

Taxonomia para Avaliação de Recursos Digitais de Aprendizagem – TARDA – Versão 2.0

Ozonias de Oliveira Brito Junior¹, Yuska Paola Costa Aguiar²,
Hermano Perrelli de Moura¹

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

¹ Centro de Informática, Universidade Federal de Pernambuco, 50740-560, Recife – PE, Brasil

² Departamento de Ciências Exatas, Universidade Federal da Paraíba, 58297-000, Rio Tinto – PB, Brasil

DOI: [10.17013/risti.42.120-135](https://doi.org/10.17013/risti.42.120-135)

Resumo: Recursos Digitais de Aprendizagem (RDAs) são projetados para implantar pedagogias baseadas no uso da tecnologia e contribuir para o processo de ensino e aprendizagem. No entanto, a diversidade de abordagens para avaliação de RDAs é uma barreira para o processo de avaliação. Diante disso, a Taxonomia para Avaliação de Recursos Digitais de Aprendizagem (TARDA – Versão 2.0) é proposta a partir de 26 abordagens descritas na literatura nacional e internacional. Esta contempla dimensões de qualidade de *software*, de uso e pedagógica. A fim de verificar a adequação da referida taxonomia junto aos potenciais avaliadores de RDAs, 32 especialistas da área de qualidade de *Software* e 45 da área de qualidade de uso responderam a dois questionários elaborados no escopo desta pesquisa. Além disso, também foram realizadas 7 entrevistas com 21 educadores. Como resultado, tem-se a TARDA – V2 composta por 4 dimensões, 17 categorias e 49 critérios para a avaliação de RDAs.

Palavras-chave: Recursos Digitais de Aprendizagem; Taxonomia para a avaliação de Recursos Digitais de Aprendizagem; Avaliação de Sistemas e Tecnologias da Informação e Comunicação.

Taxonomy for the Assessment of Digital Learning Resources – TADLR – Version 2.0

Abstract: Digital Learning Resources (DLRs) are designed to implement pedagogies based on the use of technology and contribute to the teaching and learning process. However, the diversity of approaches for the assessment of DLRs is a barrier to the evaluation process. Therefore, the Taxonomy for the Assessment of Digital Learning Resources (TADLR – Version 2.0) is proposed from 26 approaches described in national and international literature. This includes dimensions of software quality, quality of use and pedagogical quality. In order to verify the adequacy of the referred taxonomy with the potential evaluators of DLRs, 32 experts in the area of Software quality and 45 in the area of quality of use answered two questionnaires elaborated in the scope of this research. In addition, 7 interviews were also carried out with 21 educators. As a result, TADLR – V2 has 4 dimensions, 17 categories and 49 criteria for assessing RDAs.

Keywords: Digital Learning Resources; Taxonomy for the assessment of Digital Learning Resources; Assessment of Information and Communication Systems and Technologies.

1. Introdução

Com o grande avanço tecnológico e a popularização da internet, Recursos Digitais de Aprendizagem (RDAs) dão oportunidade a uma nova forma de ensino e aprendizagem (Gourlart et al., 2019). Segundo Cascini & Campos (2015), recursos educacionais digitais expandem o potencial e a eficácia do processo de ensino e aprendizagem tradicional, uma vez que a utilização correta desses recursos propicia o aprendizado lúdico do estudante, melhora a interpretação do conteúdo teórico a partir de sua aplicação prática, exercita a resolução de problemas em tempo ágil, possibilita que os discentes aprendam com seus erros, reduz a passividade dos educandos e, conseqüentemente, enriquece a prática pedagógica com a utilização de recursos multimídia.

O interesse da comunidade científica pelo desenvolvimento, pela avaliação e pelo uso de Recursos Digitais de Aprendizagem (RDAs) fica evidente diante das publicações em torno do tema (de França et al., 2014). Para Sales & Boscarioli (2020) as tecnologias digitais estão cada vez mais presentes no cotidiano das pessoas em todo o mundo, e são utilizadas em diferentes contextos e com diferentes finalidades, além de promover os processos de ensino e aprendizagem entre os estudantes. De acordo com Neves et al. (2021) a popularização, no contexto educacional, do uso das Tecnologias Digitais da Informação e Comunicação como recurso pedagógico, tem motivado a realização de pesquisa e produção de trabalhos com tal temática, sugerindo sua utilização como positiva e facilitadora no processo de ensino e aprendizagem, bem como discutindo suas limitações e possibilidades.

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), problemas de usabilidade (Silva et al., 2015), má estruturação e/ou representação didático-pedagógica (Goncalves et al., 2016), problemas de funcionamento (*bugs*) (de França et al., 2014), e falta de fundamentação pedagógica e didática adequadas ao perfil dos aprendizes (Nunes et al., 2018).

Neste contexto, vale destacar que para alcançar os benefícios supracitados, faz-se necessário que os recursos digitais de aprendizagem (i) estejam adequados pedagogicamente, (ii) tenham qualidade de acordo com os preceitos da Engenharia de *Software* e (iii) que seu projeto tenha sido direcionado às características, necessidades e habilidades de seus usuários finais, tal como recomenda a Engenharia da Usabilidade (Junior, Aguiar & de Moura, 2020).

