

Taxonomia para Avaliação de Recursos Digitais de Aprendizagem: Proposição e Análise da Adequação por Especialistas

Ozonias de O. B. Junior¹, Yuska P. C. Aguiar², Hermano P. de Moura³

¹³Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Recife, PE – Brasil

²Departamento de Ciências Exatas – Universidade Federal da Paraíba (UFPB) – Rio Tinto, PB - Brasil

oobj@cin.ufpe.br, yuska@dcx.ufpb.br, hermano@cin.ufpe.br

Abstract. *The diversity and plurality of approaches to the assessment of Digital Learning Resources is a reality. The non-existence of a consensus about the criteria to be adopted, nor an explicit relationship between these and the quality models, they are barriers to the evaluation process itself. To mitigate these problems, the Taxonomy for Assessment of DLR is proposed from the compilation of 26 national and international approaches. This organizes and conceptualizes 49 criteria in four dimensions: software quality, quality of use, pedagogical quality and hybrid quality. To analyze the adequacy of the taxonomy, online surveys and interviews were carried out. 32 Software Engineers, 45 Usability experts and 4 Educators participated.*

Resumo. *A diversidade e pluralidade de abordagens para a avaliação de Recursos Digitais de Aprendizagem é uma realidade. Não existir um consenso sobre os critérios a serem adotados, nem uma relação explícita entre estes e os arcabouços de qualidade, são barreiras para o processo de avaliação em si. Para atenuar esses problemas a Taxonomia para Avaliação de RDA é proposta a partir da compilação de 26 abordagens nacionais e internacionais. Esta organiza e conceitua 49 critérios em 4 dimensões: qualidade de software, de uso, pedagógica e híbrida. Para análise da adequação da taxonomia, foram realizados survey on-line e entrevistas. Participaram 32 Engenheiros de Software, 45 especialistas em Usabilidade e 4 Educadores.*

1 Introdução

Nos últimos anos, são notórios os esforços e o interesse pela pesquisa e desenvolvimento de recursos tecnológicos voltados para apoiar os processos de ensino e aprendizagem. Em especial, destacam-se as ferramentas computacionais, como os Recursos Digitais de Aprendizagem¹ (RDAs), que oferecem ambientes interativos, capazes de atrair a atenção de seus usuários, aumentando o interesse e favorecendo a aprendizagem [Guarda & Goulart, 2018].

O RDA é um dispositivo (ou recurso) que foi planejado e elaborado para fins didático-pedagógicos, representando um mediador de conteúdos e viabilizador do processo de ensino e aprendizagem, podendo auxiliar o educador como um recurso a

¹ Recursos Digitais de Aprendizagem, *Softwares* Educativos, *Softwares* Educacionais e Jogos Educacionais Digitais serão tratados como sinônimos no contexto deste trabalho.

mais em sua prática educativa [Medeiros, 2019]. De acordo com Mesquita (2018), a utilização de RDAs deve propiciar o aprendizado lúdico do estudante, estimular o raciocínio lógico e o pensamento crítico, possibilitar aprender com erros, reduzir a passividade, motivar e desenvolver hábitos de persistência na transposição de desafios e desenvolvimento de tarefas.

No entanto, a adoção de um RDA requer cuidados, pois vários fatores podem interferir nos objetivos pretendidos para seu uso – o aprendizado, por exemplo: inadequação do RDA ao contexto educacional ou ao conteúdo abordado [Lin & Oliveira, 2018], má estruturação e/ou representação didático-pedagógica e problemas de funcionamento (*bugs*) [França & Silva, 2014], outros. Para que os benefícios dos RDAs sejam alcançados, faz-se necessário que estes: (i) estejam adequados pedagogicamente; (ii) tenham qualidade na perspectiva da Engenharia de *Software* e (iii) que seu projeto de interface e interação estejam alinhados com recomendações da Engenharia da Usabilidade.

Nessa perspectiva, a literatura nacional e internacional apresenta uma gama de abordagens (métodos, técnicas, instrumentos etc.) para avaliação da qualidade de RDAs. No entanto, a diversidade de abordagens existentes implica em, como destacado por Pereira et al. (2016): (i) ausência de consenso ou completude de critérios; (ii) subjetividade e nível de abstração dos critérios, que leva ao fato; (iii) de os resultados serem fortemente vinculados ao avaliador; (iv) as dimensões de qualidade pedagógica, de usabilidade e de *software*, não são contempladas na mesma proporção.

