

VALIDAÇÃO DE UMA ABORDAGEM COMBINADA PARA AVALIAÇÃO DE SOFTWARE EDUCATIVO: AVANÇOS E DESAFIOS

Wendell S. Pereira¹
Raimundo J. Cardoso Filho²
Williane Rodrigues de A. Silva³
Raphael Salviano T. Da Silva⁴
Vanessa F. Dantas⁵
Yuska P. C. Aguiar⁶

RESUMO

Diante da necessidade de avaliar um *software* educacional antes de decidir sobre sua adoção, e da diversidade de abordagens de avaliação disponíveis, é comum o surgimento de dúvidas sobre a escolha da técnica mais adequada. Cada uma delas propõe critérios e formas de representação diferentes, e fica difícil escolher a mais abrangente para que aspectos pedagógicos, de qualidade de *software*, e de qualidade de uso sejam contemplados. Sendo assim, o objetivo desse trabalho é realizar a avaliação de um software educacional segundo uma abordagem combinada de técnicas, a fim de identificar equivalência de critérios entre elas e possíveis fragilidades.

Palavras-chave: avaliação de software educativo; abordagens de avaliação; qualidade de *software*.

1. Introdução

A Era da Informação influencia todos os setores da sociedade, e a inovação tecnológica tem afetado especialmente a educação, com o uso cada vez mais comum de recursos digitais no processo de ensino-aprendizagem (Costa *et al.* 2003). Segundo Freski (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 (SEs) podem auxiliar no processo de ensino-aprendizagem a partir da contextualização de determinados conteúdos para a realidade de cada aluno (Morais 2003). O SE é concebido com o objetivo principal de facilitar o processo de ensino-

-
- 1 Estudante da Licenciatura em Ciência da Computação do Departamento de Ciências Exatas (DCX), da Universidade Federal da Paraíba (UFPB), Campus IV.
 - 2 Estudante da Licenciatura em Ciência da Computação do DCX, da UFPB.
 - 3 Estudante do Mestrado em Sistemas de Informação da Universidade de São Paulo (USP).
 - 4 Estudante da Licenciatura em Ciência da Computação do DCX, da UFPB.
 - 5 Mestre em Ciência da Computação. Docente da UFPB, atuante no DCX.
 - 6 Doutora em Engenharia Elétrica. Docente da UFPB, atuante no DCX.

aprendizagem, sendo composto por um conjunto de recursos computacionais projetados com a intenção de serem usados em um contexto educacional (Sancho 1998).

A ludicidade dos recursos tecnológicos atrai os discentes, que rapidamente se envolvem no processo de ensino-aprendizagem. Porém, além das dificuldades de adaptação encontradas pelos docentes na implantação desses recursos no ambiente escolar, estes encaram ainda o desafio de escolherem SEs que correspondam às suas expectativas, que sejam adequados aos componentes curriculares a serem trabalhados, e direcionados para o público discente ao qual se destinam. Nessa perspectiva, a escolha do SE tem impacto direto no processo de ensino-aprendizagem do aluno, assim como na eficácia da metodologia utilizada pelo professor em sala de aula. Um SE escolhido sem uma avaliação prévia do professor, ou de algum agente avaliador, pode tornar-se um instrumento que dificulte o alcance dos objetivos do professor na aula que ministra, causando distração e distanciando os alunos do conteúdo que deveria ser trabalhado, por exemplo.

Portanto, para decidir sobre a adoção de um SE como recurso de apoio no processo de ensino-aprendizagem, o docente deve considerar se seus *elementos pedagógicos* estão de acordo com os propósitos dos conteúdos abordados, se a interface apresenta-se com usabilidade e ergonomia condizentes com as características, necessidades e limitações dos discentes, e se a qualidade de *software*, do ponto de vista de seu desenvolvimento, é satisfatória. Como dito anteriormente, problemas em qualquer um destes níveis podem prejudicar o processo de construção do conhecimento pelos discentes. Portanto, é necessário, antes de decidir adotar um SE, certificar-se de que este será um elemento agregador no aprendizado. Porém, a mesma dificuldade encontrada em adotar um SE pode ser encontrada na hora de buscar abordagens para avaliar o SE.