A incidência de inadequações, em qualquer uma destas dimensões, pode comprometer o uso do recurso educacional e, conseqüentemente, o seu propósito principal, potencializar o processo de ensino e aprendizagem. Portanto, a adoção destes recursos deve ser precedida por uma avaliação, para que a sua qualidade seja verificada considerando as três dimensões envolvidas no uso de RDA no contexto educacional de ensino e

aprendizagem – considerando, ainda, a sua adequação em relação ao contexto no qual se pretende aplicá-lo (de França et al., 2014).

Nessa perspectiva, a literatura nacional e internacional apresenta uma gama de abordagens (métodos, técnicas, instrumentos, metodologias e modelos) 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 levam 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.

Diante disso, o objetivo deste artigo é apresentar a evolução de uma proposição (taxonomia versão 1.0) anterior (Junior & Aguiar, 2018) desenvolvida 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 – V2 (TARDA – Versão 2.0), 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 – V2 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 (Júnior, Aguiar & Tavares, 2016).

Como resultado, TARDA – V2 tem 4 Dimensões, que são compostas por 17 Categorias, que agrupam 49 Critérios de Avaliação, associados à 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 21 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 – V2 (TARDA – Versão 2.0) 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 a versão 1.0 da taxonomia (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 1.0 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 – V2. Para tanto, duas estratégias foram adotadas: aplicação de questionários *on-line* e entrevistas semiestruturadas.

Os questionários *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) 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. Para a versão 1.0 da TARDA 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). No entanto, essa validação foi observada como uma limitação que poderia prejudicar o equilíbrio na avaliação.

Diante disso, como expansão deste trabalho de pesquisa e para garantir o equilíbrio na avaliação, foram realizadas novas entrevistas semiestruturadas que colaboraram para o aprimoramento da taxonomia, resultando assim na TARDA - Versão 2.0. Para essa

versão foram realizadas 4 reuniões de 90 minutos (28/11, 05/12, 12/12 e 19/12 de 2020), com a participação de 17 educadores (12 especialistas, 3 doutorandos, 2 doutores). Os objetivos das reuniões foram: (i) análise das nomenclaturas das categorias e disposição delas na taxonomia (primeira reunião), (ii) análise das nomenclaturas dos critérios e sua alocação dentro de cada categoria (segunda reunião), e (iii) análise da descrição de cada critério (terceira e quarta reuniões). Essa segunda rodada de entrevistas com os educadores também teve como objetivo definir cada uma das dimensões que compõem a dimensão de Qualidade Pedagógica.

3. Taxonomia para Avaliação de Recursos Digitais de Aprendizagem (TARDA – V2)

Para a concepção da Taxonomia para Avaliação de Recursos Digitais de Aprendizagem – Versão 2.0, foram consideradas as nomenclaturas e definições amplamente utilizadas na Engenharia de *Software*. Para os critérios da classe de qualidade de *software*, tomou-se como base a terminologia da norma ISO/IEC 9126-1 (2001) e do *Swebok Guide* (2003), enquanto que para os critérios de qualidade de uso (usabilidade), considerou-se o vocabulário da norma ISO/IEC 9241-1 (2003) e das heurísticas definidas por Nielsen (1994), Shneiderman & Plaisant (2004) e Bastien & Scapin (1993). Fez-se necessário eliminar as duplicidades existentes entre os modelos para favorecer o uso homogêneo de termos e de suas respectivas definições. TARDA – V2 está organizada em 4 Dimensões, que são compostas por 17 Categorias, que agrupam 49 Critérios de Avaliação, associados à respectiva Descrição.

A partir da **Figura 1** é possível identificar as 3 categorias e os 17 critérios de avaliação definidos em TARDA – V2 para a qualidade de *Software*. Cada critério da taxonomia tem destaque em **negrito**, é seguido da descrição adotada e tem seu identificador entre parênteses. Quatro critérios (Acurácia, Conformidade Funcional, Tolerância a Falhas e Maturidade) não possuem correspondentes nas abordagens para avaliação de RDAs, mas foram mantidos diante da importância destes para a qualidade de qualquer tipo de *software*, independentemente de ser ou não um Recurso Digital de Aprendizagem.

