



AVALIAÇÃO FORMATIVA ASSISTIDA POR TECNOLOGIA: APLICATIVOS E PLATAFORMAS PARA FEEDBACK CONTÍNUO

TECHNOLOGY-ASSISTED FORMATIVE ASSESSMENT: APPS AND PLATFORMS FOR CONTINUOUS FEEDBACK

DOI: 10.5281/zenodo.20371513



Sergio Augusto Ferreira Domingues¹

RESUMO

Este artigo, configurado como revisão bibliográfica analítico-crítica, examina a tensão estrutural entre a racionalidade instrumental que organiza a oferta de tecnologias avaliativas e a racionalidade pedagógica dialógica que define a efetividade da avaliação formativa. Articulando referenciais internacionais consolidados (Black e Wiliam, Hattie, Sadler), produção crítica recente sobre IA em educação (Selwyn, Williamson, Luckin, Mollick) e a tradição brasileira de avaliação (Luckesi, Hoffmann), o trabalho percorre os fundamentos teóricos do feedback efetivo, sistematiza a taxonomia de ferramentas e analisa as transformações impulsionadas pela inteligência artificial generativa. Como contribuição teórica, propõe-se o framework da Arquitetura Pedagógica Formativa Tecnicamente Mediada (APFTM), que articula cinco dimensões interdependentes (diagnóstica, dialógico-temporal, de agência e autorregulação, ético-algorítmica e de reversibilidade pedagógica) pelas quais se avalia se determinada configuração tecnológica sustenta ou descaracteriza a operação formativa. Sustenta-se que tecnologias digitais operam como amplificadores condicionais da avaliação formativa, cuja potência depende da arquitetura pedagógica em que se inscrevem, sem substituir os processos dialógicos e relacionais que sustentam sua efetividade. Conclui-se que a pergunta pedagogicamente relevante não é "que ferramentas adotar", mas "que arquitetura torna determinada adoção tecnológica efetivamente formativa", reposicionamento relevante para o contexto educacional brasileiro.

¹ Licenciado em Pedagogia pela Faculdade Aliança do Maranhão – FAMAR (2025) e Especialista em Avaliação Educacional pela Faculdade dos Vales – FACUVALE (2026). Participa como mestrando no Programa *Academic Master's Degree In Educational Sciences and Christian Ethics* da Ivy Enber Christian University, com linha de pesquisa "Metodologias e Práticas da Educação Básica e do Ensino Superior". E-mail: cap.sergiodomingues@gmail.com





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

Palavras-chave: Avaliação formativa; Feedback; Tecnologias educacionais; Avaliação adaptativa; Inteligência artificial em educação.

ABSTRACT

This article, configured as an analytical-critical literature review, examines the structural tension between the instrumental rationality that organizes the supply of assessment technologies and the dialogical pedagogical rationality that defines the effectiveness of formative assessment. By articulating consolidated international frameworks (Black and Wiliam, Hattie, Sadler), recent critical scholarship on AI in education (Selwyn, Williamson, Luckin, Mollick), and the Brazilian tradition of educational assessment (Luckesi, Hoffmann), the study explores the theoretical foundations of effective feedback, systematizes the taxonomy of technological tools, and analyzes the transformations driven by generative artificial intelligence. As a theoretical contribution, the article proposes the framework of Technologically Mediated Formative Pedagogical Architecture (TMFPA), which integrates five interdependent dimensions (diagnostic, dialogical-temporal, agency and self-regulation, ethical-algorithmic, and pedagogical reversibility) through which it becomes possible to assess whether a given technological configuration sustains or distorts formative operation. The study argues that digital technologies function as conditional amplifiers of formative assessment, whose potential depends on the pedagogical architecture within which they are embedded, without replacing the dialogical and relational processes that sustain their effectiveness. It concludes that the pedagogically relevant question is not “which tools to adopt,” but rather “which architecture makes a given technological adoption effectively formative,” a repositioning of particular relevance to the Brazilian educational context.

Keywords: Formative assessment; Feedback; Educational technologies; Adaptive assessment; Artificial intelligence in education.

