

Itens digitais no contexto de avaliações em larga escala

Technology-enhanced items in the context of large-scale assessments

Ítems mejorados con tecnología en el contexto de evaluación a gran escala

Liamara Scortegagna

Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas, Programa de Pós-graduação em Gestão e Avaliação da Educação Pública, Juiz de Fora, Minas Gerais, Brasil
liamara@ice.ufjf.br | <http://orcid.org/0000-0001-6825-4945>

Resumo

O artigo apresenta uma revisão sistemática e bibliométrica com o objetivo de identificar e apresentar características, formatos e métodos para o desenvolvimento de itens digitais no contexto de avaliações em larga escala. Para tanto, é detalhado o processo de seleção e análise dos textos a partir de duas bases de dados no período de 2000 a 2021. Os resultados mostram pontos comuns que pautam a concordância sobre a influência do avanço tecnológico para a transição das avaliações do impresso para o digital e o surgimento de novos formatos de itens com diferentes características. Ademais, trazem críticas às organizações quanto à adaptação dos itens usados em avaliações impressas para a avaliação digital sem a clara compreensão de suas diferenças e a crença de que o digital por si só agrega valor.

Palavras-chave: Itens aprimorados por tecnologias. Itens digitais. Formato de itens digitais. Avaliação em larga escala.

Abstract

The article presents a systematic and bibliometric review with the objective of identifying and presenting characteristics, formats and methods for the development of Technology-enhanced items in the context of large-scale assessments. To this end, the process of selection and analysis of texts from two databases in the period from 2000 to 2021 is detailed. The results show common points that guide the agreement on the influence of technological advances for the transition of evaluations from print to digital and the emergence of new formats of items with different characteristics. In addition, they bring criticism to organizations regarding the adaptation of items used in printed assessments to digital assessments without a clear understanding of their differences and the belief that digital in itself adds value.

Keywords: Technology-enhanced items. Digital items. Digital item format. Large-scale Assessment.

Resumen

El artículo presenta una revisión sistemática y bibliométrica con el objetivo de identificar y presentar características, formatos y métodos para el desarrollo de ítems digitales en el contexto de evaluaciones a gran escala. Para ello, se detalla el proceso de selección y análisis de textos de bases de datos en el periodo de 2000 a 2021. Los principales resultados muestran puntos en común que guían el acuerdo sobre la influencia de los avances tecnológicos para la transición de las evaluaciones de lo impreso a lo digital y el surgimiento de nuevos formatos de artículos. Además, critican a las organizaciones en cuanto a la adaptación de los ítems utilizados en las evaluaciones impresas a las evaluaciones digitales sin una comprensión clara de sus diferencias y la creencia de que lo digital en sí mismo agrega valor.

Palabras clave: Ítems mejorados por tecnologías. Ítems digitales. Formato de ítems digitales. Evaluación a gran escala.

1 Introdução

Com a crescente evolução e importância da tecnologia educacional nas últimas duas décadas, as avaliações em larga escala começaram a migrar de avaliações tradicionais impressas para avaliações baseadas digitalmente. Destarte, tal fato motivou o desenvolvimento de formatos de itens baseados em tecnologias

Artigo recebido em: 09/02/2023 | Aprovado em: 29/03/2023 | Publicado em: 08/12/2023

Como citar:

SCORTEGAGNA, Liamara. Itens digitais no contexto de avaliações em larga escala. **Pesquisa e Debate em Educação**, Juiz de Fora: UFJF, v. 13, p. 1-23, e40259. 2023. ISSN 2237-9444. DOI: <https://doi.org/10.34019/2237-9444.2023.v13.40259>.

denominados por muitos autores como itens aprimorados por tecnologias (*Technology-Enhanced Items - TEIs*), itens habilitados para tecnologia (*technology-enabled items*), itens inovadores (*innovative items*), itens inovadores aprimorados por tecnologia (*technology-enhanced innovative items*), itens baseados em computador (*computer-based items*), itens de teste informatizados inovadores (*innovative computerized test items*), itens digitais (*digital items*), entre outros.

No entanto, alguns dos formatos desses itens não são tão inovadores, pois são remanescentes das avaliações impressas e estão, agora, adaptados para avaliações digitais sem nenhum tipo de alteração. Já em outros casos, integram alguma mídia ou recursos tecnológicos, como vídeos ou gráficos. Ademais, há ainda uma série de formatos que são realmente inovadores e construídos para usar ações de resposta indisponíveis no modelo de avaliação impressa, casos das simulações de ações reais em espaços virtuais, animações e do uso de Inteligência Artificial (IA) por exemplo.

Bryant (2017), com base nos escritos de Jodoin (2003), Kane (2006), Tarrant, Knierim, Hayes e Ware (2006), Boyle e Hutchinson (2009), Parshall, Harmes, Davey e Pashley (2010), afirma que se os itens digitais forem bem planejados e desenvolvidos podem oferecer vantagens sobre os tradicionais itens de múltipla escolha ou respostas construídas, tais como: ampliar a medição do constructo; apresentar contextos mais autênticos para a demonstração de habilidades e conhecimentos; reduzir os efeitos de adivinhação aleatória; reduzir a irrelevância do constructo; aumentar as oportunidades de medição; facilitar a pontuação eficiente em termos de tempo e custo das respostas; e melhorar a motivação dos respondentes através de um maior envolvimento.

No Brasil, as avaliações em larga escala que utilizam itens digitais começam a se fortalecer, porém de forma tímida e ainda com a prevalência dos itens tradicionais projetados para o impresso e simplesmente adaptados para as plataformas digitais conforme comprovam dados de pesquisas realizadas por Bruno (2019) e Scortegagna (2020).

Dessa forma, a principal motivação para este estudo é analisar, na literatura existente, quais são as características, formatos, técnicas e métodos utilizados para o planejamento e desenvolvimento de itens digitais, buscando observar o que há de inovador para a aplicação em novos *designs* de itens digitais interativos.

A partir da utilização de métodos de revisão sistemática e bibliométrica voltados a coletar e analisar dados de diferentes bases de pesquisas, o estudo tem por objetivo investigar e apresentar características, formatos, técnicas e métodos empregados para o planejamento e desenvolvimento de itens digitais ou aprimorados por tecnologias, no contexto de avaliações em larga escala, no âmbito nacional e internacional.

Adicionalmente à seção introdutória, este artigo exhibe mais três seções. Na seção 2, é apresentada a metodologia de pesquisa. Já a seção 3 mostra os resultados das revisões bibliométrica e sistemática e suas respectivas análises. Finalmente, a seção 4 conclui o artigo com a exposição dos principais destaques da pesquisa, bem como suas limitações e considerações.

2 Metodologia

Neste trabalho, utilizamos as metodologias Revisão Sistemática da Literatura (RSL) e Revisão Bibliométrica para investigar e apresentar características, formatos, técnicas e métodos utilizados para o planejamento e desenvolvimento de itens digitais ou aprimorados por tecnologias, no âmbito nacional e internacional, no contexto de avaliações em larga escala.

A RSL é definida por Kitchenam (2004) como uma forma de identificação, avaliação e interpretação de trabalhos relevantes para uma determinada questão de pesquisa. Normalmente, as formas de utilização da RSL servem para nortear o desenvolvimento de projetos (SAMPAIO; MANCINI, 2007), pois possuem características de serem “metódicas, explícitas e passíveis de reprodução” (SAMPAIO; MANCINI, 2007, p. 83).

Além de nortear o desenvolvimento de projetos, há outras razões para a utilização da metodologia de RSL. Dentre essas, podemos citar: resumir alguma evidência existente sobre uma determinada teoria ou tecnologia; identificar pontos em aberto para a pesquisa em questão, possibilitando a definição de áreas onde mais investigações devem ser realizadas; e prover um embasamento para novas atividades de pesquisa (KITCHENHAM, 2004).

Para Paula, Araújo e Silva (2016), a Revisão Sistemática da Literatura tem por objetivo fornecer uma completa avaliação do estado da arte relacionado a um tópico de interesse. Ainda segundo os autores, a RSL pode ser considerada uma metodologia específica de pesquisa, que obedece a uma sequência de passos metodológicos rigorosa e bem definida, em acordo com um protocolo previamente desenvolvido contendo a formulação da questão central, o foco da pesquisa, as bases de conhecimento utilizadas e os critérios de filtragem na seleção de trabalhos, entre outras definições.