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. (C01)</p> <p>Acurácia: Capacidade de prover, com o grau de precisão necessário, resultados ou efeitos corretos ou conforme acordados. (C02)</p> <p>Conformidade Funcional: Capacidade de estar de acordo com normas, convenções ou regulamentações previstas em leis e prescrições similares relacionadas à funcionalidade. (C03)</p> <p>Uso de Recursos: Capacidade de usar tipos e quantidades de recursos (tecnologias, <i>frameworks</i>, bibliotecas) dentro dos limites acordados para o <i>software</i> executar suas funções sob condições estabelecidas. (C04)</p> <p>Conformidade Portátil: Capacidade de estar de acordo com normas ou convenções relacionadas à portabilidade. (C05)</p> <p>Interoperabilidade: Capacidade de interagir com um ou mais sistemas especificados. (C06)</p>	<p>Estabilidade: Capacidade de evitar efeitos inesperados decorrentes de modificações no <i>software</i>. (C07)</p> <p>Analisabilidade: Capacidade de permitir o diagnóstico de deficiências ou causas de falhas no <i>software</i>, ou a identificação de partes a serem modificadas. (C08)</p> <p>Modificabilidade: Capacidade de permitir que uma modificação especificada seja implementada. (C09)</p> <p>Substituibilidade: Capacidade de ser substituído por outro produto de <i>software</i> especificado, com o mesmo propósito e no mesmo ambiente. (C10)</p> <p>Instabilidade: Capacidade para ser instalado em um ambiente especificado. (C11)</p> <p>Testabilidade: Capacidade de se testar o sistema modificado, tanto quanto as novas funcionalidades quanto às não afetadas diretamente pela modificação. (C12)</p> <p>Maturidade: Capacidade de evitar defeitos decorrentes de falhas no <i>software</i>. (C13)</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. (C14)</p> <p>Aprensibilidade: Capacidade do produto de <i>software</i> possibilitar ao usuário entender como ele funciona e/ou aprender a usá-lo. (C15)</p> <p>Tolerância a Falhas: Capacidade de mitigar falhas e manter um nível de desempenho especificado em casos de defeitos no <i>software</i> ou em violação de sua interface especificada. (C16)</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. (C17)</p>

Figura 1 – Dimensão de Qualidade de *Software* da TARDA – V2

A partir da **Figura 2** é possível identificar as 4 categorias e os 10 critérios de avaliação definidos em TARDA – V2 para a qualidade de Uso (Usabilidade). Todos os critérios desta dimensão foram mapeados em pelo menos uma das 26 abordagens analisadas, destacando-se o critério Flexibilidade e Eficiência de Uso como sendo o mais citado nas abordagens (21) e os critérios Prevenção de Erros e Suporte para reconhecer, diagnosticar e recuperar erros (4) como os menos citados.

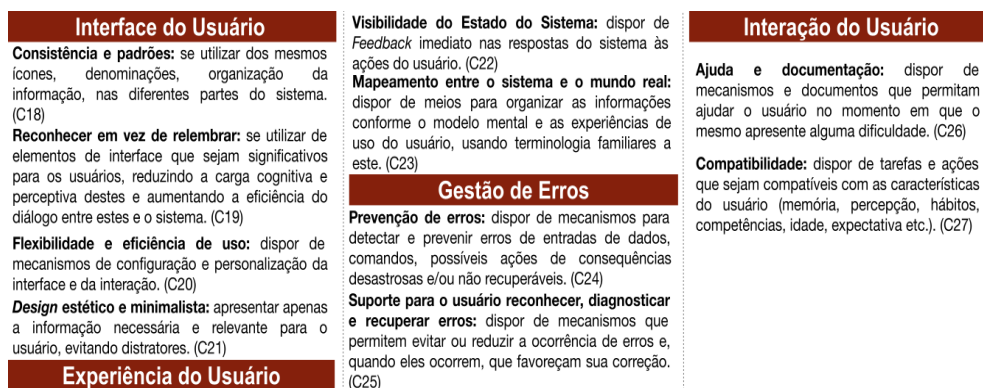


Figura 2 – Dimensão de Qualidade de Uso da TARDA – V2

Os critérios *Tolerância a Falhas* (Figura 1) e *Prevenção de Erros* (Figura 2), embora pareçam semelhantes, distinguem-se por descrição. O primeiro se refere ao sistema se manter em uso, mesmo em caso de falhas ou defeitos, enquanto o segundo consiste na disposição de mecanismos que sejam capazes de detectar e prevenir erros causados a partir das ações dos usuários.

O critério *Adaptabilidade* (Figuras 1 e 2) diz respeito à possibilidade de mudança de ambiente onde o sistema opera (Figura 1), mas também sobre como o sistema se adapta para corresponder às necessidades e características de seus usuários (Figura 2). Considerando que as dimensões da taxonomia se destinam a especialistas de áreas distintas de formação, pode-se compreender que não há sobreposição de critérios/definições neste caso.

De acordo com a **Figura 3**, é possível analisar que a dimensão de Qualidade Pedagógica é composta por 6 categorias e 15 critérios. Todos os critérios dessa dimensão foram contemplados nas 26 abordagens analisadas, destacando-se o critério Adequação do RDA ao usuário e ao currículo como sendo o mais citado nas abordagens (16) e os critérios Registro de Desempenho do Estudante e Indicação dos Conhecimentos Prévios (3) como os menos citados.

Por fim, a **Figura 4** apresenta a dimensão de Qualidade Híbrida que é composta por 4 categorias e 7 critérios. Essa categoria foi elaborada para contemplar os critérios que apresentam uma descrição ampla podendo ser utilizado para avaliar dimensões de qualidade dupla. As duas categorias presentes na primeira coluna compõem a Qualidade

Híbrida das dimensões de Qualidade de *Software* + Qualidade de Uso (3 critérios); as outras duas categorias presentes na segunda coluna compõem a Qualidade Híbrida das dimensões de Qualidade de *Software* + Qualidade de Uso (1 critério) e Qualidade de Uso + Qualidade Pedagógica (3 critérios).