1. INTRODUÇÃO

A avaliação formativa emergiu, nas últimas décadas, como uma das práticas pedagógicas de maior potencial para melhoria da aprendizagem. Diferentemente da avaliação somativa, que ocorre ao final de períodos instrucionais com propósito primário de certificação, a avaliação formativa caracteriza-se pela coleta sistemática de evidências sobre a aprendizagem durante o processo educacional, com a finalidade explícita de orientar ajustes pedagógicos e apoiar o desenvolvimento estudantil. Como demonstra a meta-análise seminal de Black e Wiliam (1998), práticas efetivas de avaliação formativa podem produzir ganhos de aprendizagem substanciais, com tamanhos de efeito que se situam entre os mais elevados identificados em pesquisas educacionais.

Revista OWL Journal, Campina Grande - PB, v.4 n.5 (2026) - ISSN 2965-2634

A Revista OWL Journal está licenciada com uma Licença Creative Commons Atribuição (CC BY)





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

A implementação efetiva da avaliação formativa, contudo, apresenta desafios práticos importantes. A coleta frequente e sistemática de evidências sobre a aprendizagem de todos os estudantes, a análise oportuna dessas evidências, e a provisão de feedback específico e acionável demandam tempo e competências profissionais que nem sempre estão disponíveis em contextos educacionais reais. Como observam Wiliam e Thompson (2007), embora os princípios da avaliação formativa sejam amplamente aceitos, sua implementação consistente e sustentável permanece como desafio significativo para muitos sistemas educacionais.

Nesse contexto, tecnologias digitais emergem como recursos potencialmente valiosos para apoiar e amplificar práticas de avaliação formativa. Ferramentas digitais podem automatizar aspectos da coleta e análise de dados, facilitar a provisão de feedback imediato, possibilitar personalização de trajetórias de aprendizagem, e liberar tempo docente para interações pedagógicas mais complexas e significativas. Bennett (2011, p. 5) argumenta que *tecnologias oferecem oportunidades de transformar avaliação formativa de aspiração pedagógica em prática sistematicamente implementável*, desde que sejam utilizadas com compreensão adequada de seus potenciais e limitações.

Paralelamente, o desenvolvimento de aplicativos educacionais, plataformas de aprendizagem adaptativa e sistemas baseados em inteligência artificial tem expandido significativamente o repertório de ferramentas disponíveis para avaliação formativa. De sistemas simples de resposta estudantil (clickers) a plataformas sofisticadas que utilizam algoritmos de aprendizagem de máquina para diagnóstico personalizado e recomendação de atividades, o ecossistema de tecnologias para avaliação formativa tornou-se diversificado e em rápida evolução (SHUTE; RAHIMI, 2017).

No contexto brasileiro, essa discussão assume contornos particulares. A Base Nacional Comum Curricular (BRASIL, 2018) institui a avaliação formativa como princípio orientador da prática pedagógica, ao enfatizar o desenvolvimento de competências e habilidades em processo. Paralelamente, sistemas como o Sistema de Avaliação da Educação Básica (SAEB) e a Lei Geral de Proteção de Dados (Lei nº 13.709/2018) estabelecem, respectivamente, parâmetros de monitoramento educacional em larga escala e exigências legais para o

Revista *OWL Journal*, Campina Grande - PB, v.4 n.5 (2026) - ISSN 2965-2634

A Revista *OWL Journal* está licenciada com uma Licença Creative Commons Atribuição (CC BY)





tratamento de dados pessoais de estudantes, fatores que condicionam a adoção de tecnologias avaliativas em escolas brasileiras. Soma-se a esse cenário uma sólida tradição nacional de pesquisa em avaliação, com contribuições de autores como Luckesi (2011) e Hoffmann (2017), cujas formulações sobre avaliação como prática diagnóstica e mediadora dialogam com referenciais internacionais e enriquecem o debate sobre integração de tecnologias a práticas formativas.

Apesar dessa expansão, observa-se uma tensão estrutural no debate contemporâneo: as tecnologias avaliativas tendem a ser apresentadas sob lógica predominantemente instrumental, centrada em escalabilidade, automação e personalização algorítmica, ao passo que a avaliação formativa, em sua formulação teórica consolidada, opera sob lógica dialógica, mediadora e processual, fundada na relação intersubjetiva entre professor e estudante. Essa dissociação entre a racionalidade técnica que organiza a oferta de ferramentas e a racionalidade pedagógica que define a efetividade do feedback constitui o problema central deste artigo.