Por sua vez, a Revisão Bibliométrica tem por objetivo explorar e avaliar os dados coletados na revisão sistemática. O princípio da bibliometria é analisar a atividade científica ou técnica por meio de estudos quantitativos de publicações. Ademais, os indicadores quantitativos, além de apoiarem a compreensão da dinâmica da ciência e da tecnologia, também funcionam como ferramentas para o planejamento de políticas e tomada de decisões nesse setor (SANTOS, 2003).

2.1 Descrição e execução

A primeira etapa de uma Revisão Sistemática e Bibliométrica corresponde à delimitação da questão que norteia o estudo, seguida da criação de alguns critérios.

A questão que norteou a investigação foi proposta a partir do objetivo do estudo e consiste na seguinte pergunta: Quais são as características, formatos, técnicas e métodos utilizados para o planejamento e desenvolvimento de itens digitais ou aprimorados por tecnologias no contexto de avaliações em larga escala?

A partir da questão de pesquisa, especificamos alguns critérios (Quadro 1) para a realização da Revisão Sistemática e Bibliométrica, como a definição das bases de dados, palavras-chave, idiomas, *string* de busca, tipos de publicações e critérios de inclusão:

Quadro 1: Critérios para a Revisão Sistemática e Bibliométrica

| Critério | Descrição |
|------------------------|---|
| Seleção de Fontes | Será fundamentada em bases de dados eletrônicas: Portal de Periódicos da CAPES e Google Acadêmico. |
| Palavras-chave | Avaliação em larga escala; avaliação educacional em larga escala; itens digitais. <i>large-scale assessments; large-scale educational assessment; technology-enhanced items.</i> |
| Idioma dos Estudos | Português e inglês. |
| <i>String</i> de busca | (“avaliação em larga escala” OR “avaliação educacional em larga escala”) AND (“itens digitais”) (“ <i>large-scale assessments</i> ” OR “ <i>large-scale educational assessments</i> ”) AND (“ <i>Technology-Enhanced Items</i> ”). |
| Tipos de publicações | Teóricos/bibliográficos, Estudos Experimentais e Estudo de Casos. |
| Critérios de Inclusão | Os textos devem: possuir resumo e texto completo disponíveis; ter relação com estudo ou experimento ou indicação de características, formatos, técnicas e métodos de planejamento ou de desenvolvimento de itens digitais no contexto de avaliações em larga escala; ter sido publicados no período de 2000 a 2021; ter sido publicados em inglês ou português. |

Fonte: da autora, 2022.

No processo de seleção dos estudos preliminares, aplicou-se a *string* de busca nas bases de dados “Periódicos da CAPES” e “Google Scholar” no período de 10 a 19 de julho de 2021.

A escolha da Base de Dados “Periódicos da CAPES” deve-se ao fato de esta possibilitar acesso a um conjunto de bases de dados nacionais e internacionais atualizadas e em todas as áreas do conhecimento. Optamos ainda pela Base de Dados Google Acadêmico, pois apresenta múltiplas fontes acessíveis gratuitamente.

Ademais, utilizamos a *string* de busca desenvolvida a partir das palavras-chave nos idiomas português e inglês, combinadas com os operadores booleanos “AND” e “OR”. Com a aplicação da *string*, foi possível obter um resultado preliminar de 139 documentos conforme descrito na Tabela 1:

Tabela 1: Estratégia de pesquisa por base, idioma e respectivos resultados

| String/Idioma | Bases de Dados | |
|--|------------------------------|------------------|
| | Porta de Periódicos da CAPES | Google Acadêmico |
| Inglês | | |
| ("large-scale assessments" OR "large-scale educational assessments") AND ("Technology-Enhanced Items") | 07 | 129 |
| Português | | |
| ("avaliação em larga escala" OR "avaliação educacional em larga escala") AND ("itens digitais") | 0 | 03 |
| Total por Base | 07 | 132 |
| Total Geral | 139 | |

Fonte: da autora, 2022.

A avaliação dos textos ocorreu a partir da análise dos títulos e resumos das publicações identificadas que preencheram os critérios de inclusão. Os resultados foram organizados e tabulados utilizando o programa *Microsoft Excel* e serão apresentados na seção a seguir.

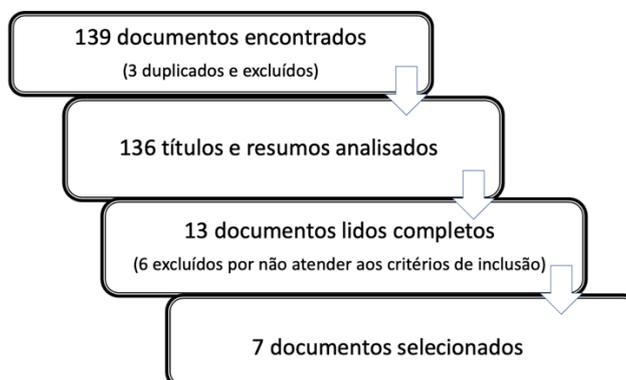
3 Apresentação e análise dos resultados

Nesta seção, trazemos os resultados da pesquisa, os quais serão divididos por tipo de análise realizada: bibliométrica e sistemática, respectivamente.

3.1 Análise bibliométrica

A pesquisa inicial obteve 139 documentos, dos quais 3 (três) estavam em duplicidade e foram excluídos. Destarte, os 136 títulos e resumos arrolados foram analisados, sendo selecionados 13 textos para leitura completa. Destes, 6 (seis) foram excluídos por não atenderem aos critérios de inclusão, restando um total de 7 (sete) documentos selecionados por satisfazerem os critérios de elegibilidade conforme demonstrado no fluxograma exposto na Figura 1 a seguir:

Figura 1: Fluxograma do estudo



Fonte: da autora, 2022.

Após as buscas e análises completas dos 7 (sete) documentos selecionados para compor o estudo, foram relacionados e são apresentados, por ordem alfabética, na Tabela 2, sendo 6 (seis) artigos científicos publicados em periódicos e uma tese de doutorado:

Tabela 2: Referências completas dos documentos selecionados

| Referências completas |
|---|
| BRYANT, William. Developing a Strategy for Using Technology-Enhanced Items in Large-Scale Standardized Tests. Practical Assessment, Research, and Evaluation : Vol. 22, Article 1. 2017. |
| BURGA LEÓN, Andrés Alberto. Aplicaciones de la tecnología a la evaluación psicométrica. Propósitos y Representaciones , 7(SPE), e318. 2019. |
| JIANG, Yang; GONG, Tao; SALDIVIA, Luis E.; CAYTON-HODGES, Gabrielle; AGARD, Christopher. Using process data to understand problem-solving strategies and processes for drag-and-drop items in a large-scale mathematics assessment. (Research). Large-scale Assessments in Education . 2021. |
| MONCALEANO, Sebastian. Examining the comparative measurement value of Technology-Enhanced Items . 2021. 261 p. PhD thesis. Boston College Lynch School. 2021. |
| MOON, Jung A.; KEEHNER, Madeleine; KATZ, Irvin R. Test Takers Response Tendencies in Alternative Item Formats: A Cognitive Science Approach. Educational Assessment , 25(3), 236–250. 2020. |
| RUSSELL, Michael. A Framework for Examining the Utility of Technology-Enhanced Items. Journal of Applied Testing Technology , Vol 17(1), 20-32, 2016. |
| WISE, Steven L.; SOLAND, James; DUPRAY, Laurence M. The Impact of Technology-Enhanced Items on Test-Taker Disengagement. Journal of Applied Testing Technology , Vol 22(1), 28-36, 2021. |

Fonte: da autora, 2022.

Já o Gráfico 1 a seguir traz o ano de publicação dos documentos selecionados no período de 2000 a 2021. É possível observar que somente no ano de 2016 iniciam-se as publicações, seguindo o número quase constante de um documento por ano e aumentando consideravelmente no ano de 2021. Isso se deve, principalmente, à evolução das tecnologias e seu uso nas avaliações em larga escala e no desenvolvimento de itens digitais ocorrido na última década. Observa-se que essa revisão foi desenvolvida no mês de julho de 2021:

Gráfico 1: Ano de publicação dos documentos



Fonte: da autora, 2022.