Para realizar avaliação de SE, é possível identificar na literatura vários métodos, metodologias e técnicas de avaliação, a exemplo de: Método de Reeves (Campos 1989); Técnica de TICESE (Gamez 1998); Técnica de Mucchielli (Silva

1998); Avaliação de LORI (Nesbit *et al.* 2002); Metodologia de Martins (Martins 2004), Método Rocha (Rocha 1992) e Modelo de Avaliação de Campos (Campos 1994). Estas são abordagens objetivas de avaliação, uma vez que se apoiam em um conjunto de critérios (*checklists*), associados a uma escala de avaliação de conformidade, e sua aplicação resulta em um diagnóstico indicando se o SE contempla ou não os critérios de interesse. Cada abordagem possui critérios, escalas e diagnósticos específicos, conforme detalhado a seguir.

O método de Reeves (Campos 1989) utiliza uma escala bidirecional, em que o avaliador deve posicionar um ponto para expressar sua opinião ao avaliar os critérios propostos pelo seu *checklist*. O critério é avaliado de forma negativa se o ponto se aproxima da extremidade esquerda; da mesma maneira, se o ponto estiver mais próximo da extremidade direita da escala, o critério é avaliado positivamente. O *checklist* é particionado em duas etapas: a primeira envolve os critérios de interface, e a segunda diz respeito aos critérios pedagógicos.

A Técnica de Inspeção Ergonômica de Software Educacional (TICESE) foi desenvolvida pelo LabiUtil – Laboratório de Utilizabilidade – Universidade Federal de Santa Catarina. Esta técnica resulta em um laudo técnico que serve de orientação para os responsáveis pela aquisição de material didático de programas de ensino (Gamez 1998). A técnica é dividida em três estágios – classificação, avaliação e contextualização. Para cada estágio, um conjunto de critérios diferentes é avaliado, e um peso é atribuído a cada critério, totalizando 17 critérios. Aplica-se peso 0 quando os critérios não se aplicam ao SE, peso 1 para os que são importantes, e 1,5 para os critérios muito importantes.

Para Silva (1998), a Técnica de Mucchielli tem como objetivo permitir a avaliação global de software considerando o público para o qual o SE foi desenvolvido. Esta abordagem utiliza como escala de mensuração um sistema de pontuação que varia de 1 (avaliação negativa) até 5 (avaliação positiva).

O instrumento de avaliação de software LORI (*Learning Object Review Instrument*) (Nesbit *et al.* 2002) desenvolvido pela *e-Learning Research and Assessment Network* é utilizado em alguns países para a avaliação de objetos de aprendizagem disponíveis na internet. O instrumento tem como objetivo avaliar a qualidade de um objeto de aprendizagem considerando apenas a parte técnica (adaptação aos recursos de hardware e software, e de usabilidade), baseado em nove critérios, de acordo com uma escala de pontuação que varia de 1 (avaliação negativa) até 5 (avaliação positiva).

A metodologia de Martins (Martins 2004) utiliza um método de inspeção (avaliação heurística) para realizar teste de usabilidade em interfaces de modo rápido e intuitivo. A avaliação foi desenvolvida para vários tipos de softwares, ou seja, a avaliação não tem o foco voltado apenas para softwares educacionais, mas abrange softwares de diferentes modalidades.

O método de Rocha (Rocha 1992) tem como objetivo avaliar a qualidade de Objetos de Aprendizagem e Softwares Educativos. Seus critérios são divididos em duas categorias: 7 critérios que avaliam os Objetos de qualidade (propriedades gerais que um SE deve ter para proporcionar qualidade do ponto de vista do usuário); e 16 critérios que avaliam os Fatores de qualidade do produto (qualidade de software). A ferramenta de avaliação utilizada para medir o grau de cada critério analisado é uma escala de 0 a 1, em que o critério pode ser avaliado como de alta qualidade (0.95 a 1), boa qualidade (0.90 a 0.94), qualidade mediana (0.60 a 0.89) ou sem qualidade (0.00 a 0.59).

O modelo de avaliação de Campos (Campos 1994) é voltado para avaliação da qualidade de SE a partir de um *checklist* técnico. O modelo propõe dez critérios de avaliação, e utiliza uma escala que varia de 0 a 1, adotando os números 0,25; 0,50 e 0,75 como grau de avaliação negativa ou positiva. A medida na escala é feita de forma que se a nota dada ao critério estiver perto do 0 é considerado negativa, e se estiver perto do 1 é positiva.

A diversidade de abordagens de avaliação de SE é uma realidade que pode tornar difícil a escolha sobre qual técnica adotar, uma vez que cada abordagem apresenta seu conjunto de critérios e seu instrumento de mensuração, assim como cada abordagem possui o seu nível de dificuldade de aplicação. Oliveira e Aguiar (2014) realizaram um estudo de análise comparativa entre um conjunto de abordagens de avaliação de SE. Este estudo comparativo contempla os métodos, metodologias e técnicas de avaliação citadas anteriormente, incluindo a Taxonomia de Bloom, que acabou sendo descartada neste trabalho, por ter sido considerada uma ferramenta inadequada para a avaliação de SE. Como resultado, os autores identificaram que: (a) existem critérios equivalentes entre abordagens distintas, embora a nomenclatura utilizada não seja a mesma (uso de sinônimos); e (b) nenhuma das abordagens contempla, igualmente, os três pilares de uma avaliação de SE: elementos pedagógicos, critérios de usabilidade e qualidade de *software*.