<p>Conhecimentos Prévios</p> <p>Indicação dos conhecimentos prévios: relacionado aos pré-requisitos necessários à compreensão do conteúdo trabalhado no RDA indicados no guia de apoio pedagógico do professor. (C28)</p> <p>Retrabalho com os conhecimentos prévios: relacionado à perspectiva recursiva, onde o RDA parte dos conhecimentos prévios disponíveis possibilitando que o usuário construa e reconstrua seus próprios conceitos. (C29)</p>	<p>Conteúdo Pedagógico</p> <p>Nível de complexidade dos conteúdos nas atividades: relacionado às atividades que estimulam o interesse, a ativação do conhecimento prévio e a utilização desse na elaboração de novos conhecimentos. (C32)</p> <p>Abordagem específica e ou interdisciplinar: relacionado à abordagem do conteúdo disciplinar e ou interdisciplinar. (C33)</p> <p>Trabalho em grupo e ou individual: relacionado às possibilidades de trabalho em grupo e ou individual na execução do RDA. (C34)</p>	<p>Clareza Epistemológica</p> <p>Explicitação da concepção pedagógica: relacionado à concepção didático-pedagógica utilizada no desenvolvimento do RDA. (C38)</p> <p>Coerência pedagógica em relação à base teórico-conceitual: relacionado à teoria pedagógica utilizada no RDA, considerando a correlação entre a teoria e a prática pedagógicas. (C39)</p>
<p>Fundamentos e Objetivos Pedagógicos</p> <p>Guia de apoio pedagógico ao professor: relacionado à presença de informações pedagógicas para que o professor amplie as possibilidades de utilização do RDA. (C30)</p> <p>Identificação dos objetivos pedagógicos: relacionado à especificação dos conhecimentos a serem elaborados pelo usuário na utilização do RDA. (C31)</p>	<p>Pertinência do Conteúdo</p> <p>Adequação do RDA ao conteúdo trabalhado: relacionado à adequação do trabalho didático-pedagógico com o conteúdo do RDA. (C35)</p> <p>Análise do conteúdo: relacionado à correção do conteúdo, sua organização lógica, representação e ausência de erros conceituais. (C36)</p> <p>Adequação do RDA ao usuário e ao currículo: relacionado à complexidade do conteúdo e sua adequação ao nível do usuário em comparação ao currículo escolar. (C37)</p>	<p>Elaboração do Conhecimento</p> <p>Feedback encorajador mediante respostas inadequadas: relacionado ao retorno positivo dado pelo RDA ao usuário, incentivando-o nas atividades sem prejudicar a elaboração do conhecimento. (C40)</p> <p>Reflexão acerca de erros e acertos: relacionado à reflexão do usuário sobre o conteúdo trabalhado, permitindo retorno às atividades e a outras formas de resolução. (C41)</p> <p>Registro de desempenho do usuário: relacionado às respostas das atividades realizadas pelo usuário no RDA. (C42)</p>

Figura 3 – Dimensão de Qualidade Pedagógica da TARDA – V2

<p>Qualidade de Software e Qualidade de Uso - Segurança do Sistema & Experiência do Usuário</p> <p>Liberdade e Controle ao Usuário: capacidade de dispor de mecanismos que permitam aos usuários agir sob o sistema para interromper, cancelar, suspender e continuar as ações do sistema. (C43)</p> <p>Suporte para o usuário reconhecer, diagnosticar e recuperar erros: capacidade de dispor de mecanismos que permitem evitar ou reduzir a ocorrência de erros e, quando eles ocorrem, que favoreçam sua correção. (C44)</p>	<p>Qualidade de Software e Qualidade de Uso - Concepção do Software & Experiência do Usuário</p> <p>Estado do Sistema: capacidade de fornecer tempos de resposta, de processamento e de taxas de transferência apropriados, ao executar suas funções nas condições estabelecidas, ou seja, dispor de <i>Feedback</i> imediato nas respostas do sistema às ações do usuário. (C46)</p>	<p>Desafios Pedagógicos: Desafios pedagógicos para manter o nível de interesse: relativo à presença de desafios capazes de provocar desequilíbrios cognitivos no aluno e de manter seu nível de interesse. (C48)</p> <p>Estímulo a participação do usuário: Interação com o usuário que estimule à participação e auxilie na superação de conflitos cognitivos: relativo a estimular a participação do usuário no prosseguimento do Recurso Digital de Aprendizagem, ajudando-o a superar conflitos cognitivos. (C49)</p>
<p>Evolução do Software & Experiência do Usuário</p> <p>Adaptabilidade: capacidade de ser adaptado para diferentes ambientes especificados, sem necessidade de aplicar outras ações ou meios além dos fornecidos para tal finalidade e dispor de meios para que o sistema respeite o nível de flexibilidade e de experiência do usuário. (C45)</p>	<p>Qualidade de Uso e Qualidade Pedagógica - Interação do Usuário & Elaboração do Conhecimento</p> <p>Atratividade: Dispor de mecanismos para despertar o interesse de uso, capacidade de despertar no usuário um interesse intrínseco pelo conteúdo em si e não por premiações ou por outras formas de manipulação de comportamento. (C47)</p>	

Figura 4 – Dimensão de Qualidade Híbrida da TARDA – V2

4. Análise da Adequação da TARDA – V2 por Especialistas em Qualidade de *Software*, Qualidade de Uso e Qualidade Pedagógica

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 para a TARDA Versão 1.0 e 17 especialistas para a TARDA Versão 2.0, foco deste trabalho.