Ressalta-se, ainda, que as pesquisas publicadas têm como origem, em sua maioria, os Estados Unidos (EUA), com 6 documentos, sendo que apenas um é originário da América Latina, mais especificamente do Peru

A avaliação dos periódicos que publicaram os documentos selecionados foi levantada a partir dos dados disponíveis nos índices de fator de impacto *Journal Citation Reports (JCR)*¹, *Scientific Journal Rankings (SJR)*² e *h-índice*. Os resultados encontram-se apresentados na Tabela 3 seguinte. Como nem todos os documentos são oriundos de periódicos, não foi possível avaliar toda a amostra. Entretanto, dos sete documentos, foi possível avaliar os periódicos de seis deles:

Tabela 3: Avaliação dos periódicos

| Título da publicação | Fator de impacto SJR | <i>h-index</i> |
|--|----------------------|----------------|
| Practical Assessment, Research, and Evaluation | 1.306 ^a | 52 |
| Journal of Applied Testing Technology | NC | NC |
| Educational Assessment (B3) | 0.806 ^b | 27 |
| Propósitos y Representaciones | NC | 33 |
| Large-scale Assessments in Education | 1.294 ^b | 12 |

^a Dado referente a 2017 / ^b Dado referente a 2020

Fonte: da autora, 2022.

O fator de impacto é considerado ainda um dos melhores instrumentos de avaliação de qualidade de revistas científicas, sendo que, quanto maior o índice, mais bem classificada é a revista. Para a análise do fator de impacto, utilizamos os indicadores constantes no JCR e o SJR. As principais diferenças entre os indicadores decorrem especialmente das diferentes bases de dados utilizadas como fontes de citações e de diferenças na estimação desses índices (FALAGAS *et al.*, 2008).

Os resultados apresentados na Tabela 3 apontam uma qualidade razoável dos periódicos medida pelo fator de impacto, destacando-se entre eles o *Practical Assessment, Research, and Evaluation*. Ressalta-se que nenhum dos periódicos consta no fator de impacto JCR.

Já o *h-index* é um complemento útil para fatores de impacto de revistas, pois combina o número de publicações (quantidade) e a taxa de citação (qualidade) de forma equilibrada, sendo calculado para o período de um ano (BRAUN; GLÄNZEL; SCHUBERT, 2006). Por exemplo, um *h-index* de 40 significa que 40 artigos foram publicados e que cada um conteve pelo menos 40 citações (BORNMANN; DANIEL, 2007). Nesse sentido, destaca-se o periódico *Practical Assessment, Research, and Evaluation* com um *h-index* de 52.

Foi realizada, também, uma avaliação da quantidade de citações que cada documento possui no *Google Acadêmico*, que pode ser observado na Tabela 4:

Tabela 4: Citações Google Acadêmico

| Referências | Google Acadêmico |
|---|------------------|
| BRYANT (2017) | 27 |
| WISE; SOLAND; DUPRAY (2021) | 0 |
| RUSSELL (2016) | 14 |
| MONCALEANO (2021) | 0 |
| MOON; KEEHNER; KATZ (2020) | 2 |
| BURGA LEÓN (2019) | 1 |
| JIANG; GONG; SALDIVIA; CAYTON-HODGES (2021) | 0 |

Fonte: da autora, 2022.

português: “medição”, “teste”, “larga escala”, “avaliações”, “adaptativo”, “computador”, “resposta”, “cognitivo”.

Ressalta-se que foram verificadas variadas formas de escrita utilizadas para se referir a itens digitais como, *Technology-Enhanced Items*, *TEIs*, *innovative items*, *digital items*, *novel item formats*, o que mostra uma falta de padrão para o termo. Entre as formas, *Technology-Enhanced* é a mais aplicada pelos autores.

3.2 Análise sistemática

Com fundamento na análise sistemática e bibliométrica, observa-se a pouca quantidade de publicações abordando os temas da pesquisa proposta. Na sequência, são apresentadas as principais contribuições e achados de cada um dos textos selecionados.

O primeiro texto, “*Developing a Strategy for Using Technology-Enhanced Items in Large-Scale Standardized Tests*”, é de autoria de William Bryant (BRYANT, 2017). O autor faz uma análise sobre itens digitais ou itens aprimorados por tecnologias (*Technology-Enhanced Items* – TEIs) à luz da necessidade dos responsáveis pela organização de avaliações em larga escala possuírem argumentos baseados em evidências para seu uso. Bryant (2017) relata que tanto as instituições responsáveis pelas avaliações quanto os desenvolvedores de itens, acreditam que os itens digitais, por si só, agregam valor, porém considera difícil de verificar e descrever que tipo de valor e se este realmente existe. Ademais, o autor pontua que as organizações de avaliações apenas incorporam os itens digitais sem a clara compreensão de suas diferenças, implicações, medições, compensações de custo-benefício ou efeitos sobre os participantes das avaliações.

A partir dessas constatações, Bryant (2017) passa a conceituar, classificar e comparar itens digitais com itens convencionais, principalmente os itens de múltipla escolha e de construção de resposta (produção textual), bem como a apresentar os objetivos e limitações dos itens digitais.

Ainda sobre o tema, o autor apresenta uma estratégia para tomada de decisões sobre o uso de itens digitais em avaliações em larga escala, ressaltando que o uso implica o desenvolvimento de critérios sobre como avaliar o valor relativo desses itens. Para Bryant (2017), os itens digitais parecem atingir seu maior potencial utilitário quando enriquecem a medição em comparação com os itens tradicionais de múltipla escolha, bem como melhoram a eficiência das avaliações e os custos de pontuação em comparação com os itens tradicionais de construção de respostas.

O segundo texto selecionado foi de Wise, Soland e Dupray (2021) – “*The Impact of Technology-Enhanced Items on Test-Taker Disengagement*”. Nele, os autores afirmam que os itens aprimorados por tecnologias são considerados mais motivadores e envolventes para os participantes de avaliações do que os itens tradicionais de múltipla escolha. No entanto, a alegação de maior envolvimento recebeu atenção limitada em pesquisas constantes na literatura. Dessa forma, realizaram um estudo sobre as taxas de comportamento de adivinhação rápida em três tipos de itens (múltipla escolha, seleção múltipla e itens aprimorados por tecnologias (TEIs)) em um teste de desempenho adaptativo comumente usado no ensino fundamental e no ensino médio nos EUA.

Para que pudessem analisar e diferenciar TEIs de itens convencionais, Wise, Soland e Dupray (2021) apresentaram algumas definições embasadas nos estudos de outros pesquisadores, como Sireci e Zenisky (2016, p. 314), que descreveram os TEIs como formatos de itens inovadores baseados em computador que: “(a) minimizam as desvantagens dos itens de múltipla escolha, (b) medem os aspectos do constructo alvo que não podem ser medidos usando itens de múltipla escolha, e (c) aumentam o engajamento e a motivação entre os examinandos” Os TEIs básicos geralmente incluem (mas não estão limitados) itens que exigem que os participantes das avaliações exibam uma variedade de ações de resposta, como arrastar e soltar conteúdo, destacar texto ou selecionar várias respostas. TEIs mais avançados podem até envolver ambientes simulados, exigindo interação com avatares e outros tipos de ações de respostas complexas. Na pesquisa desenvolvida por Wise, Soland e Dupray (2021), estes se pautaram em TEIs que usaram as ações de resposta mais básicas.

Para atingir o objetivo da pesquisa, os autores utilizaram os dados do programa de testes de realização do K-12 em leitura, matemática e ciências. Todos as avaliações foram do sistema de testes da plataforma Medidas de Progresso Acadêmico MAP® Growth™ da NWEA, usando bancos de itens contendo mais de 24 milhões de respostas de itens.

Como conclusão, Wise, Soland e Dupray (2021) afirmam que, nas três áreas temáticas, os TEIs mostraram consistentemente a menor taxa de adivinhação rápida, sugerindo que seu uso pode ajudar a mitigar o problema de fazer o teste sem envolvimento.

Russell (2016), em seu texto “*A Framework for Examining the Utility of Technology-Enhanced Items*”, relata que, na última década, houve um aumento considerável de interesse no uso de itens aprimorados pela tecnologia em avaliações. Ademais, devido ao tempo necessário para administrá-los, bem como ao alto custo para o desenvolvimento, o autor ressalta a importância de que as instituições que organizam as avaliações considerem a utilidade desses itens.