Sendo assim, os autores propuseram uma abordagem que combina critérios de vários métodos, técnicas e metodologias a fim de tornar a avaliação do SE mais abrangente. O presente trabalho tem por objetivo aplicar a abordagem combinada proposta em (Oliveira and Aguiar 2014) a fim de analisar sua aplicabilidade em termos de (i) abrangência dos critérios analisados em relação à tríade: elementos pedagógicos, critérios de usabilidade e qualidade de *software*; (ii) a equivalência entre os critérios adotados na abordagem mista; (iii) o esforço cognitivo e temporal para sua aplicação, e (iv) qualidade dos resultados alcançados.

O artigo está organizado em quatro seções, incluindo esta. Na seção 2, são descritas brevemente as técnicas aplicadas neste estudo, seus critérios e escalas de medição. Na seção 3, descreve-se a aplicação da abordagem combinada para avaliação de SE e apresentam-se os resultados alcançados. Por fim, na seção 4, estão as considerações finais, contemplando as limitações e contribuições desta pesquisa, assim como as possibilidades para continuá-la.

2. Métodos de Avaliação de Softwares Educacionais

De acordo com Frescki (2008), a expressão “avaliação de *software* educativo” consiste em analisar como um *software* pode ser aplicado na educação para ajudar o aluno na construção do conhecimento, e assim aperfeiçoar sua visão de mundo.

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). Além disso, os SEs devem ter suas funcionalidades corretas, de acordo com os preceitos da Engenharia de *Software* (Sommerville 2007), sendo importante ainda que a interatividade e a interface oferecidas pelo SE sejam adequadas ao público-alvo (Preece *et al.* 2013). Numa tentativa de agrupar melhor os critérios de avaliação de SE, Oliveira *et al.* (2001) propôs a definição de quatro categorias, a saber:

- i. *Interação aluno-SE-professor*: agrupa 21 critérios destinados a avaliar a qualidade e a importância da utilização de um dado recurso tecnológico dentro do ambiente escolar;
- ii. *Fundamentação pedagógica*: compreende 2 critérios para identificar quais foram as teorias e as opções pedagógicas selecionadas para o desenvolvimento do SE sob avaliação;
- iii. *Conteúdo*: une 12 critérios para identificar a completude e corretude do conteúdo considerando a área de conhecimento que está sendo trabalhada pelo SE; e, por fim,
- iv. *Programação*: cujos 25 critérios reunidos buscam verificar se o SE encontra-se de acordo com suas especificações (fidedignidade e integridade) e com os critérios de facilidade e flexibilidade para o contexto educacional.

Embora Oliveira *et al.* (2001) contemple um conjunto considerável de 60 critérios e o agrupamento adotado considere as quatro categorias citadas acima, não

existe uma definição clara dos critérios associados especificamente à usabilidade (qualidade de uso), encontrando-se estes diluídos entre as demais categorias. Os critérios relativos aos elementos pedagógicos estão presentes nas três primeiras categorias de forma não organizada, havendo uma identificação mais clara dos itens relacionados à qualidade de software (programação). Outro fator relevante a considerar consiste no fato do autor não indicar, explicitamente, como os critérios devem ser avaliados, não definindo, por exemplo, escalas de valores a serem associadas a cada critério. Em virtude destas características, esta proposta de organização de critérios para avaliação de SE pode ser entendida como uma abstração a ser instanciada por métodos, metodologias, e técnicas de avaliação de SE.

Neste sentido, existe uma grande variedade de técnicas para avaliação de SE descritas na literatura, e cada uma delas possui singularidades em relação (i) ao número de critérios; (ii) à natureza dos critérios (aspectos pedagógicos, qualidade de software, qualidade de uso); (iii) à natureza e granularidade das escalas adotadas na avaliação dos critérios e (iv) à qualidade dos resultados alcançados. Dentre as existentes, serão consideradas no escopo deste trabalho as seguintes abordagens, cuja caracterização está resumida nos Quadros 1 e 2: Método de Reeves; Técnica de TICESE; Técnica de Mucchielli; Avaliação de LORI; Metodologia de Martins, Método Rocha e Modelo de Avaliação de Campos. Os campos vazios correspondem a “*Não se Aplica*”, e a numeração representa a quantidade de critérios de cada técnica.