Neste sentido, como primeira tentativa de organização das abordagens existentes para a avaliação de RDAs foi proposta uma versão inicial da taxonomia considerando o corpo de publicação nacional. No entanto, o resultado parcial desta versão indicou a necessidade de 1) ampliar a análise para contemplar a literatura internacional sobre abordagens para avaliação de RDAs e 2) a importância de ampliar a análise da adequação pelos especialistas da tríade de qualidade.

Este artigo apresenta a evolução da proposição anterior [Brito Junior & Aguiar, 2018] para contemplar os critérios de avaliação propostos na literatura internacional, assim como a ampliação do processo de análise de sua adequação junto aos especialistas em Engenharia de *Software* e de Usabilidade, assim como Educadores. Como resultado, tem-se a Taxonomia para Avaliação de Recursos Digitais de Aprendizagem (TARDA²), que foi concebida a partir da combinação de um conjunto de 26 abordagens (12 internacionais), previamente identificadas em revisão sistemática da literatura. Com objetivo de diminuir a curva de aprendizagem necessária para a avaliação de RDA, TARDA se apoia nas nomenclaturas e conceitos amplamente utilizados na Engenharia de *Software*, na Engenharia de Usabilidade e na Pedagogia a partir de modelos de qualidade, normas, dicionários e heurísticas adotadas nessas áreas de conhecimento [Brito Junior & Aguiar, 2016]. Como resultado, TARDA tem 4 Dimensões, que são compostas por 17 Categorias, que agrupam 49 Crítérios de Avaliação, associados à

² De acordo com a literatura analisada, o termo mais adequado é Recursos Digitais de Aprendizagem (RDAs) e não *Software* Educativo (SE); a nomenclatura da taxonomia foi adaptada para inserção do novo termo.

respectiva Descrição. Para análise de sua adequação, dois *surveys on-line* foram respondidos por 32 Engenheiros de *Software* e 45 especialistas em Usabilidade; assim como 4 Educadores participaram de entrevistas semiestruturadas.

2 Metodologia da Pesquisa

Bayona-Oré et al. (2014) propõem 4 etapas para concepção de taxonomias, que foram instanciadas para a Taxonomia para Avaliação de Recursos Digitais de Aprendizagem (TARDA) em dois ciclos. As etapas são: **Etapa 1** Planejamento: identificação do objetivo da taxonomia, sua importância, o escopo e as atividades a serem executadas para sua elaboração; **Etapa 2** Identificação e extração de informações: busca de artigos em fontes científicas (conferências, simpósios, *journals*) e caracterização das abordagens selecionadas (critérios contemplados, instrumentos para coleta de dados, dimensões e categorias); **Etapa 3** Projeto e construção da taxonomia: cujos critérios foram mapeados de acordo com os modelos, normas, dicionários e heurísticas amplamente difundidos na Engenharia de *Software*, na Engenharia de Usabilidade e na Pedagogia, e **Etapa 4** Teste e Validação: consulta aos especialistas nas dimensões de qualidade para a obtenção de *feedbacks* sobre consistência, integridade e ambiguidade da taxonomia.

O primeiro ciclo executado contemplou a literatura nacional para extração das informações sobre critérios de avaliação de RDAs e propondo uma primeira versão da taxonomia [Brito Junior & Aguiar, 2018] e sondagem inicial de sua adequação para duas das 4 dimensões existentes. O segundo ciclo, escopo deste artigo, ampliou a base de consulta para a literatura internacional a partir de um processo de *snowball* [Wohlin, 2014] e uma revisão sistemática da literatura [Cresswell, 2013], assim como aprofundou a validação da taxonomia. Este ciclo foi conduzido por 3 pesquisadores da área de Ciência da Computação (2 professores doutores e 1 doutorando) que atuam nas áreas de Interação Humano-Computador, Qualidade de *Software* e de Uso (Usabilidade), experientes na avaliação de Recursos Digitais de Aprendizagem.

Como resultado, 26 novas abordagens foram acrescentadas na análise para o refinamento da taxonomia. O total de 724 critérios (incluindo ambos os ciclos) foram extraídos da totalidade de 26 abordagens analisadas a partir do *corpus* geral de 172 publicações nacionais e internacionais. O refinamento resultou na inclusão da dimensão de Qualidade Híbrida (QH), cujos critérios pertencem a mais de uma das dimensões já existentes na versão inicial da taxonomia (Qualidade de *Software*: QS, Qualidade de Uso: QU e Qualidade Pedagógica: QP). Para análise da taxonomia refinada, buscou-se a opinião de especialistas (potenciais avaliadores de RDAs) sobre a adequação da TARDA. Para tanto, duas estratégias foram adotadas: aplicação de questionários *on-line* e entrevistas semiestruturadas.