Pautado em Haladyna e Rodriguez (2013), Russell (2016) pondera que todos os itens de avaliações são compostos de, pelo menos, duas partes: a) um estímulo que estabelece o problema em que o candidato deve se concentrar; e b) um espaço de resposta no qual o candidato registra uma resposta ao problema. Para itens utilizados em ambientes digitais, o espaço de resposta tem sido chamado de espaço de interação, já que esta é a área em que o aluno interage com o sistema de avaliação para produzir uma resposta.

Todavia, o autor observa que há uma ampla e crescente variedade de espaços de interação empregados em itens digitais. Por exemplo, uma classe de espaço de interação apresenta palavras ou objetos que são classificados em duas ou mais categorias, arrastando-as e soltando-as em seus respectivos recipientes. Outra classe de espaço de interação requer que os participantes das avaliações criem um pequeno texto para produzir uma resposta em um determinado espaço. Já uma terceira classe de espaço de interação requer que os participantes da avaliação destaquem o conteúdo para produzir uma resposta.

Dada a crescente variedade de espaços de interação de respostas que podem ser criados em um ambiente digital, Russell (2016) projetou um

Framework/estrutura de utilidade para itens digitais com o objetivo de ajudar instituições organizadoras de avaliações a pesar os custos e benefícios de empregar uma determinada metodologia de interação de resposta para medir o conhecimento, habilidade ou capacidade de interesse. O autor define a utilidade de um espaço de interação de um item aprimorado pelas tecnologias como sua utilidade para medir conhecimentos ou habilidades específicas.

Ao considerar a utilidade de um espaço de interação em itens aprimorados pela tecnologia, Russell (2016) aponta três características que devem ser consideradas: a) fidelidade ao constructo; b) usabilidade do espaço de interação para a produção de respostas; e c) acessibilidade da interação de resposta para participantes com deficiências específicas e necessidades especiais.

Assim, o *Framework* projetado pelo autor pretende ajudar os desenvolvedores de avaliações a considerar até que ponto o uso de um determinado espaço de interação de resposta é apropriado para o constructo medido por um item e, ainda, se a forma de implementação permite que os participantes da avaliação produzam respostas com precisão e eficiência, refletindo os processos cognitivos. Embora cada um desses fatores seja examinado individualmente na pesquisa, eles foram considerados coletivamente a fim de avaliar a utilidade da interação.

Destarte, Russell (2016) conclui, após a aplicação e testes com o *Framework*, que, do ponto de vista econômico, fica claro que os novos espaços de interação empregados por itens aprimorados por tecnologias têm utilidade e maior custo. No entanto, é importante lembrar que o critério primário para incluir qualquer item em uma avaliação é sua capacidade de contribuir com evidências precisas com utilidade para informar a medida de um constructo alvo.

O *Framework* de Russell objetiva também ajudar os programas de avaliações e desenvolvedores de itens a manter o foco no valor de medição, direcionando a atenção para o uso de um determinado espaço de interação para produzir um contexto autêntico, utilizável e acessível no qual o constructo alvo é aplicado pelo candidato.

Moncaleano (2021), em sua tese de doutorado, intitulada “*Examining the comparative measurement value of Technology-Enhanced Items*”, argumenta e toma como base diversos autores que afirmam que o crescimento dos testes baseados em computador nas últimas duas décadas motivou a criação de formatos de itens inovadores. Tais pesquisadores sustentam ainda que frequentemente se argumenta que os Itens Aprimorados por Tecnologia (TEIs) fornecem uma melhor medição do conhecimento e das habilidades dos participantes, aumentando a autenticidade das tarefas apresentadas aos participantes (SIRECI; ZENISKY, 2006).

Nesse sentido, o autor afirma que, apesar da popularidade dos TEIs em avaliações, ainda existem poucas pesquisas psicométricas sobre esses formatos de itens inovadores. Dessa forma, Moncaleano (2021) investigou, a partir da proposição e aplicação de um protocolo, como julgar o Valor de Medição Comparativa (CMV) de dois formatos de itens aprimorados por tecnologias, quais sejam o de arrastar e soltar (classificação e ordenação) em relação aos itens de múltipla escolha. Os itens foram administrados a uma amostra de alunos adultos, e os resultados foram calibrados usando um modelo de IRT logístico de dois

parâmetros. Além disso, a utilidade dos TEIs foi avaliada de acordo com o instrumento “*TEI Utility Framework*” proposto por Russell (2016) e acrescenta à pesquisa, a partir do embasamento teórico, uma proposta de metodologia para comparar os TEIs aos formatos de itens tradicionais.

Para embasar o estudo, Moncaleano (2021) organizou uma ampla revisão da literatura dividida em duas partes. Na primeira, a evolução dos itens tradicionais e aprimorados por tecnologia foi descrita em detalhes, incluindo uma visão geral da transição para testes baseados em computador, definições de itens aprimorados por tecnologia e uma discussão sobre seus benefícios e limitações. Aqui, o autor destaca que os testes baseados em computador foram lentamente sendo introduzidos no âmbito educacional a partir da década de 70 (realidade dos EUA) e, desde então, simplesmente, foram replicando os itens tradicionais usados em testes impressos em uma plataforma digital. A partir desse momento, desencadeou-se um conjunto de pesquisas que examinaram a equivalência de itens tradicionais apresentados digitalmente.

Em relação aos TEIs, Moncaleano (2021) destaca que estes diferem de formatos tradicionais e têm sido referidos por vários termos nas últimas três décadas. Esses termos incluem: novos tipos de itens, itens inovadores, itens habilitados para tecnologia, itens baseados em computador, tarefas sofisticadas, itens interativos e itens aprimorados por tecnologia (BRYANT, 2017). Enquanto “novo tipo de item” foi o rótulo mais prevalente na década de 1990 e “item inovador” nos anos 2000, o “item aprimorado por tecnologia” tornou-se o rótulo preferido durante a década mais recente.

As duas abordagens para caracterizar formatos de itens inovadores apresentadas por Moncaleano (2021) são definidas como (a) definições operacionais e (b) esquemas de classificação. As definições operacionais apresentam critérios que permitem classificar um formato de item como inovador ou tradicional. Em contrapartida, os esquemas de classificação consideram a inovação como um contínuo, no qual os tipos de itens são classificados de acordo com o grau de inovação que fornecem com base em múltiplas características.

Na segunda parte da revisão bibliográfica realizada por Moncaleano (2021), o autor examinou vários esforços feitos para avaliar a qualidade dos TEIs. Os métodos e critérios usados para avaliar a qualidade de um item tradicional incluem o exame de dificuldade e discriminação do item, qualidade do distrator do item, dimensionalidade do teste, confiabilidade, ajuste do item e do modelo e informações do item. Além disso, apresenta vários trabalhos que compararam itens com base em seu formato, modo de entrega e interface. Entretanto, segundo Moncaleano (2021), a extensa literatura apresentada baseou-se principalmente em formatos de itens tradicionais, levantando a questão quanto a saber se os critérios e métodos empregados são pertinentes a itens aprimorados por tecnologia ou se novas abordagens são justificadas.

Em relação ao instrumento “*TEI Utility Framework*”, desenvolvido por Russell (2016), Moncaleano (2021) detalha que, apesar de ter introduzido o *Framework* como um instrumento para avaliar especificamente a utilidade dos TEIs com base em sua fidelidade de constructo, usabilidade e acessibilidade, o instrumento é baseado no julgamento humano e depende do *design* dos TEIs sem qualquer consideração às propriedades psicométricas do item.

Após a revisão bibliográfica, o autor apresenta a metodologia utilizada para a realização da sua pesquisa, os resultados, faz discussões acerca dos achados e conclui afirmando que, com a aplicação do protocolo desenvolvido aos itens, revelou-se que a maioria dos TEIs examinados no estudo mostraram um CMV reduzido, indicando que, em um cenário da vida real, o formato de múltipla escolha seria favorecido para a maioria desses pares de itens. Embora os resultados para itens de arrastar e soltar como uma classe sejam limitados, eles sugerem que os itens com esse formato devem ser usados com moderação devido à sua falta de valor em comparação com os itens de múltipla escolha.

