentes é professor de língua estrangeira; (iii) Escolha do SE: o Duolingo (<https://www.duolingo.com/>) é um aplicativo para o estudo de idiomas disponível nas versões *web* e *mobile*. Seu conteúdo é agrupado em níveis (iniciante, intermediário e avançado), com temas de estudo (saudações, comidas, animais, família, etc) organizados em lições e atividades. O nível de dificuldade das lições aumenta à medida que o aluno progride entre os níveis. Foram avaliadas as dez primeiras lições disponíveis no Duolingo, em ambas as versões, a fim de verificar se as abordagens são válidas independente da plataforma do SE (*web* ou *mobile*); (iv) Definição da dinâmica de avaliação: os alunos formaram duplas com diferentes configurações, alternando entre os membros, as versões do SE (*web* e *mobile*), e as abordagens de avaliação, de modo a evitar uma percepção tendenciosa e/ou limitada; e (v) Escolha de elementos de interesse da análise comparativa: facilidade de aplicação (tempo necessário e dúvidas frequentes), abrangência da abordagem frente aos critérios e facilidade de interpretação dos resultados.

#### **4. Avaliação do Duolingo: versões web e mobile**

Nas subseções seguintes, são apresentados os resultados das avaliações realizadas. A análise comparativa refere-se à aplicação das quatro abordagens consideradas neste estudo (REEVES, MUCCHIELLI, LORI e TICESE) para as versões *web* e *mobile* do Duolingo, a fim de investigar se existem diferenças no diagnóstico obtido. Por fim, tem-se uma análise geral e o relato das principais dificuldades enfrentadas na aplicação das abordagens.

##### **4.1. Resultado da aplicação de Reeves**

Os resultados apresentam semelhanças entre as versões *web* e *mobile* (Figuras 1 e 2) para todos os critérios. Em ambas as versões os aspectos de interface do usuário foram associados a conceitos positivos. A diferença nos critérios pedagógicos

acontece para a acomodação das diferenças individuais (mais próximo de “*multifacetada*” para a versão *mobile*) e aprendizado cooperativo (mais próximo de “*integral*” para a versão *web*). Portanto, pode-se afirmar que os elementos pedagógicos de ambas as versões são consistentes entre si. Esta característica faz com que o aluno possa alternar entre as versões sem prejuízo.

O fato deste método não usar uma escala numérica para quantificar resultados, apenas setas graficamente com marcações visuais, torna a avaliação muito subjetiva e difícil para avaliadores inexperientes, que ficam inseguros sobre como representar as fragilidades do SE. A aplicação do método demandou um total de 4 horas para que todos os aspectos fossem checados e a equipe de avaliação chegasse a um consenso sobre a marcação a ser associada a cada critério - evidenciando a essencia subjetiva da avaliação do SE. Além disto, percebe-se que o resultado final pode ser inconclusivo para indicar se o *software* é recomendável ou não, especialmente no caso dos critérios pedagógicos, visto que estes são conceitos antagônicos, mas não necessariamente bons ou ruins.

**Figura 1** - Critérios de Interface da aplicação de Reeves para avaliação do Duolingo nas versões Web e Mobile.

#### Critérios da interface com o usuário

##### Versão Web

1. Facilidade de uso	Difícil	←	→	Fácil
2. Navegação	Difícil	←	→	Fácil
3. Carga cognitiva	Não-gerenciável	←	→	Intuitiva
4. Mapeamento	Nenhum	←	→	Poderoso
5. Design de tela	Princípios violados	←	→	Princípios preservados
6. Compatibilidade espacial do conhecimento	Incompatível	←	→	Compatível
7. Apresentação do conhecimento	Confusa	←	→	Clara
8. Integração das mídias	Não-coordenada	←	→	Coordenada
9. Estética	Desagradável	←	→	Agradável
10. Funcionalidade geral	Não funciona	←	→	Funciona

##### Versão Mobile

1. Facilidade de uso	Difícil	←	→	Fácil
2. Navegação	Difícil	←	→	Fácil
3. Carga cognitiva	Não-gerenciável	←	→	Intuitiva
4. Mapeamento	Nenhum	←	→	Poderoso
5. Design de tela	Princípios violados	←	→	Princípios preservados
6. Compatibilidade espacial do conhecimento	Incompatível	←	→	Compatível
7. Apresentação do conhecimento	Confusa	←	→	Clara
8. Integração das mídias	Não-coordenada	←	→	Coordenada
9. Estética	Desagradável	←	→	Agradável
10. Funcionalidade geral	Não funciona	←	→	Funciona