Os questionários³ *on-line* foram utilizados para avaliação das dimensões de Qualidade de *Software*, de Uso e Híbrida (para os critérios que unem as duas dimensões citadas). A estrutura dos questionários foi de 3 seções: (i) termo de consentimento; (ii)

³ Questionários para avaliação dos critérios de qualidade de *software* (<https://forms.gle/CdWebCiB5Y9rtV786>) e de qualidade de uso (<https://forms.gle/WAK9juBTbkLF15cf9>) da TARDA.

perfil do respondente; e, (iii) avaliação da adequação da taxonomia com a apresentação do critério/descrição. Para cada critério da taxonomia o respondente deveria indicar concordância ou discordância, sendo possível, para o segundo caso, indicar sugestões de mudanças. Participaram 32 Engenheiros de *Software* e 45 especialistas em Usabilidade.

As entrevistas semiestruturadas foram adotadas para a dimensão de Qualidade Pedagógica, com a participação de especialistas em educação através de reuniões virtuais. Esta estratégia foi adotada considerando a riqueza de informações que podem ser discutidas e coletadas – essencial para compreensão de conceitos e respectivas descrições – como uma forma de balancear com o fato de a proposição da taxonomia ter sido conduzida por especialistas que não dominam a dimensão pedagógica. Foram realizadas 3 reuniões de 90 minutos (20/06, 04/07, 06/07 de 2020), com a participação de 4 educadores (1 especialista, 1 doutorando, 2 doutores). Os objetivos das reuniões foram: (i) análise das nomenclaturas das categorias e disposição delas na taxonomia, (ii) análise das nomenclaturas dos critérios e sua alocação dentro de cada categoria, e (iii) análise da descrição de cada critério (em dois momentos).

3 Taxonomia para Avaliação de Recursos Digitais de Aprendizagem (TARDA)

A Taxonomia para Avaliação de Recursos Digitais de Aprendizagem (TARDA) está organizada em 4 Dimensões, que são compostas por 17 Categorias, que agrupam 49 Critérios de Avaliação, associados à respectiva Descrição. A nomenclatura adotada e as definições foram mapeadas de acordo com aquelas amplamente utilizadas na literatura nacional e internacional para Qualidade de *Software*, de Uso e Pedagógica – tendo em vista proporcionar uma linguagem comum para conceitos já conhecidos da comunidade científica, diminuindo a curva de aprendizado para aplicação da TARDA.

Concepção do Software	Evolução do Software	Segurança do Sistema
<p>Adequação: Capacidade de prover um conjunto apropriado de funções para tarefas e objetivos dos usuários especificados.</p> <p>Acurácia: Capacidade de prover, com o grau de precisão necessário, resultados ou efeitos corretos ou conforme acordados.</p> <p>Conformidade Funcional: Capacidade de estar de acordo com normas, convenções ou regulamentações previstas em leis e prescrições similares relacionadas à funcionalidade.</p> <p>Recursos: Capacidade de usar tipos e quantidades apropriados de recursos, quando o <i>software</i> executa suas funções sob condições estabelecidas.</p> <p>Conformidade Portátil: Capacidade de estar de acordo com normas ou convenções relacionadas à portabilidade.</p> <p>Interoperabilidade: Capacidade de interagir com um ou mais sistemas especificados.</p>	<p>Estabilidade: Capacidade de evitar efeitos inesperados decorrentes de modificações no <i>software</i>.</p> <p>Analísabilidade: Capacidade de permitir o diagnóstico de deficiências ou causas de falhas no <i>software</i>, ou a identificação de partes a serem modificadas.</p> <p>Modificabilidade: Capacidade de permitir que uma modificação especificada seja implementada</p> <p>Substituibilidade: Capacidade de ser usado em substituição a outro produto de <i>software</i> especificado, com o mesmo propósito e no mesmo ambiente.</p> <p>Instalabilidade: Capacidade para ser instalado em um ambiente especificado.</p> <p>Testabilidade: Capacidade de permitir que o <i>software</i>, quando modificado, seja validado.</p> <p>Maturidade: Capacidade de evitar falhas decorrentes de defeitos no <i>software</i>.</p>	<p>Inteligibilidade: Capacidade de possibilitar ao usuário compreender se o <i>software</i> é apropriado e como ele pode ser usado para tarefas e condições de uso específicas.</p> <p>Apreensibilidade: Capacidade do produto de <i>software</i> possibilitar ao usuário entender como ele funciona e/ou aprender a usá-lo.</p> <p>Tolerância à Falhas: Capacidade de evitar falhas e manter um nível de desempenho especificado em casos de defeitos no <i>software</i> ou de violação de sua interface especificada.</p> <p>Segurança de Acesso: Capacidade de proteger informações e dados, de forma que pessoas ou sistemas não autorizados não possam lê-los nem modificá-los e que não seja negado o acesso às pessoas ou sistemas autorizados.</p>