A partir desse problema, o presente trabalho se configura como revisão bibliográfica de natureza analítico-crítica, articulando literatura internacional consolidada (Black, Wiliam, Hattie, Sadler), formulações da tradição brasileira (Luckesi, Hoffmann) e produção recente sobre tecnologias educacionais e inteligência artificial generativa, com o objetivo de examinar criticamente como aplicativos e plataformas tecnológicas podem (e em que condições efetivamente conseguem) apoiar práticas de avaliação formativa. A tese sustentada é a de que tecnologias digitais ampliam significativamente a capacidade operacional da avaliação formativa em frequência, escala e imediatez do feedback, mas não substituem os processos pedagógicos, dialógicos e relacionais que historicamente sustentam sua efetividade; antes, sua incorporação produtiva exige subordinação a uma racionalidade pedagógica explícita, sob risco de converter a avaliação formativa em prática meramente diagnóstica de baixa potência transformadora.

Para sustentar essa tese, o artigo organiza-se em seções progressivamente mais analíticas: a seção 2 explicita o percurso metodológico que orientou a construção do corpus e a análise da literatura; a seção 3 estabelece os fundamentos teóricos que definem o que torna a





avaliação formativa efetiva; a seção 4 sistematiza, em chave crítica, a taxonomia de ferramentas disponíveis; a seção 5 examina as transformações em curso no campo da inteligência artificial, com ênfase na ruptura introduzida pelos grandes modelos de linguagem; a seção 6 propõe e formaliza um framework analítico para a Arquitetura Pedagógica Formativa Tecnicamente Mediada (APFTM), articulando cinco dimensões interdependentes pelas quais é possível avaliar se determinada adoção tecnológica sustenta ou descaracteriza a operação formativa; e a seção 7 problematiza limites e tensões persistentes que delimitam o alcance dessa integração.

2. PROCEDIMENTOS METODOLÓGICOS

Este estudo configura-se como revisão bibliográfica de natureza analítico-crítica, modalidade que se distingue tanto da revisão narrativa de caráter meramente descritivo quanto da revisão sistemática orientada por protocolos quantitativos de inclusão e exclusão (LIMA; MIOTO, 2007; WHITTEMORE; KNAFL, 2005). A escolha por essa modalidade decorre da natureza da pergunta de pesquisa que estrutura o trabalho: não se trata de mapear exaustivamente a produção empírica sobre tecnologias avaliativas, mas de examinar criticamente as articulações conceituais entre três tradições teóricas distintas (a tradição anglo-saxã da formative assessment, a tradição brasileira de avaliação mediadora e diagnóstica e a literatura recente sobre inteligência artificial em educação), à luz de uma tensão estrutural que se manifesta no campo. Para questões dessa natureza, a opção metodológica adequada é a hermenêutico-dialética, conforme formulada por Minayo (2014), por permitir o tratamento integrado de aspectos espaço-temporais, racionalidade das fontes e intencionalidades subjacentes às formulações teóricas em diálogo.

A constituição do corpus bibliográfico seguiu três etapas sucessivas, cujo mapeamento quantitativo aproximado é sintetizado no Quadro 1 ao final desta seção. Na primeira etapa, procedeu-se à identificação de obras fundacionais sobre avaliação formativa e feedback efetivo, a partir de levantamento em bases de dados internacionais (Scopus, ERIC e Web of





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

Science) e nacionais (SciELO, Portal de Periódicos da CAPES e Biblioteca Digital Brasileira de Teses e Dissertações, BDTD), utilizando os descritores em inglês “formative assessment”, “assessment for learning”, “feedback” e “educational technology”, e em português “avaliação formativa”, “avaliação mediadora”, “feedback” e “tecnologias educacionais”, combinados pelos operadores booleanos AND e OR. As buscas iniciais retornaram aproximadamente 240 referências potencialmente relevantes. Aplicaram-se como critérios de inclusão: trabalhos publicados em periódicos revisados por pares, capítulos de obras de referência reconhecidas no campo e documentos institucionais de organismos com autoridade pública (UNESCO, MEC, Cetic.br). Como critérios de exclusão, descartaram-se trabalhos não revisados por pares, materiais de divulgação sem ancoragem científica e estudos cuja análise se restringia a contextos particulares sem articulação com debates conceituais mais amplos. Após triagem por leitura de títulos e resumos, restaram aproximadamente 60 obras analisadas em profundidade.