**Figura 2 - Critérios Pedagógicos da aplicação de Reeves para avaliação do Duolingo nas versões Web e Mobile**



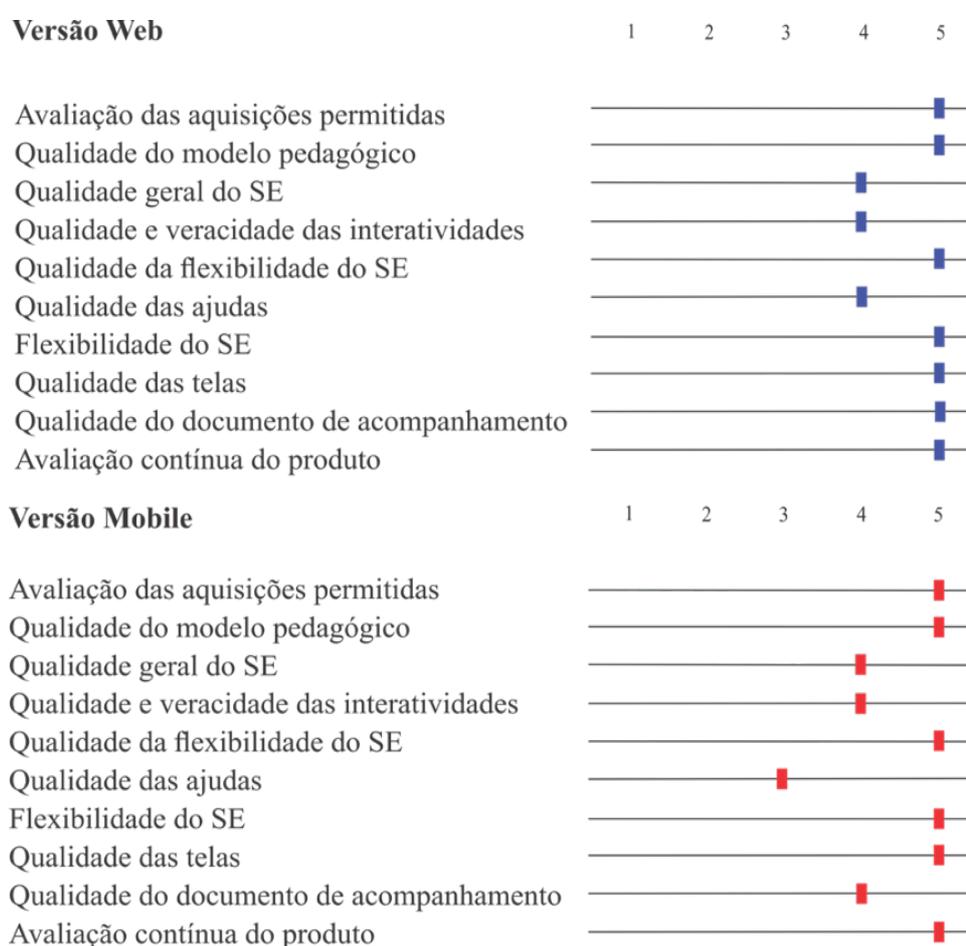
#### 4.2. Resultado da aplicação de Mucchielli

Os 10 critérios de Mucchielli foram avaliados de forma positiva para ambas as versões recebendo conceito máximo em sua maioria (Figura 3). Entretanto, dois critérios apresentam discrepâncias que influenciam na percepção do resultado: *a qualidade das ajudas* e *a qualidade do documento de acompanhamento*. Na versão *web*, o site possui informações úteis ao usuário, como detalhes sobre como funcionam as lições, enquanto que na versão *mobile* têm-se apenas informações técnicas (desenvolvedor, versão, etc.) disponíveis “loja” a partir da qual se adquire o aplicativo. A versão *web* obteve 47 pontos e a *mobile* 45, de um total de 50 possíveis.

Diferente de Reeves, esta técnica não agrupa os critérios em pedagógicos e de usabilidade, exigindo do avaliador mais atenção ao analisar os resultados antes de afirmar que os problemas encontrados no SE são prioritariamente de interface ou

pedagógicos. Além disso, é interessante observar que 7 dos 10 critérios utilizam o termo *qualidade*, que por si só é subjetivo. A aplicação de Mucchielli demandou um total de 3 horas para que os aspectos de usabilidade, qualidade de software e pedagógicos fossem checados e a equipe de avaliação chegasse a um consenso sobre a pontuação a ser associada a cada critério. A qualidade percebida por um avaliador experiente certamente será diferente daquela observada por um principiante, visto que o primeiro poderá ter uma gama diversificada de referenciais que afetarão seu julgamento.