Figura 1. Dimensão de Qualidade de *Software* da TARDA

Nas Figuras 1, 2, 3 e 4, é possível conhecer os detalhes para as dimensões de Qualidade de *Software* (3 categorias e 17 critérios), Qualidade de Uso (4 categorias e 10 critérios), Qualidade Pedagógica (6 categorias e 15 critérios) e Qualidade Híbrida (4 categorias e 7 critérios), respectivamente, onde, as categorias têm contraste de cor de fundo, os critérios estão em negrito, seguidos pelas descrições. A evolução da taxonomia anterior para a atual (TARDA) versou sobre: a modificação na nomenclatura da mesma; a organização em dimensões e categorias; a quantidade de critérios em cada

dimensão e categoria (menos critérios); as nomenclaturas dos critérios de acordo com as análises realizadas pelos especialistas; a inserção das dimensões de qualidade pedagógica e híbrida e a ampliação do quantitativo de critérios para avaliação.

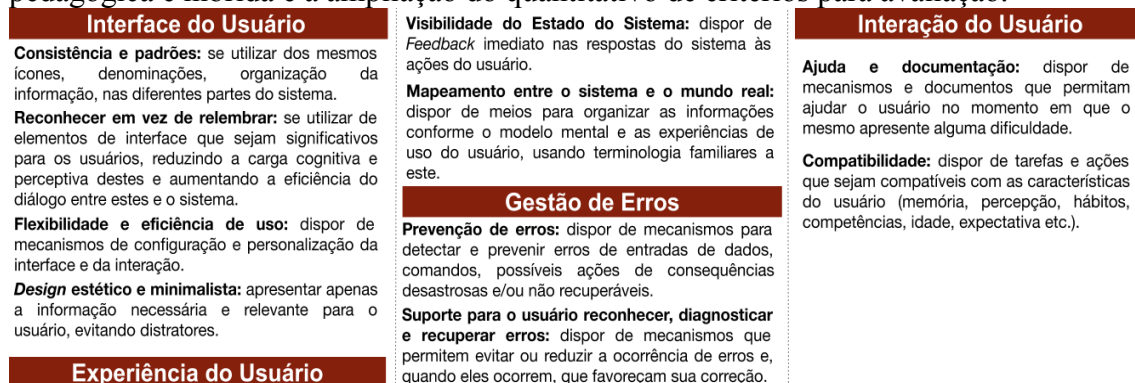


Figura 2. Dimensão de Qualidade de Uso da TARDA



Figura 3. Dimensão de Qualidade Pedagógica da TARDA

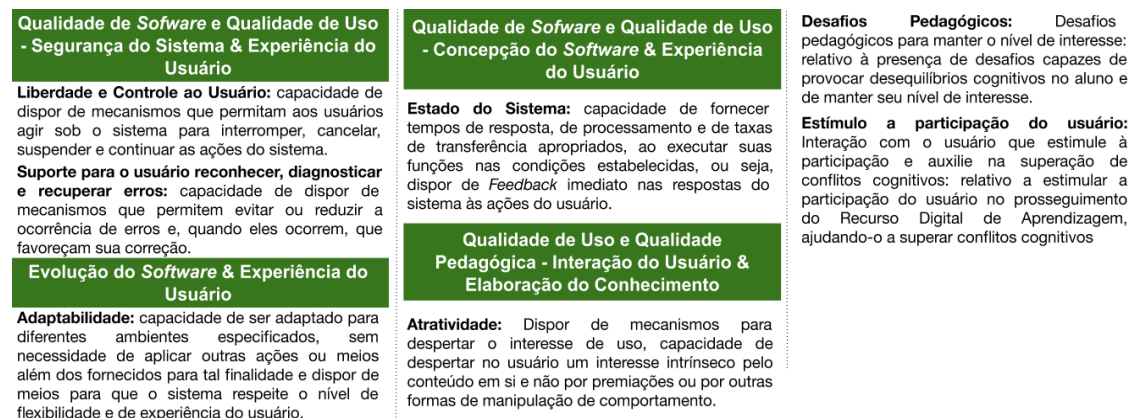


Figura 4. Dimensão de Qualidade Híbrida da TARDA

4 Análise da Adequação da TARDA por Especialistas

Foram obtidas 32 respostas para a dimensão de Qualidade de *Software* e 45 para a dimensão de Qualidade de Uso (Usabilidade), e seus respectivos critérios híbridos. Para Qualidade Pedagógica, e respectivos critérios híbridos, foram entrevistados 4 especialistas em educação e pedagogia.

4.1 Análise de 32 Especialistas em Qualidade de Software

Perfil dos respondentes: 60% doutores, 25,7% doutorandos, 5,7% mestres e 8,6% mestrandos. Atividade principal de ensino para 56,2% para disciplinas de Engenharia de Software, da Computação, de Requisitos, Qualidade de Software, Jogos Educativos, Informática na Educação e Avaliação de Software. Atuação no mercado: 21,8% são engenheiros ou analistas de software, 12,5% são desenvolvedores de software e 9,3% em formação de pós-graduação.