Na segunda etapa, em razão da centralidade conferida à inteligência artificial generativa, complementou-se o corpus com produção específica sobre Grandes Modelos de Linguagem em educação, publicada entre 2023 e 2026, recuperada nas mesmas bases. O recorte temporal pós-2022 não se justifica por presunção de obsolescência da produção anterior, mas pela transformação do regime de problematização do campo: a partir do lançamento público do ChatGPT, em novembro de 2022, as perguntas pedagógicas relevantes sobre feedback automatizado deslocaram-se de questões sobre análise estatística da escrita (predominantes na produção de processamento de linguagem natural até 2022) para questões sobre feedback dialógico gerado por modelos linguísticos de uso geral. A produção de 2020 a 2022 sobre PLN educacional permanece teoricamente relevante e foi incorporada ao corpus por duas vias complementares: pelas referências cruzadas dos trabalhos pós-2022 (que retomam contribuições anteriores como base) e pelo rastreamento de referências descrito a seguir, evitando-se assim sua exclusão por mero critério cronológico. Para os fundamentos teóricos da avaliação formativa e do feedback, não se aplicou recorte temporal restritivo, dada a natureza fundacional de obras como as de Scriven (1967), Bloom (1968), Sadler (1989),

Revista *OWL Journal*, Campina Grande - PB, v.4 n.5 (2026) - ISSN 2965-2634

A Revista *OWL Journal* está licenciada com uma Licença Creative Commons Atribuição (CC BY)





Black e Wiliam (1998), Hattie e Timperley (2007), Luckesi (2011) e Hoffmann (2017), cuja relevância permanece atual no debate contemporâneo.

Na terceira etapa, procedeu-se ao rastreamento de referências cruzadas (snowballing), técnica de busca reversa que permitiu incorporar contribuições relevantes citadas pelas obras inicialmente selecionadas, acrescentando aproximadamente 25 obras ao corpus. Adotou-se como critério de parada a saturação argumentativa: o rastreamento foi encerrado quando novas referências passaram a reiterar argumentos já presentes no corpus, sem aportar elementos teóricos ou empíricos adicionais à discussão. Para mitigar o risco de viés de citação característico de buscas reversas em cadeia, limitou-se o rastreamento a dois níveis de profundidade (referências das referências), e cada obra incorporada foi avaliada autonomamente quanto a sua relevância argumentativa, e não por sua frequência de citação. Reconhece-se, contudo, como limitação metodológica do procedimento, a tendência inerente do snowballing a reforçar redes citatórias já estabelecidas, o que pode subrepresentar tradições teóricas periféricas em relação ao mainstream anglófono. Essa limitação foi parcialmente compensada pela busca direta em bases nacionais (SciELO, BDTD), que permitiu incorporar produção brasileira que de outro modo permaneceria invisível ao rastreamento de referências internacionais.

A análise do corpus organizou-se em três movimentos articulados, em consonância com a tradição hermenêutico-dialética. O primeiro movimento, descritivo, consistiu em identificar e sistematizar as formulações conceituais sobre avaliação formativa, feedback e tecnologias educacionais presentes nas obras selecionadas, atendo-se à racionalidade interna de cada tradição teórica em seu contexto de produção. O segundo movimento, comparativo, voltou-se à identificação de convergências, divergências e tensões entre as tradições internacionais e brasileira de pesquisa em avaliação, bem como entre as formulações pedagógicas consolidadas e as proposições recentes sobre IA em educação. O terceiro movimento, interpretativo, articulou os achados em uma síntese crítica, organizada em torno da tese central do trabalho, com atenção especial à tensão estrutural entre racionalidades instrumental e pedagógico-dialógica. Como reconhecem Lima e Mioto (2007), pesquisa





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

bibliográfica não consiste em mera compilação, mas em movimento de apreensão, questionamento e interlocução crítica com o material, no qual o objeto de estudo é constantemente revisto à medida que o pesquisador o aprofunda. Assume-se, portanto, que os resultados aqui apresentados não esgotam o tema, mas constituem síntese interpretativa orientada pela pergunta de pesquisa formulada, passível de validação e reformulação por outros pesquisadores que mobilizem corpus semelhante e método análogo.

Quadro 1 – Síntese da constituição do corpus bibliográfico

Etapas	Procedimento	Resultado aproximado
1. Busca por descritores	Levantamento em Scopus, ERIC, Web of Science, SciELO, Periódicos CAPES e BDTD	~240 referências
2. Triagem inicial	Aplicação dos critérios de inclusão e exclusão por leitura de títulos e resumos	~60 obras retidas
3. Recorte temático sobre LLMs	Complementação com produção pós-2022 sobre IA generativa em educação	~15 obras adicionadas
4. Snowballing	Rastreamento de referências cruzadas (até 2 níveis), com critério de saturação argumentativa	~25 obras incorporadas
Corpus final analisado	Obras submetidas aos três movimentos analíticos (descritivo, comparativo e interpretativo)	~100 obras

Fonte: Elaborado pelo autor (2026).