**Figura 3** - Aplicação de Mucchielli para avaliação do Duolingo nas versões Web e Mobile

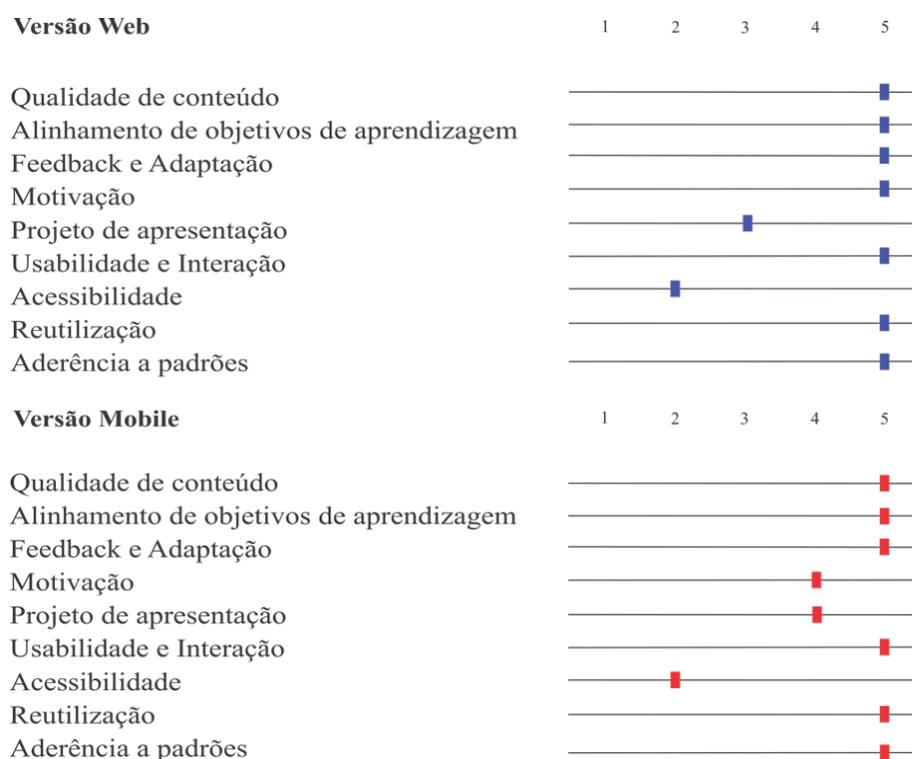


### 4.3. Resultado da aplicação de LORI

O resultado diverge levemente entre as versões *web* e *mobile* em relação aos critérios “*motivação*” e “*projeto e apresentação*” (Figura 3). Na versão *mobile*, a motivação é menos favorável devido ao fato do usuário perder pontos quando comete erros em suas respostas. Entretanto, nesta mesma versão, o recurso de voz utilizado para submeter respostas contribuiu para uma pontuação mais alta no aspecto projeto e apresentação. Embora o somatório obtido para ambas as versões seja o mesmo (39 pontos de um total de 50), percebe-se que a versão *web* teve mais critérios com pontuação máxima.

A aplicação do método demandou um total de 3 horas para que todos os aspectos fossem checados e a equipe de avaliação chegasse a um consenso sobre a pontuação associada a cada critério. Além disto, percebe-se que a análise do resultado final é conclusiva, por se tratar de valores numéricos.

**Figura 4** - Aplicação de Lori para avaliação do Duolingo nas versões Web e Mobile



Andres e Cybis (2000) “Um Estudo Teórico sobre as Técnicas de Avaliação de Software Educacional”. In VI Congresso Argentino de Ciências de la Computación.