Quantitativo da concordância e discordância (Gráfico 1): Considerando uma taxa de pelo menos 25% de discordância entre os respondentes (4,8 discordâncias), 5 critérios foram mencionados como passíveis de melhoria em termos de sua nomenclatura e descrição: Tolerância a Falhas (5), Testabilidade (5), Substituibilidade (7), Recursos (7) e Maturidade (9).

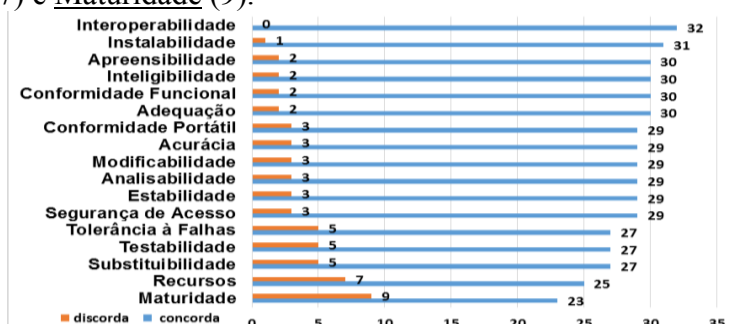


Gráfico 1. Qualidade de Software: Quantitativo de concordância e discordância

Sugestões de melhoria e nova descrição dos critérios:

Tolerância a Falhas (5 sugestões): (i) “recuperar as falhas”, (ii) “capacidade de evitar falhas e manter um nível de desempenho especificado em casos de defeitos no software ou em violação de sua interface especificada”, (iii) “capacidade do sistema continuar operando mesmo em caso de falhas”, (iv) “capacidade de mitigar falhas e manter um nível de desempenho especificado em casos de defeitos no software ou de violação de sua interface especificada”, (v) “capacidade de recuperação e resiliência quando ocorrerem falhas”. Nova descrição: Capacidade de mitigar falhas e manter um nível de desempenho especificado em casos de defeitos no software ou em violação de sua interface especificada.

Maturidade (5 sugestões): (i) “estabilidade”, (ii) “resiliência”, (iii) “maturidade de prevenção a falhas”, (iv) “capacidade de evitar defeitos decorrentes de falhas no software”, (v) “condição adquirida por um software após vários ciclos de desenvolvimento”. Nova descrição: Capacidade de evitar defeitos decorrentes de falhas no software.

Testabilidade (4 sugestões): (i) “capacidade do software de ser testado”, (ii) “capacidade de se testar o sistema modificado, tanto quanto as novas funcionalidades quanto as não afetadas diretamente pela modificação”, (iii) “capacidade de permitir que o software, quando desenvolvido ou modificado, seja validado”, (iv) “capacidade de permitir que o software, quando modificado, seja verificado e validado”. Nova descrição: Capacidade de se testar o sistema modificado, tanto quanto as novas funcionalidades quanto as não afetadas diretamente pela modificação.