O quinto texto analisado foi *“Test Takers Response Tendencies in Alternative Item Formats: A Cognitive Science Approach”*, cuja autoria é de Moon, Keehner e Katz (2020). Os autores destacam que as avaliações em larga escala estão adotando novos formatos de itens e que estes são, por vezes, denominados de itens aprimorados por tecnologia ou itens inovadores. Tais itens possuem em sua estrutura imagens, tabelas ou, ainda, formas diferentes de o aluno responder, como arrastar e soltar em comparação com os tradicionais itens de múltipla escolha por exemplo. Destacam também que a transição para avaliações baseadas em computador e as muitas críticas a itens de múltipla escolha são dois dos principais motivos para o uso crescente de novos formatos de itens.

Baseando-se em autores como Bryant (2017), Huff e Sireci (2001) e Parshall e Harmes (2014), Moon, Keehner e Katz (2020) ponderam que esses formatos de itens fornecem potencial para melhorar a validade das inferências sobre os resultados das avaliações, ou, ainda, que eles podem introduzir uma variância no constructo-irrelevante devido à não familiaridade dos usuários das avaliações com tipos específicos de interfaces.

Dessa forma, para aproveitar os potenciais benefícios dos novos formatos de itens e minimizar seus efeitos negativos, os autores realizaram uma pesquisa sobre como os formatos de itens afetam a maneira como os participantes das avaliações pensam e respondem os itens.

Moon, Keehner e Katz (2020) analisaram a cognição dos alunos em dois formatos de itens: múltipla escolha e grade. No formato múltipla escolha, os alunos selecionam uma ou mais opções dentre uma lista de alternativas. Já no formato de grade, recebem um conjunto de alternativas apresentadas nas linhas de uma grade e selecionam células que se aplicam às opções dadas nas colunas de grade.

Mais especificamente, os autores pesquisaram como esses formatos de itens poderiam afetar a tendência dos alunos em responder em tom incerto, bem como se poderiam afetar as pontuações das avaliações e o tempo gasto para o desenvolvimento. Além de testar empiricamente essas questões, aplicaram ainda duas teorias da ciência cognitiva para fornecer uma caracterização mais detalhada dos processos cognitivos dos participantes do teste nesses formatos de itens: a teoria das possibilidades de Gibson (1979) – para explicar como esses formatos de itens poderiam levar os alunos a exibir diferentes tendências de resposta; e a teoria da detecção de sinal de Macmillan e Creelman (1991) – para examinar se esses formatos de itens afetariam a tendência de resposta dos alunos, suas capacidades de fazer julgamentos, ou ambos.

Participaram da pesquisa adultos entre 20 e 29 anos, alunos inscritos na plataforma *Amazon Mechanical Turk*⁴. Os critérios de participação foram aplicados porque os itens desenvolvidos para o presente estudo foram estimados no nível de dificuldade adequada para adultos com escolaridade universitária. Duzentos e noventa e dois participantes (idade média de 26,8 anos e 44% mulheres) completaram o estudo.

Os resultados da investigação demonstraram que os formatos dos itens podem afetar as tendências de resposta do participante da avaliação quando ele está com dúvidas, ou seja, sob incerteza. A interação entre formatos de itens e a pontuação indicam que os formatos de itens afetaram seleções afirmativas, em particular entre aqueles com o menor nível de conhecimento prévio do que os participantes de alto conhecimento prévio. Os resultados ainda sustentam que a presença de uma grade aumenta a seleção de opções afirmativas de alunos sob incerteza. Os alunos optaram por respostas mais afirmativas em itens de grade do que em itens de múltipla escolha, embora ambos os formatos tivessem opções de escolha não forçada. Os resultados também mostraram que os itens da grade de escolha forçada resultaram em maiores respostas afirmativas sob incerteza.

Os formatos de itens, no entanto, podem fornecer pistas diferentes sobre o que os alunos podem fazer quando não têm certeza. A presença de opções de escolha forçada sinaliza que uma decisão explícita entre Verdadeiro/Falso precisa ser tomada em resposta a um item. A presença de células vazias em uma grade, mesmo quando não há escolha forçada a ser feita, parece sinalizar que algumas dessas células precisam ser preenchidas, solicitando, assim, seleções em vez de omissões. O formato de múltipla escolha que não apresenta opções de escolha forçada, nem uma grade, pode sinalizar que a omissão é permitida.

Assim, Moon, Keehner e Katz (2020) concluem que os formatos de itens podem desempenhar um papel poderoso nas interações dos alunos com itens de avaliação, moldando sua percepção das possibilidades de ação.

Burga León (2019) em seu texto, *“Aplicaciones de la tecnología a la evaluación psicométrica”*, aponta a importância e a aplicabilidade da tecnologia e, principalmente, o uso do computador na avaliação psicométrica. Apesar de a avaliação psicométrica não estar em total consonância com o tema central deste estudo, as abordagens feitas pelo autor auxiliam no entendimento de conceitos e classificações que são fundamentais. Dessa forma, foi considerada nesta revisão.

Em relação a itens aprimorados por tecnologia, Burga León (2019) traz o conceito de Sireci e Zenisky (2006), que os denominaram como “itens com formatos inovadores”. Porém, em muitos casos, trata-se de formatos já pré-existentes que foram melhorados graças a tecnologia, permitindo a inclusão de aspectos como gráficos de alta qualidade, áudio, vídeo e hiperlinks. Assim Sireci e Zenisky (2006), após uma década da primeira denominação, promoveram, em 2016, uma atualização, passando para “itens aprimorados com tecnologia”. É importante ressaltar que esta não é a única nomenclatura pela qual os itens são conhecidos, pois também são chamados de facilitados ou habilitados pela tecnologia, bem como de inovadores computadorizados ou digitais.

Ademais, Burga León (2019), pautado em Sireci e Zenisky (2006), ressalta que a grande vantagem de usar a tecnologia fornecida por um computador é que,

graças a ele, é possível realizar atividades mais semelhantes às situações da vida real. Já com base em Bryant (2017), o autor apresenta algumas vantagens dos itens digitais e afirma que estes têm uma menor probabilidade de acerto por acaso (depende do tipo de item), além de favorecer o envolvimento dos alunos da avaliação com o item, uma vez que o consideram mais autêntico. Outro aspecto positivo destacado por Burga León (2019) baseia-se em Strain-Seymour, Way e Dolan (2009), cujo estudo aponta que, em geral, os itens digitais fornecem informações diagnósticas mais ricas e podem mostrar não apenas o produto, mas também o processo de resolução.

Nesse sentido, o autor apresenta também uma classificação dos itens digitais pautados em Parshall, Harmes, Davey e Pashley (2010), trazendo alguns exemplos de itens aprimorados por tecnologia, e ressalta que outros exemplos podem ser encontrados na pesquisa de Sireci e Zenisky (2016).

O próximo tema abordado por Burga León (2019) refere-se a bancos de itens, com ênfase nas grandes possibilidades que estes oferecem, não apenas em termos de serem um repositório organizado, mas em termos das diversas funcionalidades, vinculadas à elaboração, revisão e aplicação de itens (MUCKLE, 2016). Com isso, contribuirão para elaborar, de forma mais eficiente, os itens que compõem os diferentes instrumentos de medição.

O autor finaliza sua pesquisa descrevendo os testes adaptativos informatizados e destaca que a utilização de banco de itens calibrados torna uma aplicação de testes psicométricos mais eficiente, obtendo uma redução do erro de medição padrão. Isso é possível graças ao fato de que os computadores facilitam, entre outras coisas, o registro do tempo de reação, além da resposta emitida.

Para Burga León (2019), a transição para avaliações digitais e a introdução de novos tipos de itens também criaram oportunidades para a coleta de importantes dados – como os registros detalhados das interações do usuário com o sistema digital e os cronogramas desses eventos gerados pelo usuário ou servidor –, que não estão disponíveis em avaliações tradicionais impressas. Ademais, os dados produzidos no processo das avaliações educacionais em larga escala digitais oferecem a oportunidade de estudar os caminhos dos participantes para uma solução e inferir os processos cognitivos e metacognitivos em que se envolvem em um nível mais apurado que as respostas, por si sós, não poderiam revelar.