Quadro 1 - Abordagens para Avaliação de SE

		Reeves	Mucchielli	Martins	LORI	TICESE	Campos	Rocha
Aspectos de avaliação	Pedagógico	14			1	1		
	Qualidade de software	10	10	4	4	6	9	15
	Qualidade de uso			6	4	9	1	1
	Total	24	10	10	9	16	10	16

Quadro 2 - Caracterização das estratégias de avaliação por Abordagem

	Forma de avaliação	Valor mínimo	Valor máximo	Escala
Reeves	Escala bidirecional não numérica	Esquerda: negativo	Direita: positivo	
Mucchielli	Escala numérica de 5 pontos	1 = negativo máximo	5 = positivo máximo	(1, 2, 3, 4, 5)
Martins		1 = mais difícil	5 = mais fácil	(1, 2, 3, 4, 5)
LORI		1 = negativo máximo	5 = positivo máximo	(1, 2, 3, 4, 5)
TICESE	Escala numérica de 3 pontos	0 = não aplicável	1.5 = muito importante	(0, 1, 1.5)
Campos	Escala numérica de 5 pontos	0 = negativo máximo	1 = positivo máximo	(0; 0,25; 0,50; 0,75; 1)
Rocha	Escala numérica de 4 pontos	0.00 a 0.59 sem qualidade	0.95 a 1 alta qualidade	(0.00 a 0.59; 0.60 a 0.89; 0.90 a 0.94; 0.95 a 1)

Realizando uma análise inicial das abordagens e de seus critérios, ao comparar com a proposta de Oliveira *et al.* (2001), é possível perceber que existe uma diferença significativa em relação à quantidade de critérios definidos por este (60) e pelas abordagens, sendo o Método de Reeves a que apresenta maior quantidade (24). Além disso, ao observar a distribuição dos critérios nos aspectos de avaliação, apenas duas abordagens (LORI e TICESE) contemplam critérios na tríade: elementos pedagógicos, qualidade de uso e de *software*, embora a quantidade de critérios seja baixa (9 e 16, respectivamente). Essas observações iniciais incitam o questionamento sobre a possibilidade de utilizar estas abordagens de forma combinada, aumentando o número de critérios e a abrangência do resultado em relação aos aspectos de avaliação.

Neste sentido, um estudo aprofundado sobre a completude das abordagens em relação à proposta de (Oliveira *et al.* 2001) foi realizada e descrita em (Oliveira and Aguiar 2014). Foi realizado um mapeamento dos critérios destas abordagens em relação aos critérios abstratos de (Oliveira *et al.* 2001), a fim de identificar a completude das abordagens, assim como a equivalência entre os critérios destas. Como resultado da análise, foi possível identificar: i) a pobreza de critérios relacionados aos aspectos

pedagógicos, comparado a critérios de qualidade de uso (usabilidade) e qualidade de *software* (programação) – exceto em Reeves; ii) a ausência de uma abordagem que, sozinha, cobrisse um conjunto abrangente e significativo de critérios. Com estas duas percepções, Oliveira e Aguiar (2014) constataram que os resultados alcançados na avaliação de um SE a partir da aplicação destas abordagens poderiam ser insuficientes, podendo prejudicar a tomada de decisão sobre a adoção de um SE. A fim de minimizar este problema, Oliveira e Aguiar (2014) propuseram uma abordagem para avaliação de SE que combina o conjunto total de critérios mapeados no estudo. Foi utilizada como base a classificação de Oliveira *et al.* (2001) para agrupar a combinação de critérios por abordagem, e os operadores lógicos AND (indica o uso de ambas as abordagens) e OR (indica a escolha de apenas uma das abordagens) para evidenciar a equivalência entre critérios:

- *Interação aluno-SE-professor: REEVES and MARTINS and TICESE;*
- *Fundamentação pedagógica: REEVES or MUCCHIELLI;*
- *Conteúdo: TICESE and LORI and MARTINS;*
- *Programação: ROCHA and (TICESE or MARTINS) and LORI and REEVES and CAMPOS.*

Embora a proposta de (Oliveira *and* Aguiar 2014) seja bastante interessante, os autores não apresentam uma validação da abordagem combinada, a partir de sua aplicação na avaliação de um SE real. Sendo assim, não existe informação sobre a viabilidade da proposta, ou das limitações impostas pela associação de abordagens com critérios e escalas de avaliação distintas, ou ainda sobre a qualidade dos resultados obtidos. Portanto, o objetivo do presente trabalho consiste em analisar estes aspectos diante da avaliação de um SE amplamente utilizado para o estudo de idiomas, o Duolingo.