Substituibilidade (4 sugestões): (i) “*adaptabilidade*”, (ii) “*capacidade de ser abrangente o suficiente para substituir outro produto de software especificado, com o mesmo propósito e no mesmo ambiente*”, (iii) “*capacidade de ser substituído por outro produto de software especificado, com o mesmo propósito e no mesmo ambiente*”, (iv) “*capacidade de adaptação perante as mudanças*”. Nova descrição: Capacidade de ser substituído por outro produto de *software* especificado, com o mesmo propósito e no mesmo ambiente.

Recursos (4 sugestões): (i) “*uso de recursos*”, (ii) “*funcionalidades*”, (iii) “*capacidade de usar tipos e quantidades de recursos dentro dos limites acordados para o software executar suas funções sob condições estabelecidas*”, (iv) “*tecnologias, frameworks, bibliotecas, todo o suporte para o bom funcionamento do software em questão*”. Nova nomenclatura e descrição: Uso de Recursos - Capacidade de usar tipos e quantidades de recursos (tecnologias, *frameworks*, bibliotecas) dentro dos limites acordados para o *software* executar suas funções sob condições estabelecidas.

Mesmo para os critérios com taxa de discordância abaixo de 15%, alguns comentários foram realizados, para: Adequação: (i) “*adequação dentro de um contexto*”, (ii) “*capacidade de prover um conjunto apropriado de funções específicas para tarefas e objetivos de determinados usuários*”; Conformidade Portátil: (i) “*portabilidade*”; Acurácia: “*capacidade de prover, com o grau de precisão necessário, resultados e efeitos corretos, ou, em conformidade a acordos pré-estabelecidos*”; Modificabilidade: (i) “*manutenibilidade*”; Analisabilidade: “*capacidade de permitir o diagnóstico de deficiências ou causas de falhas no software*”; Estabilidade: (i) “*continuidade*”, (ii) “*capacidade contínua de prover serviço mediante eventuais eventos*”, (iii) “*capacidade de respeitar os requisitos projetados*”; Segurança de Acesso: (i) “*segurança de dados e acessos*”, (ii) “*capacidade de proteger informações e dados, de forma que pessoas ou sistemas não autorizados não possam lê-los, nem modificá-los e que não seja negado o acesso às pessoas ou sistemas autorizados*”. No entanto, devido à baixa taxa de discordância, esses critérios não sofreram modificações nas suas nomenclaturas e nem descrições.

4.2 Análise de 45 Especialistas em Qualidade de Uso (Usabilidade)

Perfil dos respondentes: 42,6% doutores, 24,1% doutorandos, 18,5% mestres, 9,3% mestrandos e 5,6% graduados. Atividade principal de ensino para 37,7% para disciplinas de Usabilidade, Interação Humano-Computador, Qualidade de Uso e Softwares Educativos. Atuação no mercado: 44,3% são especialistas *User Experience* (UX) e analistas de testes de qualidade de uso, 8,8% são desenvolvedores de *software* e 4,4% em formação de pós-graduação.



Gráfico 2. Qualidade de Uso: Quantitativo de concordância e discordância

Quantitativo da concordância e discordância (Gráfico 2): Considerando uma taxa de pelo menos 6,5% de discordância entre os respondentes (2,92 discordâncias), 7 critérios foram mencionados como passíveis de melhoria em termos de sua descrição: Compatibilidade (3), Prevenção de erros (3), Design estético e minimalista (3), Visibilidade do estado do sistema (3), Suporte para o usuário reconhecer, diagnosticar e recuperar erros (4), Flexibilidade e eficiência de uso (4) e Mapeamento entre o sistema e o mundo real (4).

Sugestões de melhoria e nova descrição dos critérios:

Suporte para o usuário reconhecer, diagnosticar e recuperar erros (3 sugestões): (i) “validação”, (ii) “recuperação de erros”, (iii) “dispor de mecanismos que permitem identificar erros e, quando eles ocorrem, que favoreçam sua correção”. Flexibilidade e eficiência de uso (3 sugestões): (i) “customização”, (ii) “dispor de mecanismos de configuração e personalização da interface e da interação”, (iii) “dispor de mecanismos de configuração e personalização da interface e da interação, e a utilização eficiente e eficaz relacionada à experiência do usuário”.