3. FUNDAMENTOS TEÓRICOS DA AVALIAÇÃO FORMATIVA

A avaliação formativa possui raízes conceituais que remontam aos trabalhos de Scriven (1967) e Bloom (1968), mas sua formulação contemporânea deve muito às contribuições de pesquisadores como Black, Wiliam, Hattie e Sadler. Compreender os fundamentos teóricos dessa abordagem é essencial para avaliar criticamente como e quando tecnologias podem efetivamente apoiar suas práticas.

3.1 Conceituação e Elementos Essenciais

Black e Wiliam (1998, p. 2) definem avaliação formativa como *todas as atividades realizadas por professores e/ou por seus alunos, que fornecem informação a ser usada como*

Revista *OWL Journal*, Campina Grande - PB, v.4 n.5 (2026) - ISSN 2965-2634

A Revista *OWL Journal* está licenciada com uma **Licença Creative Commons Atribuição (CC BY)**





feedback para modificar as atividades de ensino e aprendizagem nas quais estão engajados. Essa definição enfatiza três aspectos centrais: a coleta de informação sobre aprendizagem, o uso dessa informação como feedback, e a modificação de práticas pedagógicas baseada nesse feedback.

Wiliam e Thompson (2007) elaboram essa conceituação identificando cinco estratégias-chave de avaliação formativa: clarificação e compartilhamento de intenções de aprendizagem e critérios de sucesso com estudantes; engendramento de discussões, tarefas e atividades que evidenciem aprendizagem; provisão de feedback que mova os aprendizes adiante; ativação dos estudantes como recursos instrucionais uns dos outros; e ativação dos estudantes como donos de sua própria aprendizagem. Essas estratégias destacam que avaliação formativa transcende a mera aplicação de testes, envolvendo transformações em culturas de sala de aula e papéis de professores e estudantes.

No Brasil, essa conceituação dialoga com tradição consolidada de pesquisa em avaliação educacional. Luckesi (2011) propõe a noção de avaliação diagnóstica como prática inclusiva, voltada à compreensão dos processos de aprendizagem e à tomada de decisões pedagógicas, em oposição a modelos classificatórios e excludentes. De forma complementar, Hoffmann (2017) desenvolve o conceito de avaliação mediadora, que enfatiza o caráter dialógico e processual da avaliação, valorizando ritmos individuais de aprendizagem e a relação intersubjetiva entre professor e estudante. Essas formulações, embora desenvolvidas em referenciais teóricos distintos dos anglo-saxões, convergem com a literatura internacional ao concebê-la não como mensuração terminal, mas como dispositivo a serviço da aprendizagem, oferecendo base sólida para a integração de tecnologias avaliativas em contextos brasileiros.

3.2 Feedback: Componente Central da Avaliação Formativa

O feedback constitui elemento central de processos de avaliação formativa. Hattie e Timperley (2007) propõem modelo influente que identifica quatro níveis nos quais feedback





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

pode operar: nível da tarefa (correção de respostas específicas), nível do processo (estratégias e processos utilizados), nível de autorregulação (automonitoramento e ajuste metacognitivo), e nível do self (elogios pessoais gerais). Os autores argumentam que feedback é mais efetivo quando focado nos níveis de processo e autorregulação, em contraste com feedback meramente corretivo ou centrado no self.

Sadler (1989) oferece análise complementar, argumentando que feedback efetivo deve auxiliar estudantes a: compreender o que constitui bom desempenho (o objetivo); identificar onde estão atualmente em relação ao objetivo (a lacuna); e entender como fechar essa lacuna (as ações necessárias). Essa conceituação enfatiza que feedback não é mera provisão de informação pelo professor, mas processo que deve envolver ativamente o estudante na compreensão de seu próprio aprendizado e na tomada de ações para seu avanço.

Crucialmente, como observa Nicol (2010), feedback é efetivo não porque é dado, mas porque é usado. Pesquisas documentam que estudantes frequentemente não compreendem, não concordam com, ou simplesmente ignoram feedback recebido. Portanto, práticas formativas efetivas devem não apenas prover feedback, mas também criar condições para que estudantes se engajem produtivamente com ele, o que pode incluir oportunidades de revisão, discussão sobre feedback, e aplicação de sugestões em trabalhos subsequentes.