3. Aplicação da abordagem combinada para avaliação do Duolingo

O Duolingo é um aplicativo para estudo de idiomas (inglês, francês, espanhol e alemão), disponível nas versões *web* e *mobile*, e com usuários de diferentes nacionalidades. Neste SE, o conteúdo é agrupado em níveis (iniciante, intermediário e avançado), com temas de estudo (saudações, comidas, animais, família, etc). Os temas são organizados em lições e atividades. O nível de dificuldade das lições aumenta à medida que o aluno progride. Foram avaliadas as dez primeiras lições disponíveis na versão *web* do Duolingo para ensino-aprendizagem de inglês.

A avaliação do software foi feita por duas duplas de alunos, todos discentes do curso de Licenciatura em Ciência da Computação da Universidade Federal da Paraíba – Campus IV, Rio Tinto, sob a supervisão de duas professoras especialistas em qualidade de Software e Interação Homem-Computador. As duplas iniciaram a avaliação após um estudo bibliográfico sobre os métodos de avaliação presentes na combinação de abordagens proposta, e a aplicação de algumas delas em atividades da disciplina *Softwares Educativos*, revezando-se na aplicação das técnicas. Num primeiro momento, o SE foi avaliado considerando cada técnica isoladamente, com seus critérios e seus respectivos instrumentos de mensuração. Os resultados foram registrados em relatórios, contendo também detalhes da avaliação como: tempo e dificuldade de compreensão da abordagem, tempo e esforço cognitivo exigido no momento da aplicação da abordagem. Em seguida, iniciou-se o mapeamento entre os critérios propostos em (Oliveira *et al.* 2001) e a combinação de abordagens proposta por (Oliveira and Aguiar 2014), produzindo tabelas únicas que concentravam todas as informações das avaliações. O passo seguinte consistiu em acrescentar em cada tabela os resultados das avaliações, a fim de identificar equivalências e fragilidades. A organização tabular da informação considerou as “fórmulas” para combinação das abordagens. Nas subseções seguintes, são apresentados os resultados obtidos com a avaliação a partir da abordagem combinada para a versão *web* do Duolingo. Com o intuito de facilitar a compreensão dos resultados, estes estão apresentados de acordo com a categorização proposta em (Oliveira *et al.* 2001).

3.1. Interação Aluno-SE-Professor = REEVES and MARTINS and TICESE

Observando a Figura 1, é possível perceber que muitos critérios são cobertos por apenas uma das técnicas, o que justifica a necessidade de combiná-las para se ter um resultado abrangente. Entretanto, ainda há critérios (2, 4, 11, 15, 16, 19, 20 e 21) que não são cobertos por nenhuma das técnicas, deixando lacunas na avaliação. Além disso, os resultados obtidos em técnicas distintas nem sempre foram compatíveis, como fica evidente para o critério 9: mal avaliado em REEVES, e bem avaliado em MARTINS.

Figura 1 - Avaliação da Interação Aluno-SE-Professor para o Duolingo Web.

		AND		
		Reeves	TICESE	Martins
1	Instruções claras e objetivas;	Facilidade de uso ←————→		
2	Ícones e botões para interação;			
3	Auxílio e dicas para esclarecimento de dúvidas;	Facilidade de uso ←————→	Significado dos códigos (82%)	
4	Linguagem (vocabulário) direcionada ao público-alvo;			
5	Universalidade da linguagem para favorecer o uso por um amplo público;		Adaptabilidade (59,5%)	
6	Estrutura de navegação não linear;	Navegação ←————→		
7	Navegabilidade que dê acesso a todas as partes do SE;	Navegação ←————→		
8	Mapeamento para esclarecimento sobre percursos percorridos;	Mapeamento ←————→		Visibilidade do sistema (5)
9	Memória do SE para localizar a posição do usuário dada uma interrupção.	Sequenciamento instrucional ←————→		Reconhecimento em lugar de lembrança (5)
10	Atratividade para despertar o interesse de uso;	Motivação ←————→	Motivação e compreensão dos conteúdos (84%)	
11	Desafios pedagógicos para manter nível de interesse;			
12	Interação com o usuário que estimule a participação e auxilie na superação de conflitos cognitivos	Atividade do usuário ←————→		Flexibilidade e eficiência do uso (5)
13	Layout das telas com visual esteticamente adequado;	Design de tela ←————→		
14	Carga cognitiva reduzida com o uso adequado de imagens, textos e animações.	Carga cognitiva ←————→		
15	Adequação dos recursos de hipertexto às atividades pedagógicas, em quantidade e qualidade;			
16	Adequação dos recursos de imagem e animação às atividades pedagógicas, em quantidade e qualidade			
17	Interação intragrupo	Aprendizado cooperativo ←————→		
18	Interação intergrupos	Aprendizado cooperativo ←————→		
19	Interação transgrupos			
20	Orientação didático-pedagógica			
21	Inclusão de múltiplos recursos			