4.1. Análise de 32 Especialistas em Qualidade de Software

Dos respondentes especialistas em Engenharia de *Software*, 60,0% possuíam doutorado completo, 25,7% doutorado incompleto, 5,7% mestrado completo e 8,6% mestrado incompleto. Destes, 56,2% se dedicam à atividade de ensino lecionando disciplinas de Engenharia de *Software*, da Computação, de Requisitos, Qualidade de *Software*, Jogos Educativos, Informática na Educação e Avaliação de *Software*, 21,8% atuam no mercado de trabalho como engenheiro ou analista de *software*, 12,5% são desenvolvedores de *software*. Os 9,3% restantes são pós-graduandos nos níveis de mestrado ou doutorado.

Quantitativo da concordância e discordância (Figura 5): 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).

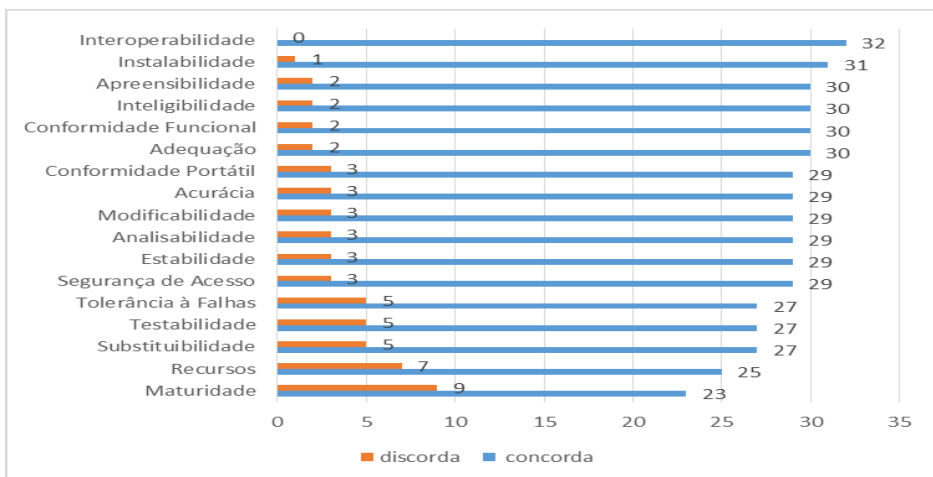


Figura 5 – Qualidade de Software: Quantitativo de concordância e discordância

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)

Dentre os respondentes especialistas em Engenharia de Usabilidade, 42,6% são doutores, 24,1% possuíam doutorado incompleto, 18,5% são mestres, 9,3% possuíam mestrado incompleto e 5,6% são graduados em Ciência da Computação e áreas afins. Estes atuam como professores de nível superior (38,7%) nas disciplinas de Usabilidade, Interação Humano-Computador, Qualidade de Uso e *Softwares* Educativos, especialistas em tecnologias educacionais (37,5%), em *User eXperience* (UX) e Interação Humano-Computador (IHC) (9,8%), em desenvolvedores de *software* (8,8%) e estudantes de pós-graduação (5,2%).

Quantitativo da concordância e discordância (Figura 6): 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).

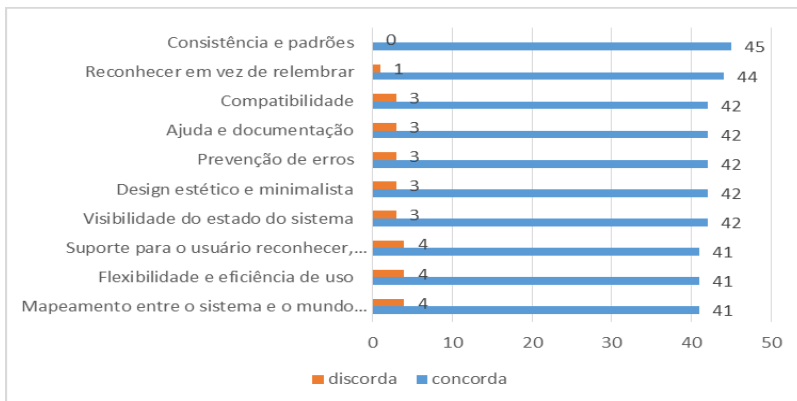


Figura 6 – Qualidade de Uso: Quantitativo de concordância e discordância

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*”. **Nova descrição:** Dispor de mecanismos que permitem evitar ou reduzir a ocorrência de 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*”. **Nova descrição:** Dispor de mecanismos de configuração e personalização da interface e da interação.

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*”. **Nova descrição:** Dispor de mecanismos para detectar e prevenir os erros de entradas de dados, comandos, possíveis ações de consequências desastrosas e/ou 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, a maioria dos critérios não sofreu modificações nas suas nomenclaturas e nem descrições, os critérios que sofreram estão destacados em **negrito** e sublinhado.