3.3 Evidências de Efetividade

Meta-análises e revisões sistemáticas têm consistentemente demonstrado impactos positivos substanciais de práticas de avaliação formativa sobre aprendizagem. A revisão de Black e Wiliam (1998) identificou ganhos de aprendizagem equivalentes a 0.4 a 0.7 desvios-padrão - entre os maiores efeitos documentados em intervenções educacionais. Hattie (2009), em síntese de múltiplas meta-análises, reporta tamanho de efeito médio de 0.79 para feedback, novamente situando-o entre os fatores de maior influência sobre aprendizagem.





Importantemente, pesquisas sugerem que benefícios de avaliação formativa não se distribuem uniformemente. Kingston e Nash (2011), em meta-análise de 42 estudos, encontraram que práticas formativas tendem a ser especialmente benéficas para estudantes de menor desempenho inicial, sugerindo potencial para redução de desigualdades. No entanto, os autores também identificam variabilidade substancial em efeitos, indicando que a qualidade de implementação - não apenas a presença nominal de práticas formativas - é determinante de resultados.

A leitura conjunta dessas contribuições permite extrair uma conclusão de natureza propositiva, central para a argumentação subsequente: aquilo que confere efetividade à avaliação formativa não é a coleta de evidências em si, mas a articulação entre três operações pedagógicas indissociáveis, a saber, a clarificação de critérios, a identificação da lacuna entre desempenho atual e desejado e a mobilização do estudante para fechá-la. Trata-se, portanto, de um processo essencialmente dialógico e mediado, e não de um procedimento de mensuração. Essa constatação tem implicação direta sobre a discussão tecnológica que se segue: ferramentas digitais serão tanto mais úteis quanto mais sustentarem essas três operações, em vez de substituí-las. É à luz desse critério, e não da sofisticação técnica das plataformas, que a próxima seção examina a taxonomia disponível.

4. TAXONOMIA DE FERRAMENTAS TECNOLÓGICAS PARA AVALIAÇÃO FORMATIVA

O ecossistema de tecnologias que podem apoiar avaliação formativa é amplo e diversificado. Esta seção apresenta taxonomia organizacional dessas ferramentas, examinando funcionalidades principais, aplicações pedagógicas e evidências sobre efetividade quando disponíveis.

4.1 Sistemas de Resposta Estudantil





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

Sistemas de resposta estudantil, popularmente conhecidos como clickers ou sistemas de resposta pessoal, permitem que professores colem respostas de toda a turma a questões de múltipla escolha ou verdadeiro/falso em tempo real. Versões contemporâneas baseadas em web e aplicativos móveis - como Kahoot, Mentimeter, Socrative e Poll Everywhere - expandiram funcionalidades além de questões fechadas, incorporando nuvens de palavras, escalas de opinião e questões abertas breves (BENNETT, 2011).

Pedagogicamente, esses sistemas facilitam a implementação de técnicas como peer instruction, na qual estudantes primeiro respondem individualmente a questões conceituais, depois discutem com colegas e respondem novamente. A visualização agregada e anônima de respostas permite que professores identifiquem rapidamente conceitos mal compreendidos e ajustem instrução. Como observam Caldwell (2007), a anonimidade pode também encorajar participação de estudantes que hesitariam em responder publicamente.

Evidências empíricas sobre impacto de clickers são mistas. Meta-análise de Hunsu, Adesope e Bayly (2016) encontrou efeito positivo pequeno mas estatisticamente significativo sobre aprendizagem. Crucialmente, os autores identificaram que benefícios dependem de como sistemas são utilizados: mero uso de clickers para verificação de presença ou coleta de respostas sem discussão subsequente produz efeitos negligenciáveis; em contraste, uso integrado a estratégias de discussão entre pares e feedback formativo produz ganhos substanciais.