3.2. Fundamentos Pedagógicos = REEVES or MUCCHIELLI

Para fundamentação pedagógica, houve discrepâncias dos resultados obtidos com Reeves (avaliação negativa) em comparação com Mucchielli (avaliação positivas), como pode ser visto na Figura 2. Este conflito de resultados é preocupante, uma vez que as abordagens podem ser utilizadas de forma alternada (devido ao operador lógico OR). A não equivalência dos resultados entre elas pode influenciar a avaliação feita, e induzir o avaliador na tomada de decisão sobre a adoção do SE.

Figura 2 - Avaliação dos Fundamentos Pedagógicos para o Duolingo Web.

		OR	
		Reeves	Mucchielli
22	Fundamentos pedagógicos que embasam o SE, opção pedagógica utilizada no desenvolvimento do SE	Filosofia pedagógica ↔	Qualidade do modelo pedagógico (5)
23	Consistência pedagógica, escolha da teoria pedagógica utilizada no SE.	Filosofia pedagógica ↔	Qualidade do modelo pedagógico (5)

3.3. Conteúdo = TICESE and LORI and MARTINS

Embora seja proposta a combinação de três técnicas para avaliação do conteúdo, ainda há quatro critérios (25, 31, 34, 35) que não puderam ser avaliados por nenhuma delas. Além disso, LORI compreende apenas dois critérios, e ambos são mapeados por TICESE, levando a uma reflexão sobre a real necessidade de seu uso. No caso de Martins, são apenas dois critérios, porém um deles (30) não é avaliado pelas outras duas técnicas. Ainda assim, num universo dos 12 critérios destacados pela Figura 3, cabe a reflexão se o uso combinado dessas técnicas seria realmente necessário, ou se apenas TICESE seria suficiente para avaliar a qualidade do conteúdo do SE.

Figura 3 - Avaliação do Conteúdo para o Duolingo Web.

		AND		
		LORI	TICESE	Martins
24	Adequação do SE ao conteúdo	Qualidade do conteúdo (5)	Carga de trabalho (93%)	
25	Utilização do SE como ferramenta didática			
26	Correção do conteúdo		Condução (83%)	
27	Correção da lógica do conteúdo		Recursos de avaliação do aprendizado (94%)	
28	Correção da representação do conteúdo		Condução (83%)	
29	Correção das simplificações do conteúdo		Recursos de avaliação do aprendizado (94%)	
30	Atualidade do conteúdo			Compatibilidade entre o sistema e o mundo real (5)
31	Atualidade da metodologia			
32	Adequação do conteúdo ao público-alvo	Feedback e Adaptação (5)	Motivação e compreensão dos conteúdos (84%)	Liberdade e controle do usuário (4)
33	Adequação do conteúdo ao currículo escolar		Adequabilidade (92%)	
34	Indicação dos conhecimentos prévios			
35	Retrabalho com os conhecimentos prévios			

3.4. Programação = ROCHA and (TICESE or MARTINS) and LORI and REEVES and CAMPOS

A avaliação dos aspectos de programação de um Software Educativo é bastante complexa, uma vez que os critérios utilizados são transversais à aplicação e requerem não apenas uma combinação de abordagens, mas também um conhecimento técnico aprofundado. Entretanto, é comum que os avaliadores não tenham acesso aos artefatos de código dos aplicativos, e por isso sejam incapazes de avaliá-los completamente.

As abordagens de Rocha e Campos focam especialmente em critérios de qualidade não-funcionais, que demandariam uma análise detalhada do código da aplicação, e a realização de testes complexos. O fato do código-fonte do Duolingo não ser disponibilizado impediu que os critérios de programação dessas abordagens fossem avaliados. Por isso, na Figura 4, esses itens aparecem indicados apenas para evidenciar a correspondência (ou o mapeamento) de critérios, mas sem valores nas escalas

numéricas, pois não foi possível apreciá-los. Esses itens estão destacados com um contorno diferenciado.