4.3. Análise de 21 Especialistas em Qualidade Pedagógica

A análise da dimensão de Qualidade Pedagógica foi realizada através de entrevistas semiestruturadas (reuniões virtuais) com a participação de 21 educadores (13 especialistas, 4 doutorandos, 4 doutores) que atuam como professores (14), pedagogos (2), orientadores (2) e supervisores escolares (3).

A primeira análise dos especialistas de QP para versão 1.0 da TARDA contou com a participação de 4 educadores que observaram as nomenclaturas das categorias e disposição delas na taxonomia, que 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 e 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*”.

Sugestões de melhoria e nova descrição dos critérios:

A segunda análise dos especialistas de QP para versão 2.0 da TARDA contou com 17 educadores que observaram as descrições das categorias e dos critérios, além de algumas

pequenas alterações nas nomenclaturas dos critérios. As descrições das categorias foram: *conhecimentos prévios* - referente aos conhecimentos prévios necessários para o trabalho com o conteúdo proposto no RDA; *fundamentos e objetivos pedagógicos* - referente aos fundamentos da concepção de ensino e aprendizagem inter-relacionada aos objetivos do RDA; *conteúdo pedagógico* - referente à coerência entre a concepção de ensino e aprendizagem e os objetivos do RDA; *pertinência do conteúdo* - referente ao conteúdo educativo do RDA; *clareza epistemológica* - referente à base teórico-conceitual que permeia o desenvolvimento do RDA; e, *elaboração do conhecimento* - referente à interação autônoma do usuário no processo de elaboração do conhecimento.

Também foram indicadas algumas mudanças nas nomenclaturas. As indicações de mudanças foram: Explicitação da fundamentação pedagógica → “*explicitação da concepção pedagógica*”; Consistência pedagógica → “*coerência pedagógica em relação à base teórico-conceitual*”; Feedback encorajador e isento de carga negativa mediante respostas inadequadas → “*feedback encorajador mediante respostas inadequadas*”; Interpretação do usuário acerca de seus erros e acertos → “*reflexão sobre erros e acertos*”; Registro de desempenho do estudante → “*registro de desempenho do usuário*”.

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

A dimensão de Qualidade Híbrida foi avaliada juntamente com os questionários que foram submetidos para os especialistas de QS e QU. Logo, o perfil dos respondentes para essa dimensão segue a distribuição anteriormente apresentada. Com relação à concordância e discordância dos especialistas em relação aos critérios desta dimensão a Figura 7 e 8 apresentam a distribuição dessas análises. Para os especialistas em QS em relação aos de QU a dimensão de QH teve um maior nível de concordância.

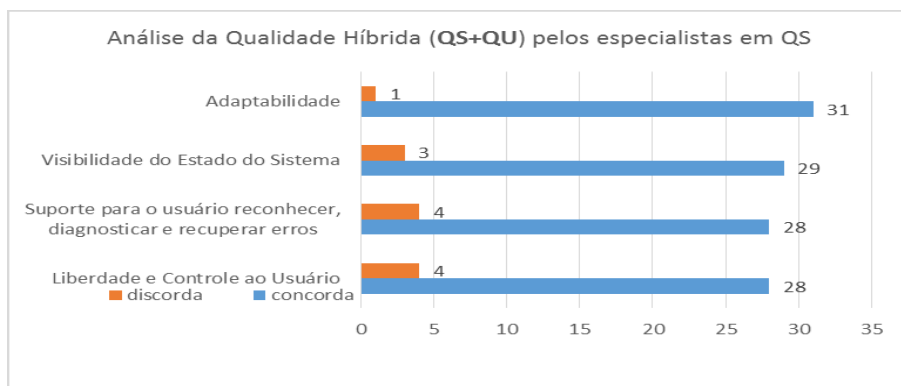


Figura 7 – Quantitativo de concordância e discordância com os critérios de Qualidade Híbrida da TARDA – V2 pelos especialistas em Qualidade de *Software*

Mesmo assim, algumas indicações de mudanças para nomenclaturas dos critérios foram sugeridas pelos respondentes de QS, a saber: Visibilidade do estado do sistema → *auditabilidade* ou *dependabilidade*; Suporte para o usuário reconhecer, diagnosticar e recuperar erros → *suporte para reconhecer, diagnosticar e recuperar falhas*; e, Liberdade e controle ao usuário → *liberdade e controle* ou *configurabilidade*. Também foi indicada uma única sugestão de melhoria para a descrição do critério: Liberdade e controle ao usuário → “*capacidade de o usuário configurar o sistema conforme seu gosto de uso e necessidade*”.