O último texto analisado nesta revisão foi *“Using process data to understand problem-solving strategies and processes for drag-and-drop items in a large-scale mathematics assessment”* de Jiang, Gong, Saldivia e Cayton-Hodges (2021). Os autores apresentam um estudo exploratório, no qual investigaram os processos de resolução de problemas e estratégias aplicados pelos alunos da quarta e oitava séries dos EUA durante a Avaliação Nacional do Progresso Educacional (NAEP) em 2017, analisando os dados coletados durante suas interações com dois itens aprimorados por tecnologia (um item para cada série) do tipo arrastar e soltar na área da matemática. Especificamente, foram desenvolvidas medidas a partir dos dados tratados/processados para caracterizar as estratégias de resolução de problemas dos alunos e sua alocação de tempo durante os processos de resposta.

Os autores iniciam o texto ressaltando que o rápido avanço tecnológico contribui para o processo de ensino da matemática, bem como se baseiam em diversos outros estudos para afirmar que, com a crescente importância da tecnologia educacional, as avaliações em larga escala começaram a passar de avaliações tradicionais impressas para avaliações digitais (HEI; BORGONOVÍ; PACCAGNELLA, 2019; SCALISE; GIFFORD, 2006; ZENISKY; SIRECI 2002), citando como exemplo a NAEP.

A NAEP iniciou as avaliações com uso de dispositivos tecnológicos no ano de 2017, quando da transição do impresso para o digital. A utilização do digital oferece oportunidades de inovações através da introdução de novos tipos de itens interativos e aprimorados por tecnologia, bem como ferramentas matemáticas. Para ilustrar tal afirmação, Jiang, Gong, Saldivia e Cayton-Hodges (2021) apresentam alguns tipos de itens aprimorados por tecnologia, tais como múltipla escolha, arrastar e soltar, zonas, grade e escolha na linha (*inline*), além de ferramentas matemáticas como, régua interativa, calculadora digital, diagrama de caixas, bloco de anotações (*Scratchpad*) e editor de equações, que foram utilizadas na avaliação da NAEP em 2017 para o público investigado na pesquisa.

Os autores mencionam que os itens de arrastar e soltar têm sido cada vez mais utilizados em avaliações digitais. Ademais, afirmam que, em comparação com os itens convencionais de múltipla escolha, estes têm sido usados para reduzir o efeito da adivinhação aleatória, fortalecer a medição e melhorar o engajamento e a motivação dos participantes das avaliações. Pautados em Arslan *et al.* (2020), Bryant (2017) e Scalise e Gifford (2006), atestam que os itens do tipo arrastar e soltar possuem potencial para melhor representar habilidades relevantes do constructo relacionadas à correspondência, categorização, (re)ordenação/(re)organização e sequenciamento. Nesse sentido, relatam ainda que os dados resultantes do processo de resolução dos itens de arrastar e soltar são ricos, pois incluem registros detalhados das interações dos alunos com o sistema, como suas ações de respostas e os cronogramas dessas ações.

Para Jiang, Gong, Saldivia e Cayton-Hodges (2021), um dos objetivos importantes para a educação matemática no nível K-12 é ajudar os alunos a desenvolverem conhecimentos e habilidades necessárias para a resolução de problemas matemáticos. Além disso, há evidências extensas de que indivíduos que aplicam estratégias eficientes de resolução de problemas são mais propensos a ter sucesso em tarefas de desempenho e aprendizagem.

Apesar da eficácia das metodologias e das medidas tradicionais, estas podem não refletir necessariamente formas autênticas de resolução de problemas matemáticos. Os dados coletados de avaliações em larga escala digitais, por outro lado, fornecem informações refinadas sobre como os alunos planejam, selecionam e executam várias estratégias de resolução de problemas para encontrar uma solução e como eles monitoram e refletem sobre sua resposta de forma discreta e maneira escalável.

Os dados utilizados na pesquisa foram coletados de uma amostra representativa de estudantes do quarto e do oitavo anos nos EUA que fizeram a avaliação matemática do NAEP em 2017 e completaram dois itens selecionados: o item G4, que avaliou os conhecimentos e habilidades dos alunos sobre a área de conteúdo matemático “propriedades numéricas e operações”; e o item G8, que

solicitava aos participantes organizarem um determinado conjunto de dígitos para produzir dois fatores que se multiplicam a um determinado produto.

Nesse sentido, os autores desenvolveram uma lista de medidas a partir dos dados coletados para inferir os processos e estratégias de resolução de problemas dos alunos. Essas medidas foram combinadas com as pontuações de resultado dos itens para entender os vários processos cognitivos e metacognitivos. Ademais, foram analisadas as sequências de ações executadas pelos alunos quando responderam a cada item, o comportamento de mudança de respostas realizadas pelos alunos e o tempo utilizado para responder cada item.

Como resultados, Jiang, Gong, Saldivia e Cayton-Hodges (2021) ressaltam que os participantes que alcançaram um nível mais alto de precisão em um item aplicaram estratégias de resolução de problemas que eram mais eficientes ao responder aos itens de arrastar e soltar. Os alunos também passaram mais tempo se engajando em comportamentos metacognitivos, como revisar uma solução enviada anteriormente. Ao contrário, os alunos com pontuação mais baixa tenderam a usar estratégias que são menos eficientes ao tentar resolver o item e passaram menos tempo se engajando em comportamentos de monitoramento metacognitivo.

Por fim, ressaltam ainda que os resultados também podem ser valiosos para a prática educacional, pois podem auxiliar no processo de tomada de decisão dos desenvolvedores de avaliações ao projetar itens digitais. Ou seja, os desenvolvedores devem ser encorajados a propor e usar tipos de itens mais interativos e tecnológicos, incluindo – mas não se limitando a eles – itens de arrastar e soltar para fazer uso completo dos dados e inferir melhor os processos e habilidades de resolução de problemas matemáticos.

4 Considerações finais

Este estudo teve como objetivo investigar e apresentar características, formatos, técnicas e métodos utilizados para o planejamento e desenvolvimento de itens digitais ou aprimorados por tecnologias no contexto de avaliações em larga escala, no âmbito nacional e internacional, já explorados na literatura.

Para atingir o intuito pretendido, técnicas de revisão sistemática e bibliométrica da literatura foram empregadas, de modo que as buscas se deram em duas bases de publicações no período de 2000 a 2021.

A análise bibliométrica encontrou algumas particularidades acerca dos documentos selecionados, como a pouca quantidade de publicações abordando o assunto do estudo, porém com uma tendência de crescimento a partir do ano de 2016. Tal tendência é motivada pela evolução das tecnologias e seu uso nas avaliações em larga escala e no desenvolvimento de itens digitais com formatos inovadores ocorridos nos últimos anos. Quanto à origem das publicações, observa-se a concentração nos Estados Unidos (EUA), sendo apenas um originário da América Latina.

Os documentos selecionados foram publicados quase na totalidade em periódicos de médio fator de impacto. No entanto, destaca-se a notável incidência de citações para documento de Bryant (2017), demonstrando a relevância e o potencial que possui a pesquisa envolvendo a temática abordada. Entretanto, é

importante ressaltar que o desenvolvimento de novas pesquisas, com citações recentes, envolve um tempo de, no mínimo, um ou dois anos para publicações. Dessa forma, não podemos aqui desconsiderar a relevância de tais documentos apenas por não possuírem citações.

Já em relação aos termos utilizados nos títulos, resumos e palavras-chave dos documentos, aqueles exibidos com maior ocorrência foram “*items*”, “*item*”, “*TEIs*”, “*formats*”, “*technology-enhanced*” e “*tests*”. Durante o estudo, foram observadas variadas formas de escrita utilizadas para se referir a itens digitais, tais como *Technology-Enhanced Items*, *TEIs*, *innovative items*, *digital items*, *novel item formats*, o que denota uma falta de padrão para o termo. Entre os formatos, *Technology-Enhanced Items* é o mais aplicado pelos autores.

Com fundamento na análise sistemática, alguns pontos comuns foram averiguados em quase a totalidade das publicações. Em geral, os documentos relatam que, com a evolução tecnológica na área educacional, as avaliações em larga escala estão em transição do impresso para o digital, e, com isso, surgem os novos formatos de itens aprimorados por tecnologia.

Um segundo ponto em comum são as críticas a aspectos das organizações de avaliações, tais como: a adaptação ou replicação dos itens tradicionais usados em avaliações impressas para uma plataforma digital; as avaliações que possuem apenas formatos de itens convencionais do tipo múltipla escolha; a incorporação de itens digitais sem a clara compreensão de suas diferenças, implicações, medições, custo-benefício ou efeitos sobre os participantes das avaliações; e a posição crítica aos desenvolvedores de itens, que acreditam que os itens digitais, por si sós, agregam valor.