#### **4.4. Resultado da aplicação de TICESE**

Nas versões *web* e *mobile* os valores relativos à avaliação contextual foram idênticos. Porém, algumas situações de inadequação (ou insuficiência) foram observadas. Na versão *web* houve uma baixa avaliação, inferior a 75%, para os itens: *dados de identificação* (29%), *controle explícito* (36,5%), *adaptabilidade* (46.5%) e *gestão de erros* (66%). Na versão *mobile* o mesmo aconteceu para os critérios de *dados de identificação* (51.7%), *adaptabilidade* (59,5%) e *gestão de erros* (58%). Vale ressaltar, no entanto, que os dois primeiros foram melhor avaliados em relação à versão *web*. Percebe-se uma queda significativa no critério de *avaliação da aprendizagem* (de 94% na versão *web* para 72% na versão *mobile*). Observa-se, ainda, uma elevação no critério da *documentação* (de 64.5% na versão *web* para 70.45% na versão *mobile*) e uma diminuição para a *avaliação geral do produto* (de 88.5% na versão *web* para 82.5% na versão *mobile*). Portanto, considerando a percepção da qualidade do produto pelo usuário, a versão *web* possivelmente seria considerada mais completa e melhor recomendada em relação à *mobile*.

É importante ressaltar que os valores percentuais obtidos são calculados a partir de subcritérios, e de suas respectivas escalas, associados a cada item presente na Figura 4. O detalhamento de critérios em subcritérios guia o avaliador e minimiza o nível de abstração envolvido na aplicação do método. No entanto, esta técnica exigiu o maior tempo de dedicação (cerca de 14 horas) por parte do avaliador, foi necessário o uso de planilhas auxiliares para guiar o acompanhamento de cada critério, e para a execução dos cálculos exigidos pela técnica. A utilização da planilha para a realização dos cálculos e organização dos critérios, facilitam a leitura e visualização para o avaliador. O fato dos resultados serem apresentados em valores percentuais pode facilitar na interpretação do diagnóstico uma vez que é usual trabalhar esta forma de representação numérica.

**Figura 5** - Aplicação de Ticese para avaliação do Duolingo nas versões Web e Mobile.

<b>Avaliação da Documentação</b> Versão Mobile		<b>Avaliação da Documentação</b> Versão Web	
( 51,75% ) Dados de identificação	70,45%	( 29% ) Dados de identificação	64,50%
( 89,16% ) Qualidade da informação impressa		( 100% ) Qualidade da informação impressa	
<b>Avaliação do Produto</b>		<b>Avaliação do Produto</b>	
( 83% ) Condução	82,50%	( 75,25% ) Condução	88%
( 59,50% ) Adaptabilidade		( 46,50% ) Adaptabilidade	
( 85% ) Controle explícito		( 36,50% ) Controle explícito	
( 78% ) Recursos de apoio à compreensão dos conteúdos		( 84% ) Recursos de apoio à compreensão dos conteúdos	
( 58% ) Gestão de erros		( 66% ) Gestão de erros	
( 72% ) Avaliação da aprendizagem		( 94% ) Avaliação da aprendizagem	
( 92,50% ) Carga de trabalho		( 93% ) Carga de trabalho	
( 82% ) Significado dos códigos e denominações		( 92% ) Significado dos códigos e denominações	
( 85% ) Homogeneidade		( 93% ) Homogeneidade	
( 100% ) Compatibilidade		( 100% ) Compatibilidade	
<b>Avaliação Contextual</b>	92%	<b>Avaliação Contextual</b>	92%
( 92% ) Adequabilidade		( 92% ) Adequabilidade	

#### 4.7. Análise geral dos resultados

De modo geral, as avaliações foram positivas para o SE independente da plataforma ser *mobile* ou *web*, assim como da abordagem aplicada. Entretanto, faz-se necessário observar alguns aspectos qualitativos, muitas vezes não revelados explicitamente nos diagnósticos resultantes da aplicação das abordagens. Na Tabela 1 têm-se informações relevantes para esta pesquisa considerando as (i) dimensões consideradas pelos critérios adotados nas abordagens (pedagógico, de usabilidade e de qualidade de software); (ii) a dificuldade de compreensão dos critérios, suas escalas e interpretação do diagnóstico resultante, e (iii) do tempo necessário para aplicação de cada abordagem.

A quantidade de critérios entre as técnicas varia de 9 (LORI) à 24 (Reeves), que pode representar uma carga de trabalho bastante diferenciada para os avaliadores. A escolha por uma abordagem com menos critérios pode induzir o pensamento sobre a rapidez em aplicar a avaliação.