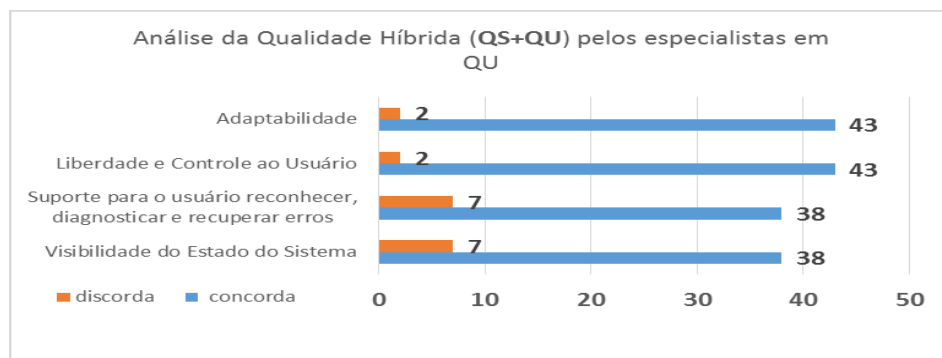


Figura 8 – Quantitativo de concordância e discordância com os critérios de Qualidade Híbrida da TARDA – V2 pelos especialistas em Qualidade de Uso

As indicações de mudanças para nomenclaturas dos critérios pelos especialistas em QU foram: Adaptabilidade → *acessibilidade*; Liberdade e controle ao usuário → *gerenciamento de tarefas*; Suporte para o usuário reconhecer, diagnosticar e recuperar erros → *relatórios de erros ou feedback imediato*; e, Visibilidade do estado do sistema → *informação extra para o usuário ou relatório de sistema ou qualidade de serviço*. Também foram indicadas melhorias para a descrição dos seguintes critérios: Visibilidade do estado do sistema → “*dispor de Feedback imediato nas respostas do sistema às ações do usuário*”; Suporte para o usuário reconhecer, diagnosticar e recuperar erros → “*dispor de mecanismos que permitem identificar erros e, quando eles ocorrem, que favoreçam sua correção*”.

4.5.Representação Gráfica da Taxonomia para Avaliação de Recursos Digitais de Aprendizagem – Versão 2.0

Considerando as análises realizadas pelos especialistas nas 4 dimensões da TARDA - V2, algumas mudanças foram realizadas como já descrevemos nos resultados deste artigo. A Figura 9 destaca a representação gráfica da TARDA – V2 como um todo, com todas as suas 4 dimensões, 17 categorias e 49 critérios de avaliação.

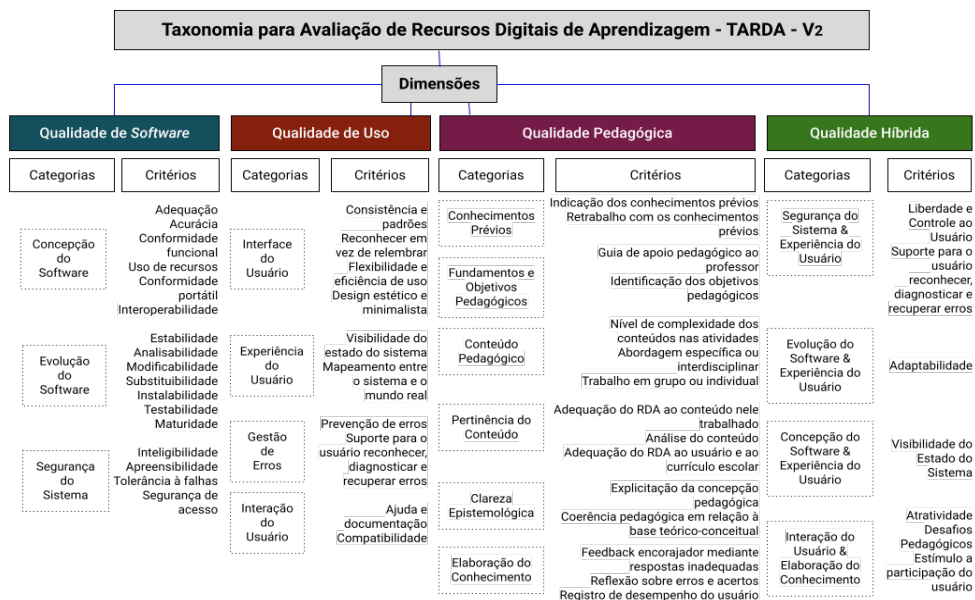


Figura 9 – Representação gráfica da TARDA – V2

5. Considerações finais

A concepção da TARDA – V2 se deu dentro de uma pesquisa realizada ao longo de 4 anos e considerou a análise de 724 critérios de avaliação de Recursos Digitais de Aprendizagem, extraídos de 26 abordagens disponíveis na literatura nacional e internacional. De forma a favorecer a homogeneidade sobre o conhecimento nesta área interdisciplinar, conceitos amplamente utilizados na Engenharia de *Software*, Engenharia de Usabilidade e na Perspectiva Pedagógica foram adotados para a proposição de TARDA – V2.

A fim de avaliar a taxonomia proposta, foram aplicados questionários com especialistas de Engenharia de *Software*, de Usabilidade e de Educação/ Pedagogia, a fim de identificar a concordância (ou discordância) destes diante das categorias, nomenclaturas e descrições. Foram obtidas 98 respostas válidas, 32 para o primeiro perfil, 45 para o segundo e 21 para o terceiro, com a difusão dos questionários em listas de e-mail da comunidade científica que ficaram disponíveis durante quatro semanas. Os resultados obtidos indicam maior concordância dos participantes para os itens que compõem as categorias de Qualidade de Uso e Qualidade Híbrida, sendo os critérios de Qualidade de *Software* dessa dimensão os que tiveram o maior nível de concordância em relação aos de Qualidade de Uso. O maior grau de mudanças foi na dimensão de QP. A dimensão de QS teve 29,4% de mudanças e as alterações realizadas de acordo com a opinião dos especialistas causaram um impacto geral de mudanças na taxonomia como um todo.