Outra informação relevante consiste no fato de alguns dos critérios serem cobertos por apenas uma das técnicas usadas: o critério 40 é abordado apenas por Reeves; o critério 51 só tem correspondente em Campos; os critérios 39, 41, 49 e 53 são mapeados somente em Rocha; e os critérios 59 e 60 estão presentes apenas em LORI. Sendo assim, a análise completa desses aspectos demanda um avaliador que conheça todas as técnicas e suas escalas de avaliação. Esse excesso de demandas de conhecimento para um uso muito pontual interfere na sua curva de aprendizado de um avaliador que deseje aplicar essa abordagem combinada.

Além disso, é importante perceber que a alternância entre as abordagens de TICESE e Martins (pelo operador lógico OR) leva necessariamente à não verificação de um critério. Isso porque, caso a opção do avaliador seja por TICESE, o item 55 (Modularidade) não será avaliado. Caso sua preferência seja por Martins, o critério 36 (Correção) não será considerado. Mesmo com uma abrangente combinação de técnicas, é possível perceber que os critérios 37 e 44 não puderam ser mapeados para nenhuma das abordagens, evidenciando falhas de cobertura nessa proposta.

Figura 4 - Avaliação da Programação para o Duolingo Web.

		AND				OR	
		Rocha	LORI	Reeves	Campos	TICESE	Martins
36	Correção					Gestão de erros (66%)	
37	Atualidade						
38	Precisão	Eficiência e Precisão		Objetividade ←→	Operacionalidade		
39	Completeza	Completeza					
40	Simplicidade			Apres. da Informação ←→			
41	Concisão	Concisão					
42	Robustez	Robustez			Robustez		Recup. de erros (2)
43	Segurança	Segurança			Integridade	Gestão de erros (66%)	Prev. contra erros (4)
44	Clareza						
45	Estrutura	Estrutura		Estruturação ←→			
46	Rastreabilidade	Rastreabilidade		Mapeamento ←→			
47	Alterabilidade	Manutenabilidade		Validade Experimental ←→			
48	Compatibilidade	Operacionalidade			Adequação ao ambiente	Compatibilidade (100%)	Flexibilidade e eficiência do uso (5)
49	Oportunidade	Oportunidade					
50	Reutilizabilidade	Reutilizabilidade	Reusabilidade (5)				
51	Economia do processamento				Eficiência		
52	Rentabilidade	Rentabilidade			Rentabilidade		
53	Verificabilidade	Avaliabilidade					
54	Validabilidade	Valiabilidade		Valoração do Erro ←→			
55	Modulandade						Consistência a padrões (5)
56	Manual Técnico			Funcionalidade geral ←→		Organização da documentação impressa (100%)	Ajuda e documentação (4)
57	Guia de apoio pedagógico					Organização da documentação impressa (100%)	Ajuda e documentação (4)
58	Manual do aluno			Funcionalidade geral ←→		Organização da documentação impressa (100%)	Ajuda e documentação (4)
59	Informações de capa		Projeto de apresentação (3)				
60	Apresentação da capa do SE		Projeto de apresentação (3)				

4. Considerações finais

Diante da necessidade de avaliar um *software* educativo, o educador se depara com uma grande variedade de abordagens, cada uma com critérios próprios e notações diferenciadas, sendo difícil escolher qual deverá ser usada. Além disso, algumas abordagens possuem critérios muito subjetivos e com alto nível de abstração, não apresentando explicações ou exemplos sobre como estes devem ser avaliados. Como consequência, os resultados alcançados em uma avaliação ficam fortemente vinculados ao avaliador que aplicou a abordagem. De forma complementar, as avaliações de SE devem contemplar critérios da tríade elementos pedagógicos, qualidade de uso (usabilidade) e qualidade de *software* (desenvolvimento e programação). A multidisciplinaridade envolvida no processo de avaliação também exige do avaliador uma gama de conhecimentos particulares – deixando explícita a necessidade de *expertise* e *know-how* – impactando na curva de aprendizado para realização das avaliações. Além disso, as abordagens para avaliação de SE não contemplam, sozinhas, os três pilares que consolidam a qualidade de um *software* educativo. Sendo assim, é importante considerar a adoção combinada de mais de uma técnica para avaliação de SE.