Compatibilidade (2 sugestões): (i) “nivelamento cognitivo”, (ii) “compatibilidade com o usuário”. Prevenção de erros (2 sugestões): (i) “validação de erros”, (ii) “dispor de mecanismos para detectar e prevenir os erros de entradas de dados, comandos, possíveis ações de consequências não recuperáveis”. Design estético e minimalista (2 sugestões): (i) “design minimalista”, (ii) “design estético”. Visibilidade do estado do sistema (2 sugestões): (i) “rápido feedback”, (ii) “visibilidade do status”.

Mapeamento entre o sistema e o mundo real (1 sugestão): (i) “organização entre as interações humano-máquinas, conforme as experiências do usuário, usando terminologias familiares a este”. Percebe-se um alto nível de concordância para esta dimensão da taxonomia; logo, esses critérios não sofreram modificações nas suas nomenclaturas e nem descrições.

4.3 Análise de 4 Especialistas em Qualidade Pedagógica

A análise da dimensão de Qualidade Pedagógica foi realizada através de entrevistas semiestruturadas (reuniões virtuais) com a participação de 4 educadores (1 especialista, 1 doutorando, 2 doutores) que atuam como professores (2) e supervisores escolares (2). A análise dos educadores com relação à nomenclatura das categorias e disposição delas na taxonomia resultou em algumas sugestões e modificações; estes destacaram que a categoria Recursos Motivacionais e Responsivos deveria ser renomeada para “Elaboração do Conhecimento” e a ordem das categorias na taxonomia deveria ser: *conhecimentos prévios* → *fundamentos e objetivos pedagógicos* → *conteúdo pedagógico* → *pertinência do conteúdo* → *clareza epistemológica* → *elaboração do conhecimento*.

Para a análise das nomenclaturas dos critérios os especialistas consideraram a relação critério/descrição e sua disposição nas categorias. Durante as discussões estes analisaram a concordância de cada critério e a coesão e coerência desses critérios em relação à categoria e descrição. Diante disso, eles indicaram várias mudanças nas nomenclaturas, as indicações de mudanças foram: Guia de apoio pedagógico → “*guia de apoio pedagógico ao professor*”; Identifica os objetivos pedagógicos presentes no

mesmo → “*identificação dos objetivos pedagógicos*”; Possui diferentes graus de complexidade de conteúdo nas atividades → “*nível de complexidade dos conteúdos nas atividades*”; Apresenta uma abordagem interdisciplinar → “*abordagem específica ou interdisciplinar*”; Favorece o trabalho em grupo, mas também pode ser utilizado individualmente → “*trabalho em grupo e ou individual*”; Adequação do RDA ao conteúdo → “*adequação do RDA ao conteúdo trabalhado*”; e, Correção do conteúdo → “*análise do conteúdo*”.

As descrições dos critérios foram todas modificadas por eles, pois estes discordaram de todas as descrições (concordância, coesão e coerência), que foram apresentadas para os critérios desta dimensão; as discordâncias também versaram após as mudanças realizadas nas nomenclaturas dos critérios e a alocação destes dentro das categorias.

4.4 Análise para os critérios da Dimensão Híbrida

Perfil dos respondentes: já descritos anteriormente.

Quantitativo da concordância e discordância: Considerando uma taxa de pelo menos 6% de discordância entre os respondentes (4,62 discordâncias), 3 critérios foram mencionados como passíveis de melhoria em termos de sua nomenclatura e descrição: Visibilidade do estado do sistema (5), Suporte para o usuário reconhecer, diagnosticar e recuperar erros (11) e Liberdade e controle ao usuário (11).

Sugestões de melhoria pelos especialistas de Qualidade de Software:

Liberdade e controle ao usuário (3 sugestões): (i) “*Liberdade e controle*”, (ii) “*configurabilidade*”, (iii) “*capacidade de o usuário configurar o sistema conforme seu gosto de uso e necessidade*”. Visibilidade do estado do sistema (2 sugestões): (i) “*auditabilidade*”, (ii) “*dependabilidade*”. Suporte para o usuário reconhecer, diagnosticar e recuperar erros (1 sugestão): (i) “*suporte para reconhecer, diagnosticar e recuperar falhas*”.