Cumprê destacar que, ao lado das plataformas comerciais de origem internacional, têm ganhado relevância no contexto educacional brasileiro iniciativas nacionais e soluções de código aberto que ampliam o escopo geográfico e tecnológico das ferramentas disponíveis. A Khan Academy Brasil, em parceria com a Fundação Lemann, oferece versão localizada de exercícios adaptativos com feedback imediato e suporte a professores da educação básica. O Plurall, desenvolvido pela Somos Educação, integra avaliações formativas a trilhas de aprendizagem alinhadas à Base Nacional Comum Curricular (BNCC). No campo do código aberto, o Moodle, amplamente adotado em universidades brasileiras, oferece módulos nativos de questionários, feedback automatizado e análise de aprendizagem (Learning Analytics),

Revista *OWL Journal*, Campina Grande - PB, v.4 n.5 (2026) - ISSN 2965-2634

A Revista *OWL Journal* está licenciada com uma Licença Creative Commons Atribuição (CC BY)





enquanto o H5P permite a criação de atividades interativas embarcáveis em ambientes virtuais de aprendizagem. Soluções como LimeSurvey e Open edX também viabilizam a implementação de avaliações formativas escaláveis sem dependência de licenças proprietárias, fator particularmente relevante para instituições públicas e contextos de recursos limitados.

O Quadro 2 apresenta uma síntese comparativa entre ferramentas comerciais internacionais e alternativas brasileiras ou de código aberto, evidenciando características relevantes para a tomada de decisão pedagógica e institucional.

Quadro 2 – Comparativo entre ferramentas internacionais, brasileiras e de código aberto para avaliação formativa

Critério	Comerciais internacionais (Kahoot, Mentimeter)	Plataformas brasileiras (Plurall, Khan Academy Brasil)	Código aberto (Moodle, H5P)
Custo	Versão gratuita limitada; planos pagos para recursos avançados	Modelos mistos: gratuito (Khan) e licenciamento institucional (Plurall)	Gratuito; custos restritos a hospedagem e suporte
Alinhamento curricular	Genérico; demanda adaptação docente	Alinhadas à BNCC e ao currículo nacional	Customizável conforme contexto institucional
Soberania de dados	Dados armazenados em servidores estrangeiros	Adesão à LGPD; possibilidade de hospedagem nacional	Controle institucional total sobre dados
Customização	Limitada às funcionalidades do fornecedor	Moderada; trilhas predefinidas	Ampla; código aberto permite extensões

Fonte: Elaborado pelo autor (2026).

4.2 Plataformas de Avaliação Diagnóstica e Adaptativa

Plataformas de avaliação adaptativa utilizam algoritmos para ajustar dificuldade de questões baseando-se em respostas prévias do estudante, permitindo estimação mais eficiente e precisa de níveis de conhecimento. Sistemas como Khan Academy, DreamBox, ALEKS (Assessment and Learning in Knowledge Spaces) e i-Ready incorporam assessment





adaptativo integrado a recomendações personalizadas de atividades de aprendizagem (SHUTE; RAHIMI, 2017).

A lógica subjacente a sistemas adaptativos é que, ao ajustar dinamicamente dificuldade de itens, pode-se obter informação diagnóstica mais rica com menos questões, mantendo estudantes em zona de desafio apropriado (nem excessivamente fácil nem frustrante). Adicionalmente, muitas plataformas fornecem dashboards para professores visualizarem progressão de estudantes, identificarem tópicos que demandam reforço, e acessarem relatórios detalhados de desempenho (BENNETT, 2011).

Evidências sobre efetividade de plataformas adaptativas são variadas. Cheung e Slavin (2013), em revisão de programas de matemática apoiados por tecnologia, encontraram que sistemas adaptativos bem implementados podem produzir ganhos modestos mas consistentes, com tamanhos de efeito tipicamente entre 0.10 e 0.20. No entanto, os autores enfatizam que tecnologia per se não garante resultados; benefícios dependem de integração a práticas pedagógicas estruturadas, capacitação docente e suporte técnico adequado.

4.3 Ferramentas de Feedback Automatizado

Sistemas de feedback automatizado analisam produções estudantis - tipicamente escritas ou código de programação - e fornecem feedback sem intervenção docente. Ferramentas como Turnitin, Grammarly e sistemas de correção automatizada de redações utilizam processamento de linguagem natural para identificar questões gramaticais, coerência argumentativa, ou similaridade com fontes externas. Em ensino de programação, sistemas como CodeLab e WebWork fornecem feedback imediato sobre correção de código (SHUTE; RAHIMI, 2017).