Em relação à categorização e classificação de formatos de itens digitais, observou-se que a maioria toma como base aquela apresentada por Bennett (1993), com algumas alterações e formas diferentes de representar os formatos de itens. Com a crescente implementação de avaliações digitais, há diferentes formatos de itens que agregam cada categoria apresentada pela literatura, ampliando a possibilidade de propor e criar novos formatos a partir dos já elencados.

Devido ao foco deste estudo estar nas características, nos formatos, técnicas e métodos para o planejamento e desenvolvimento de itens digitais e pelo fato de a literatura existente abordar o tema de forma adjacente, considera-se o resultado com algumas limitações. Porém, ainda assim, a pesquisa empreendida pode ser considerada de relevada importância como embasamento para o planejamento e novas propostas de *designs* de itens digitais interativos e inovadores nas diversas áreas do conhecimento.

Agradecimentos

Agradeço à Fundação Centro de Políticas Públicas e Avaliação da Educação (Fundação CAEd) pela concessão da Bolsa de Pesquisa que possibilitou a realização desse estudo.

Referências

ARSLAN, Burcu; JIANG, Yang; KEEHNER, Madeleine; GONG, Tao; KATZ, Irvin R.; YAN, Fred. The Effect of Drag-and-Drop Item Features on Test-Taker Performance and Response Strategies. **Educational measurement: issues and practice**, v. 39, n. 2, p. 96-106, 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/emip.12326>. Acesso em: 12 ago. 2021.

BENNETT, Randy Elliot. On the meaning of constructed response. In BENNETT, Randy Elliot; WARD, W.C. (Eds.) **Construction versus choice in cognitive measurement: issues in constructed response, performance testing, and portfolio assessment**. Hillsdale, NJ: Lawrence Erlbaum Associates. 1993. p. 1-27.

BOYLE, Andrew, HUTCHINSON, Dougal. Sophisticated tasks in e-assessment: what are they and what are their benefits? **Assessment & evaluation in higher education**, v. 34 n. 3, p. 305- 319, 2009.

BRAUN, Tibor; GLÄNZEL, Wolfgang; SCHUBERT, Andras. A Hirsch-type index for journals. **Scientometrics**, v. 69, n. 22, p. 169-173, 2006. Disponível em: https://www.researchgate.net/publication/220365396_A_Hirsch-type_index_for_journals. Acesso em: 20 out. 2021.

BRYANT, William. Developing a Strategy for Using Technology-Enhanced Items in Large-Scale Standardized Tests. **Practical assessment, research, and evaluation**, v. 22, 2017.

BRUNO, Adriana Rocha. **Relatório de atividades de pesquisa em avaliação 2018-2019/outubro**. CAEd, 2019

BURGA LEON, Andrés Alberto. Aplicaciones de la tecnología a la evaluación psicométrica. **Propós. represent.** [online], v.7, p. 318-318, 2019. Disponível em: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2307-79992019000400006&lng=es&nrm=iso. Acesso em: 20 jul. 2021.

CRABTREE, Ashleigh. R. **Psychometric properties of technology-enhanced item formats: An evaluation of construct validity and technical characteristics**. 2016. Doctoral dissertation, 181 The University of Iowa. The University of Iowa's Institutional Repository. Disponível em: <https://iro.uiowa.edu/esploro/outputs/doctoral/Psychometric-properties-of-technology-enhanced-item-formats/9983777220902771>. Acesso em: 30 jul. 2021.

FALAGAS, Matthew E.; KOURANOS, Vasilios D.; ARENCIBIA-JORGE, Ricardo; KARAGEORGOPOULOS, Drosos E. Comparison of SCImago journal rank indicator with journal impact factor. **The FASEB journal**, v. 22, n. 8, p. 2623-2628, 2008.

GIBSON, James J. **The ecological approach to visual perception**. Boston, MA: Houghton-Mifflin. 1979.

HALADYNA, Thomas; RODRIGUEZ, Michael. **Developing and validating Test Items**. New York, NY: Routledge. 2013. Disponível em: https://www.researchgate.net/publication/346346355_Developing_and_Validating_Test_Items. Acesso em: 25 jul 2021.

HEI, Qiwei, BORGONOV, Francesca; PACCAGNELLA, Marco. Using process data to understand adults' problem-solving behaviour in the Programme for the International Assessment of Adult Competencies (PIAAC): Identifying generalised patterns across multiple tasks with sequence mining. **OECD education working papers**, OECD Publishing, Paris, n. 205, 2019. Disponível em: https://www.oecd-ilibrary.org/education/using-process-data-to-understand-adults-problem-solving-behaviour-in-the-programme-for-the-international-assessment-of-adult-competencies-piaac_650918f2-en. Acesso em: 30 jul. 2021.

HUFF, Kristen L.; SIRECI, Stephen G. Validity issues in computer-based testing. **Educational measurement: issues and practice**, v. 20, n. 3, p. 16-25, 2001. Disponível em: https://www.researchgate.net/publication/229806166_VValidity_Issues_in_Computer-Based_Testing. Acesso em: 30 jul. 2021.

JIANG, Yang; GONG, Tao; SALDIVIA, Luis E.; CAYTON-HODGES, Gabrielle; AGARD, Christopher. Using process data to understand problem-solving strategies and processes for drag-and-drop items in a large-scale mathematics assessment. **Large-scale assessments in education**, v. 9, n. 2, 2021. Disponível em: <https://largescaleassessmentsineducation.springeropen.com/articles/10.1186/s40536-021-00095-4#citeas>. Acesso em: 20 jul. 2021.

JODOIN, Michael G. Measurement Efficiency of Innovative Item Formats in Computer-Based Testing. **Journal of educational measurement**, v. 40, n. 1, p. 1-15, 2003. Disponível em: <http://www.jstor.org/stable/1435051>. Acesso em: 20 dez. 2021.

KANE, M. Content-related validity evidence in test development. In: DOWNING, Steven M.; HALADYNA, Thomas M. (Eds.) **Handbook of test development**. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2006. p. 131-154.

KITCHENHAM, Barbara. Procedures for performing systematic reviews. **Keele university**, v. 33, p. 1-26, 2004.

KOCH, D.A. Testing goes graphical. **Journal of interactive instruction development**, v. 5, p. 14-21, 1993.

MACMILLAN, Neil A.; CREELMAN, C. Douglas. **Detection theory: A user's guide**. New York, NY: Cambridge University Press. 1991. Disponível em: <https://psycnet.apa.org/record/1991-97801-000>. Acesso em: 30 jul. 2021.

MONCALEANO, Sebastian. **Examining the comparative measurement value of technology-enhanced Items**. 2021. 261 p. PhD thesis. Boston College Lynch School. 2021. Disponível em: <https://dlib.bc.edu/islandora/object/bc-ir:109086>. Acesso em: 20 jul. 2021.

MOON, Jung A.; KEEHNER, Madeleine; KATZ, Irvin R. **Test takers response tendencies in alternative item formats: a cognitive science approach**. *educational assessment*, v. 25, n. 3, p. 236-250, 2020. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/10627197.2020.1804350>. Acesso em: 20 jul. 2021.

MUCKLE, Timothy J. Web-based item development and banking. V In: LANE, Suzanne; RAYMOND, Mark R.; HALADYNA, Thomas M. (org.). **Handbook of test development**. 2. ed. New York, NY: Routledge, 2016. p. 241-258.

PARSHALL, Cynthia G.; DAVEY, Tim; PASHLEY, Peter J. Innovative item types for computerized testing. In LINDEN, Wim J. van der; GLAS, Cees A.W. (Eds.), **Computerized adaptive testing: theory and practice**. Kluwer Academic Publishers. 2000. p. 129-148.

PARSHALL, Cynthia G.; HARMES, J. Christine. Improving the quality of innovative item types: Four tasks for design and development. **Journal of applied testing technology**, v. 10, n. 1, p. 1-20, 2014.