Sugestões de melhoria pelos especialistas de Qualidade de Uso:

Visibilidade do estado do sistema (3 sugestões): (i) “*informação extra para o usuário*”, (ii) “*relatório de sistema ou qualidade de serviço*”, (iii) “*dispor de feedback imediato nas respostas do sistema às ações do usuário*”. Suporte para o usuário reconhecer, diagnosticar e recuperar erros (2 sugestões): (i) “*relatórios de erros*”, (ii) “*feedback imediato*”, (iii) “*dispor de mecanismos que permitem identificar erros e, quando eles ocorrem, que favoreçam sua correção*”. Liberdade e controle ao usuário (1 sugestão): (i) “*gerenciamento de tarefas*”.

5 Considerações finais

A análise das 4 dimensões pelos especialistas impactaram em algumas mudanças na representação, nas categorias, nas nomenclaturas e descrições dos critérios da TARDA. O maior grau de mudanças foi na dimensão de QP (33%) e o menor nas dimensões de QU (0) e QH (0). A dimensão de QS teve 29,4% de mudanças e o impacto geral de mudanças na taxonomia como um todo foi correspondente a 12,2%. A nova versão da taxonomia está disponível no [link](#); juntamente com ela também está representada a versão inicial e versão com o rastreamento das mudanças.

Entende-se como limitação desta pesquisa o fato de o estudo ter sido desenvolvido apenas por um aluno; no entanto, para minimizar a limitação, o estudo foi acompanhado por 2 professores com experiência nas áreas de IHC, Qualidade de *Software*, Qualidade de Uso e Avaliação de RDAs. Outra limitação foi referente à avaliação da dimensão de QP realizada apenas por 4 especialistas; para minimizar este impacto, as entrevistas foram realizadas de forma conjunta com os 4 educadores ao mesmo tempo.

Para continuação deste trabalho de pesquisa, pretende-se desenvolver um *Guideline*, que oriente a avaliação de RDAs, composto pela taxonomia proposta, uma escala de avaliação para cada um dos critérios que gere um *score* final da avaliação, e um método de aplicação (passo a passo) para orientar pesquisadores no planejamento, execução, análise e avaliação de RDAs. Além disso, serão realizadas novas entrevistas com especialistas em educação e pedagogia e uma nova análise na taxonomia.

Referências

- Bayona-Oré, Sussy et al. (2014). Critical success factors taxonomy for software process deployment. *In: Software Quality Journal*, v. 22, n. 1, p. 21-48.
- Brito Junior, O. de O., Aguiar, Y. P. C. (2016). Abordagens para Avaliação de *Softwares* Educativos e sua Coerência com os Modelos de Qualidade de *Software*. *In: Simpósio Brasileiro de Informática na Educação*, v. 27, n. 1, p. 270-280.
- Brito Junior, O. de O., Aguiar, Y. P. C. (2018). Taxonomia de Critérios para Avaliação de *Software* Educativo - TaCASE. *In: Simpósio Brasileiro de Informática na Educação*, v. 29, n. 1, p. 298-307.
- Cresswell, John W. (2013). *Research Design: qualitative, quantitative, and mixed methods approaches*. (4 ed.). Sage, Los Angeles.
- França, R. S. de and Silva, A. C. B. da (2014). Avaliação de softwares educativos para o ensino de Língua Portuguesa. *In: Revista Brasileira de Informática na Educação*, v. 22, n. 03, p. 23-33.
- Guarda, G., and Goulart, I. (2018). Jogos lúdicos sob a ótica do pensamento computacional: Experiências do projeto logicamente. *In: Simpósio Brasileiro de Informática na Educação*, v. 29, n. 1, p. 486-496.
- Lin, Yu Han and Oliveira M. C. de (2018). Avaliação de Software utilizando MAEP: Uma Análise do Protótipo de um Simulador Imersivo de Realidade Virtual. *In: Anais SULCOMP*, v. 9.
- Medeiros, Angélica. (2019). YPEDUC: Uma adaptação de Metodologia Ágil para o Desenvolvimento de Software Educativo. *In: Simpósio Brasileiro de Informática na Educação*, v. 30, n. 1, p. 379-389.
- Mesquita, H. O., Aranha, E. and Silva, T. R. (2018). Uma Abordagem para o Desenvolvimento de Jogos Digitais Educativos no Ensino Básico. *In: Simpósio Brasileiro de Informática na Educação*, v. 29, n. 1, p. 874-884.
- Pereira, W. S. et al. (2016). “Validação de uma abordagem combinada para avaliação de Software Educativo: avanços e desafios”. *In: Revista Tecnologias na Educação*, v. 8, n. 16.
- Wohlin C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *In: Proceedings of the 18th international conference on evaluation and assessment in software engineering*, p. 1-10.