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. 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.

Referências

- Bastien, C., & Scapin, D (1993). RT-0156 – Ergonomic criteria for the evaluation of human-computer interfaces. In: Rapport technique de l'INRIA.
- Bayona-Oré, S., Calvo-Manzano, J. A., Cuevas, G., & San-Feliu, T. (2014). Critical success factors taxonomy for software process deployment. *Software Quality Journal*, 22(1), 21-48. <https://doi.org/10.1007/s11219-012-9190-y>
- Cascini, D., & Campos, S. (2015). Avaliação de jogos educacionais multiusuários: Uma revisão sistemática da literatura. In *Brazilian Symposium on Computers in Education* (Vol. 26, No. 1, p. 519). <https://doi.org/10.5753/cbie.sbie.2015.519>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- de França, R. S., & da Silva, A. C. B. (2014). Avaliação de softwares educativos para o ensino de Língua Portuguesa. *Revista Brasileira de Informática na Educação*, 22(03), 23. <https://doi.org/10.5753/RBIE.2014.22.03.23>
- de Sales, A. B., & Boscaroli, C. (2020). Uso de Tecnologias Digitais Sociais no Processo Colaborativo de Ensino e Aprendizagem. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (37), 82-98. <https://doi.org/10.17013/risti.36.1-14>.
- Goulart, J., Oliveira, F., & Pinto, P. (2019). Avaliação Sistemática de Eficácia na Aprendizagem de Algoritmos com o uso do Tupy Online. In *Anais do XXVII Workshop sobre Educação em Computação* (pp. 218-232). <https://doi.org/10.5753/wei.2019.6631>.
- Gonçalves, A. F., Miranda, G. L., & Barreira, N. (2016). B-learning, Recursos Educativos Digitais e Ensino Profissional: Uma estratégia de apoio ao desenvolvimento da Prova de Aptidão Profissional. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (20), 131-146. <https://doi.org/10.17013/risti.20.131-146>
- Junior, O. D. O. B., Aguiar, Y. P. C., & de Moura, H. P. (2020). Taxonomia para Avaliação de Recursos Digitais de Aprendizagem: Proposição e Análise da Adequação por Especialistas. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação* (pp. 232-241). <https://doi.org/10.5753/cbie.sbie.2020.232>.
- Junior, O., & Aguiar, Y. P. C. (2018). Taxonomia de critérios para avaliação de software educativo-TaCASE. In *Brazilian Symposium on Computers in Education* (Vol. 29, No. 1, p. 298). <https://doi.org/10.5753/cbie.sbie.2018.298>.

- Júnior, O. B., Aguiar, Y., & Tavares, T. (2016). Abordagens para Avaliação de Softwares Educativos e sua Coerência com os Modelos de Qualidade de Software. In *Brazilian Symposium on Computers in Education* (Vol. 27, No. 1, p. 270). <https://doi.org/10.5753/cbie.sbie.2016.270>.
- Lin, Y. H., & de Oliveira, M. C. (2018). Avaliação de Software utilizando MAEP: Uma Análise do Protótipo de um Simulador Imersivo de Realidade Virtual. *Anais SULCOMP*, 9.
- Neves, N. N., & dos Santos, A. R. (2021). O uso das tecnologias digitais da informação e comunicação para a experimentação no ensino de química: uma proposta usando sequências didáticas. *Scientia Naturalis*, 3(1).
- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
- Norma, I. S. O. IEC 9126-1. (2001). Software engineering–Product quality–Part 1: Quality model. In: *Associação Brasileira de Normas Técnicas*.
- Norma, I. S. O. IEC 9214-1. (2003), Ergonomic Requirements for office work with visual display terminals – part 11. In: *Associação Brasileira de Normas Técnicas*.
- Nunes, R., & Santos, I. (2018). A importância da avaliação ergonômico-pedagógica de aplicativos educacionais e os desafios encontrados no campo da aprendizagem móvel. In *Brazilian Symposium on Computers in Education* (Vol. 29, No. 1, p. 41). <https://doi.org/10.5753/cbie.sbie.2018.41>.
- Pereira, W. S., Cardoso Filho, R. J., Silva, W. R. D. A., Da Silva, R. S. T., Dantas, V. F., & Aguiar, Y. P. (2016). Validação de uma abordagem combinada para avaliação de Software Educativo: avanços e desafios. *Revista Tecnologias na Educação*, 8(16).
- Shneiderman, B., & Plaisant, C. (2004). Designing the user interface: strategies for effective human-computer interaction. 4. ed. Addison-Wesley Publishing Company.
- Silva, A. C. B. D., & Gomes, A. S. (2015). Conheça e utilize software educativo: avaliação e planejamento para a educação básica. *Recife: Pipa Comunicação*, 83, 3322-3222.
- Swebok. (2004). Guide to the Software Engineering Body of Knowledge. Version Project of the *IEEE Computer Society Professional Practices Committee*.
- 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* (pp. 1-10). <https://doi.org/10.1145/2601248.2601268>.