A fim de tornar a avaliação de SE mais abrangente, uma abordagem combinada de vários métodos, técnicas e metodologias foi elaborada por (Oliveira and Aguiar 2014). Para análise da viabilidade de avaliar um SE a partir desta proposição, realizou-se, neste trabalho, a sua aplicação para o *software* para aprendizado de idiomas Duolingo, amplamente utilizado mundialmente. Os resultados detalhados foram apresentados na seção anterior, mas, de forma geral, foi possível identificar as seguintes fragilidades da proposta: (i) Embora se constate uma maior abrangência dos critérios em relação à tríade, quando comparada com as abordagens individuais, ainda é possível perceber uma predominância dos critérios de qualidade de uso (usabilidade). Além disso, os critérios relativos à qualidade de *software* dependem do acesso ao código-fonte do SE, informação que nem sempre está disponível. Ainda que esteja disponível, a avaliação dos critérios não é trivial, exigindo do avaliador conhecimento detalhado

sobre programação e engenharia de *software*; (ii) Considerando a equivalência entre os critérios adotados na abordagem mista (uso do OR), é importante destacar que os resultados obtidos por abordagens diferentes para os mesmos critérios não foram compatíveis (Figura 1 e Figura 2). Este fato pode levar a uma reflexão sobre o uso alternado desses critérios, ou ainda, levantar um questionamento sobre o nível de abstração necessário para a avaliação destes critérios – o que poderia ter levado os avaliadores a resultados opostos; (iii) O esforço cognitivo para sua aplicação é demasiado, uma vez que os avaliadores se deparam com um conjunto relativamente grande de critérios a serem avaliados, e eles são mensurados a partir de estratégias diferentes (escalas e valores). A alternância constante entre as abordagens leva ao cansaço mental dos avaliadores. Além disso, a ausência de uma ferramenta que auxilie no processo de avaliação torna esta tarefa mais demorada, exigindo um alto grau de concentração por parte de quem a aplica; e, por fim, (iv) A qualidade dos resultados alcançados se restringe ao mapeamento dos critérios associados aos valores obtidos na avaliação (Figuras 1, 2, 3 e 4). A análise deste conjunto de informações não favorece a percepção dos avaliadores sobre quais são suas principais deficiências, não auxiliando na tomada de decisão sobre a adoção ou não do SE.

Tendo compreensão da validade da proposta de abordagens combinadas para avaliação de SE, e conhecendo suas atuais limitações, acredita-se que o ponto de partida para auxiliar os avaliadores de SE, em sua maioria docentes, que atuam nas mais diversas áreas de conhecimento, consiste na padronização dos critérios em termos de instrumentos de mensuração. Em seguida, faz-se necessário tornar a aplicação da abordagem mais exata (menos dependente do conhecimento do avaliador) e menos custosa (tempo e esforço). Estes dois objetivos podem ser alcançados tendo como apoio a padronização dos critérios-base para permitir o desenvolvimento de uma ferramenta que guie o processo de avaliação de SE. Com o auxílio de uma ferramenta computacional, seria possível, a partir da caracterização do SE (*mobile, web, desktop*), dos seus propósitos pedagógicos (EaD, simulação, resolução de exercícios, etc), e do seu uso (jogo, uso colaborativo, etc.) sugerir critérios a serem analisados. Os critérios

poderiam ser mensurados de acordo com uma escala comum e os resultados poderiam ser exibidos para os avaliadores considerando a proporção de critérios satisfeitos, e alertas sobre aspectos relevantes para o contexto geral de SE que não sejam contemplados pelo SE em análise.

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.

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. p.110, 1989.

CAMPOS, G. H. B. **Metodologia para avaliação da qualidade de software educacional. Diretrizes para desenvolvimento e usuários**. Rio de Janeiro, 1994.

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*. Vol. 1, No. 1, p. 545-554, 2003.

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, Cascavel, 2008.

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

MARTINS, M. de L. O. **O papel da usabilidade no ensino a distância mediado por computador. Dissertação de Mestrado**. Centro Federal de Educação tecnológica de Minas Gerais. Minas Gerais, 2004.

MORAIS, R. X. T. **Software educacional: a importância de sua avaliação e do seu uso nas salas de aula**. Monografia (Bacharel em ciências da computação) - Faculdade Lourenço Filho. 51p. Fortaleza, 2003.

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

Revista Tecnologias na Educação – Ano 8 – Número/Vol.16 – Edição Temática – Congresso Regional sobre Tecnologias na Educação (Ctrl+E 2016) – Setembro2016 – tecnologiasnaeducacao.pro.br

OLIVEIRA B JUNIOR, O.; AGUIAR, Y. P. C. Análise de abordagens objetivas para avaliação de softwares educativos. *In: Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems*. Sociedade Brasileira de Computação, 2014).

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, Em Aberto, n. 57, v.12, 1992.

SANCHO, J. **Para uma tecnologia educacional**. Porto Alegre: ArtMed, 1998.

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. 113p. Florianópolis, 1998.

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