PARSHALL, Cynthia G.; HARMES, J. Christine; DAVEY, Tim; PASHLEY, Peter J. Innovative items for computerized testing. In LINDEN, Wim J. van der; GLAS, Cees A.W. (Eds.) **Computerized adaptive testing: theory and practice**. 2. ed. Norwell, MA: Kluwer Academic Publishers. 2010. Disponível em: https://link.springer.com/chapter/10.1007/978-0-387-85461-8_11. Acesso em: 21 dez. 2021.

PARSHALL, Cynthia G.; STEWART, Rob; RITTER, Judy. Innovations: Sound, graphics, and alternative response modes. [Paper Presentation]. **National council on measurement in**

education annual meeting, New York, EUA. 1996. Disponível em: <https://files.eric.ed.gov/fulltext/ED421524.pdf>. Acesso em: 22 jul. 2021.

PAULA, Samantha; ARAÚJO, Marco Antônio; SILVA, Júlio César da. Pesquisa científica baseada em uma revisão sistemática da literatura. **RECM - Revista de educação, ciências e matemática**, v. 6, n. 2. 2016. Disponível em <http://publicacoes.unigranrio.edu.br/index.php/recm/article/view/4058>. Acesso em: 20 dez. 2021.

QIAN, Hong; WOO, Ada; KIM, Doyoung. Exploring the psychometric properties of innovative items in computerized adaptive testing. In JIAO, Hong; LISSITZ, Robert W. (Eds.), **Technology enhanced innovative assessment: development, modeling, and scoring from an interdisciplinary perspective**. Information Age Publishing. 2017. p. 97-118.

RUSSELL, Michael. A Framework for Examining the Utility of Technology-Enhanced Items. **Journal of applied testing technology**, v. 17, n. 1, p. 20-32, 2016. Disponível em: <http://jattjournal.net/index.php/atp/article/view/89189>. Acesso em: 20 jul. 2021.

SAMPAIO, Rosana Ferreira; MANCINI, Marisa Cotta. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Revista brasileira de fisioterapia**, v. 11, p. 83-89, 2007.

SANTOS, Raimundo Nonato Macedo dos. Produção científica: por que medir? o que medir? **Revista digital de biblioteconomia e ciência da informação**, v. 1, n. 1, p. 22-38, 2003. Disponível em: <https://brapci.inf.br/index.php/res/v/39961>. Acesso em: 21 dez. 2021.

SCALISE, Kathleen; GIFFORD, Bernard. Computer-based assessment in E-learning: A framework for constructing “intermediate constraint” questions and tasks for technology platforms. **The journal of technology, learning, and assessment**, v. 4, n. 6, p. 1-43, 2006. Disponível em: <https://ejournals.bc.edu/index.php/jtla/article/view/1653>. Acesso em: 20 jul. 2021.

SCORTEGAGNA, Liamara. **Relatório de atividades de pesquisa em avaliação 2020-2021/março**. Juiz de Fora: CAEd, 2020.

SIRECI, Stephen G.; ZENISKY, April L. Innovative Item Formats in Computer-Based Testing: In Pursuit of Improved Construct Representation. In DOWNING, Steven M.; HALADYNA, Thomas M. (Eds.), **Handbook of test development**. Lawrence Erlbaum Associates Publishers, 2006. p. 329-347.

SIRECI, Stephen G.; ZENISKY, April L. Computerized innovative item formats: Achievement and Credentialing. In: LANE, Suzanne; RAYMOND, Mark R.; HALADYNA, Thomas M. (org.). **Handbook of test development**. 2. ed. New York, NY: Routledge, 2016. p. 313-334.

STRAIN-SEYMOUR, Ellen; WAY, Walter; DOLAN, Robert P. **Strategies and processes for developing innovative items in large-scale assessments**. 2009. Disponível em: https://www.researchgate.net/publication/349104257_Strategies_and_Processes_for_Developing_Innovative_Items_in_Large-Scale_Assessments. Acesso em: 30 jul. 2021.

TARRANT, Marie; KNIERIM, Aimee; HAYES, Sasha.K.; WARE, James. The frequency of item writing flaws in multiple-choice. **Nurse education today**, v. 26, n. 8, p. 662-671, 2006. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0260691706001067>. Acesso em: 20 dez. 2021.

WAN, Lei; HENLY, George A. Measurement properties of two innovative item formats in a computer based test. **Applied measurement in education**, v. 25, n. 1, p. 58-78, 2012.

WISE, Steven L.; SOLAND, James; DUPRAY, Laurence M. The Impact of Technology-Enhanced Items on Test-Taker Disengagement. **Journal of applied testing technology**, v. 22, n.1, p. 28-36, 2021. Disponível em: <https://eric.ed.gov/?id=EJ1296001>. Acesso em: 20 jul. 2021.

ZENISKY, April L; SIRECI, Stephen G. Technological innovations in large-scale assessment. **Applied measurement in education**, v.15, n. 4, p. 337-362, 2002. Disponível em: https://www.researchgate.net/publication/248580109_Technological_Innovations_in_Large-Scale_Assessment. Acesso em: 08 ago. 2021.

Notas

¹ <https://jcr-clarivate.ez25.periodicos.capes.gov.br/jcr/home>

² <https://www.scimagojr.com/>

³ Disponível em: <https://www.jasondavies.com/wordcloud/#%2F%2Fwww.jasondavies.com%2Fwordcloud%2Fabout%2F>.

⁴ É uma plataforma de *crowd-sourcing* na qual trabalhadores adultos podem realizar tarefas online e recebem pagamento.

Informações complementares

Financiamento

Fundação Centro de Políticas Públicas e Avaliação da Educação (Fundação CAEd)

Contribuição de autoria

Concepção e elaboração do manuscrito: Liamara Scortegagna

Coleta de dados: Liamara Scortegagna

Análise de dados: Liamara Scortegagna

Discussão dos resultados: Liamara Scortegagna

Revisão e aprovação: Liamara Scortegagna

Preprint, originalidade e ineditismo

O artigo é original, inédito e não foi depositado como *preprint*.

Verificação de similaridades

O artigo foi submetido ao iThenticate, em 6 de dezembro de 2023, e obteve um índice de similaridade compatível com a política antiplágio da revista Pesquisa e Debate em Educação.

Consentimento de uso de imagem

Não se aplica.

Aprovação de Comitê de Ética em Pesquisa

Não se aplica.

Conflito de interesse

Não há conflitos de interesse.

Conjunto de dados de pesquisa

Não há dados disponibilizados. Observa-se que a revista Pesquisa e Debate em Educação estimula a publicação dos dados da pesquisa. Os dados podem ser disponibilizados em

bancos ou repositórios próprios para tal finalidade ou, até mesmo, na nuvem (com acesso público).

Licença de uso

Os autores cedem à Revista Pesquisa e Debate em Educação os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution \(CC BY\) 4.0 International](#). Esta licença permite que terceiros remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os autores têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

Publisher

Universidade Federal de Juiz de Fora (UFJF), Faculdade de Educação (FACED), Centro de Políticas Públicas e Avaliação da Educação (CAEd), Programa de Pós-Graduação Profissional em Gestão e Avaliação da Educação Pública (PPGP). Publicação no Portal de Periódicos da UFJF. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

Editores

Frederico Braida; Wagner Silveira Rezende.

Formato de avaliação por pares

Revisão duplamente cega (*Double blind peer review*).

Sobre os autores

Liamara Scortegagna

Graduada em Informática (UnC). Mestre em Ciência da Computação (UFSC). Doutora em Engenharia de Produção – Mídia e Conhecimento (UFSC). Professora Associada III do Departamento de Ciência da Computação (DCC) da Universidade Federal de Juiz de Fora (UFJF). Atua nos Programas de Pós-Graduação *Stricto Sensu* em Educação Matemática (PPGEM) e em Gestão e Avaliação da Educação Pública (PPGP) da UFJF. É pesquisadora da Fundação CAEd/UFJF, coordenadora do curso de Graduação em Licenciatura em Computação (LiCOMP/UFJF) e membro dos grupos de pesquisa: GRUPAR – Grupo de Pesquisa Aprendizagem em Rede, NIDEEM - Núcleo de investigação, Divulgação e Estudos em Educação Matemática - Tecnologias de Informação e Comunicação na Educação Matemática e LApIC - Laboratório de Aplicações e Inovação em Computação. É autora do livro: *Objetos de Aprendizagem* pela editora CEAD/UFJF.

Currículo Lattes: <http://lattes.cnpq.br/9104271477506670>