Tabela 1 - Análise comparativa entre as abordagens

		Abordagens aplicadas para avaliação do Duolingo (Web e Mobile)			
		Reeves	TICESE	Mucchielli	LORI
Abrangência dos critérios	Pedagógico	14	3	3	1
	Usabilidade	10	3	7	4
	Qualidade de SW	0	7	0	4
Quantidade Total de critérios		<b>24</b>	<b>13</b>	<b>10</b>	<b>9</b>
Dificuldade para compreender	Crítérios	Média	Média	Média	Média
	Instrumentos de medição	Baixa	<b>Alta</b>	Baixa	Baixa
	Diagnóstico	Média	Média	Baixa	Baixa
Tempo empregado para	Aprendizado	2h	4h	1h	1h
	Aplicação	4h	10h	2h	2h
Tempo Total		<b>6h</b>	<b>14h</b>	<b>3h</b>	<b>3h</b>

Além da quantidade de critérios, quais aspectos podem ser medidos pelas abordagens é algo bastante relevante. O fato de algumas abordagens não considerarem critérios de qualidade do *software*, como ocorre com Reeves e Mucchielli, pode induzir o avaliador a escolher um SE que seja muito bom do ponto de vista pedagógico, mas que não seja bem documentado ou apresente erros de funcionamento, o que interfere no processo ensino-aprendizagem. O mesmo equívoco poderia acontecer caso fossem adotadas abordagens que privilegiam aspectos técnicos e que dão pouca ênfase às questões pedagógicas, como LORI. O avaliador poderia escolher um excelente *software*, mas que deixasse a desejar como instrumento de apoio à educação.

É válido ressaltar que a aplicação de cada abordagem permitiu identificar limitações em diferentes aspectos do *software*. Enquanto Mucchielli identificou

restrições sobre ajuda para o usuário, LORI indica deficiências na motivação e no projeto de apresentação, e TICESE demonstrou ressalvas sobre o controle do usuário e a documentação do SE. Sendo assim, percebe-se que a escolha da abordagem de avaliação utilizada, influencia na percepção que se tem sobre o SE, explicitando certos aspectos ou ocultando fragilidades.

Outro aspecto que não está explícito nos resultados quantitativos das avaliações, é a influência da experiência e conhecimento prévios dos avaliadores na escolha e aplicação das abordagens. A equipe envolvida na avaliação do Duolingo era composta por quatro alunos com conhecimento intermediário em avaliação de SE, e mesmo assim o tempo demandado para o estudo teórico das abordagens exigiu dedicação da parte dos envolvidos. A cada nova abordagem aplicada, foi necessário estudar e entender os critérios, familiarizar-se com gráficos e escalas numéricas, e ao mesmo tempo explorar o *software* em busca da verificação da conformidade (ou não) aos critérios.

Num contexto real, em que o avaliador dispõe de prazos para avaliar os *softwares* educativos disponíveis no mercado e escolher qual deles adotar, é pouco provável que ele possa se dedicar ao domínio de diferentes abordagens avaliativas. Sendo assim, existe uma predisposição ao uso de uma abordagem com a qual o avaliador já tenha certa experiência, ou cuja aplicação seja mais simples e rápida, de modo a otimizar seu trabalho. Porém, a análise isolada dos resultados de uma única abordagem, como já discutido anteriormente, pode implicar em equívocos e julgamentos imprecisos.

Para a avaliação específica do Duolingo (*mobile* e *web*), embora seja possível indicar aspectos distintos para a aplicação de cada abordagem (abrangência dos critérios, tempo de aplicação e qualidade dos resultados alcançados) o diagnóstico final apresentado para o Duolingo foi satisfatório, sugerindo sua adoção. Este resultado pode ser decorrente do fato deste *software* ser estável e largamente utilizado para o estudo de línguas, realidade diferente para boa parte de *Softwares* Educativos.

## 5. Conclusão

A avaliação de um SE é uma etapa de extrema importância para justificar sua adoção como recurso educacional. Diversas abordagens de avaliação estão disponíveis na literatura para apoiar essa atividade, mas, a complexidade de seus critérios e a subjetividade de seus resultados muitas vezes podem confundir o avaliador e interferir na decisão sobre adoção ou não de um SE.

Este trabalho relata a aplicação de quatro abordagens de avaliação de um SE para aprendizado de idiomas, nas versões *web* e *mobile*. A análise realizada evidencia de forma geral que o SE obteve bons resultados independente da abordagem aplicada, e mostra também a subjetividade dos critérios adotados nas abordagens. Estes dois aspectos consistem em desafios na aplicação das abordagens e na interpretação dos diagnósticos alcançados, o que impacta na decisão dos avaliadores sobre a adoção do SE em questão.

Ainda se faz necessário destacar que cada abordagem possui enfoque particular visando avaliar aspectos específicos do SE. Com base nessa percepção, é interessante destacar que, a escolha da abordagem de avaliação pode variar de acordo com o objetivo definido pelo agente avaliador. A escolha equivocada da abordagem pode resultar na ausência da observação de critérios definidos como importantes pelo próprio agente avaliador, uma vez que deve haver uma comunhão entre o que se pretende avaliar e os critérios presentes na técnica escolhida.

Considerando a experiência vivenciada e o conhecimento adquirido, pretende-se direcionar a pesquisa para a elaboração de um guia de aplicação para avaliação de SE que combine aspectos e critérios de diferentes abordagens. Objetiva-se fornecer aos avaliadores um instrumento que diminua a necessidade de *expertise* e *background* dos avaliadores, diminua a abstração e subjetividade dos critérios analisados e resulte em um diagnóstico cuja interpretação seja mais clara, direta e objetiva.

## Referências

ANDRES, D. P.; CYBIS, W. A. **Um Estudo Teórico sobre as Técnicas de Avaliação de Software Educacional**. In VI Congresso Argentino de Ciências de la Computación, 2000.

BRITO, C. L., ALMEIDA, I. A.; CAVALCANTI, L. B. **O que se avalia e o que é Preciso Avaliar em um Software Educativo?** Anais do Workshop de Informática na Escola, 2003. Vol. 1. No. 1. p. 334-344.

CAMPOS, G. H. B. **Construção e validação de ficha de avaliação de produtos educacionais para microcomputadores**. Dissertação de Mestrado – Faculdade de Educação, Universidade Federal do Rio de Janeiro, 1989. p.110.

CHACON E. P., BORGES M. N., SILVA C. R. C., CLUA E. W. G. **Check-List: um formulário para avaliação de softwares educativos**. In III Encontro Nacional de Ensino de Ciências da Saúde e do Meio Ambiente (ENECIÊNCIAS), 16 a 19 de Maio, 2012, Niterói - RJ. [http://www2.ic.uff.br/~ccaetano/artigos/Artigo\\_T-210\\_ENECIENCIAS.pdf](http://www2.ic.uff.br/~ccaetano/artigos/Artigo_T-210_ENECIENCIAS.pdf). Acessado em maio de 2015.

COSTA, V. M., RAPKIEWICZ, C. E., de QUEIRÓS FILHO, M. G.; CANELA, M. C. **Avaliação de sites educacionais de Química e Física: um estudo comparativo**. In Anais do Workshop de Informática na Escola, 2003. Vol. 1, No. 1, p. 545-554.

FRESCKI, F. B. **Avaliação da qualidade de softwares educacionais para o ensino de álgebra**. Centro de Ciências Exatas e Tecnológicas, Universidade Estadual do Oeste do Paraná, campus Cascavel, 2008.

GAMEZ, L. TICESE. **Técnica de inspeção de conformidade ergonômica de software educacional**. Dissertação de Mestrado submetida à Universidade do Minho. Portugal, 1998.

NESBIT, J; BELFER, K; LEACOCK, T. **Learning Object Review Instrument (LORI)**, <http://cenlinu1.centennialcollege.ca/aahs/LORI/help.php>., 2002.

OLIVEIRA, C. C. O., COSTA, J. W.; MOREIRA, M. **Ambientes informatizados de aprendizagem: produção e avaliação de software educativo**. Papirus, 2001.

PREECE, J.; ROGERS, Y.; SHARP, E. **Design de Interação, além da interação humano-computador**. Editora Bookman, 3ª Edição, 2013.

ROCHA, A. R.; CAMPOS, G.B. **Avaliação de Qualidade de Software Educacional**. São Paulo, 1992. Em Aberto, n. 57, v.12.

SILVA, C. R. **Bases Pedagógicas e Ergonômicas para Concepção e Avaliação de Produtos Educacionais Informatizados**. Dissertação de Mestrado submetida à Universidade Federal de Santa Catarina. Florianópolis, 1998.

SILVA, C. R.; VARGAS, C. L. S. **Avaliação de Qualidade de Software Educacional**. Anais do XIX Encontro Nacional de Engenharia de Produção e V International Congress of Industrial Engineering, Rio de Janeiro, 1999.

SOMMERVILLE, Ian. **Engenharia de Software**. Ed Person Education. 8ª Edição, 2007.