A vantagem principal é escalabilidade: feedback pode ser fornecido instantaneamente a múltiplas submissões. Hattie e Timperley (2007) observam que feedback imediato pode ser particularmente benéfico para tarefas procedurais, onde correção oportuna previne consolidação de erros. No entanto, limitações importantes devem ser reconhecidas: sistemas





automatizados frequentemente focam aspectos superficiais (gramática, formato) em detrimento de dimensões mais profundas (qualidade argumentativa, criatividade); e feedback genérico pode ser menos efetivo que comentários personalizados e específicos de professores humanos.

Nicol (2010) argumenta que feedback automatizado deve ser concebido como complemento, não substituto, de feedback docente, sendo mais apropriado para estágios iniciais de desenvolvimento de trabalhos, permitindo que professores concentrem esforços em feedback sobre dimensões mais complexas em estágios posteriores.

4.4 Aplicativos e Plataformas Gamificadas

Ferramentas gamificadas incorporam elementos de design de jogos - pontos, níveis, badges, placares - a contextos de aprendizagem e avaliação. Plataformas como Quizizz, Duolingo, Classcraft e Prodigy utilizam mecânicas de jogos para aumentar engajamento e motivação. Do ponto de vista de avaliação formativa, essas ferramentas frequentemente oferecem feedback imediato, múltiplas oportunidades de prática, e progressão adaptativa através de níveis de dificuldade crescente (WILLIAM; THOMPSON, 2007).

Evidências sobre gamificação em educação são mistas. Meta-análise de Sailer e Homner (2020) encontrou efeito positivo pequeno a moderado sobre aprendizagem, mas com variabilidade substancial entre estudos. Os autores identificaram que gamificação tende a ser mais efetiva quando elementos de jogo estão integrados significativamente a objetivos de aprendizagem, em contraste com aplicação superficial de pontos e badges a atividades convencionais.

Kingston e Nash (2011) alertam para riscos potenciais: ênfase excessiva em recompensas extrínsecas pode minar motivação intrínseca; e competição via placares pode ser contraproducente para estudantes de menor desempenho. Portanto, design cuidadoso que equilibre elementos de colaboração, progressão individual e feedback formativo é essencial.





Olhando-se transversalmente as quatro categorias examinadas, emerge um padrão interpretativo relevante: as evidências empíricas convergem ao demonstrar que efeitos atribuíveis às ferramentas, isoladamente consideradas, são tipicamente modestos e fortemente modulados pelo modo de uso pedagógico. Os tamanhos de efeito reportados (entre 0.10 e 0.20 para plataformas adaptativas; pequenos a moderados para gamificação; pequenos para sistemas de resposta estudantil) situam-se sistematicamente abaixo do efeito de 0.79 atribuído ao feedback de qualidade na síntese de Hattie (2009) (CHEUNG; SLAVIN, 2013; HUNSU; ADESOPE; BAYLY, 2016; SAILER; HOMNER, 2020). Esse contraste é teoricamente significativo: indica que a tecnologia, por si, não replica a potência da avaliação formativa bem implementada; ela amplia o alcance operacional de uma prática cuja efetividade depende de variáveis pedagógicas que precedem e excedem o instrumental empregado. Tal constatação reforça a tese aqui sustentada e prepara a discussão da seção seguinte, na qual se examina como a inteligência artificial generativa altera os termos desse debate.

5. INTELIGÊNCIA ARTIFICIAL EM SISTEMAS DE AVALIAÇÃO FORMATIVA

Avanços em inteligência artificial (IA) têm possibilitado desenvolvimento de sistemas de avaliação formativa cada vez mais sofisticados, capazes de análises complexas de padrões de aprendizagem, geração de feedback personalizado e recomendação adaptativa de recursos. Esta seção examina aplicações contemporâneas de IA, explorando potencialidades e limitações.

5.1 Sistemas de Tutoria Inteligente

Intelligent Tutoring Systems (ITS) utilizam IA para fornecer instrução e feedback personalizados, adaptando-se dinamicamente a necessidades individuais de estudantes. Sistemas como AutoTutor, Carnegie Learning e Cognitive Tutor têm sido desenvolvidos para domínios como matemática, ciências e alfabetização. Esses sistemas tipicamente incorporam





modelos do domínio de conhecimento, modelos do estudante (representações de conhecimento atual e equívocos), e modelos pedagógicos que determinam quando e como intervir (SHUTE; RAHIMI, 2017).

Meta-análises de efetividade de ITS reportam resultados promissores. VanLehn (2011) identificou que sistemas de tutoria inteligente bem desenvolvidos podem produzir ganhos de aprendizagem equivalentes a tutoria humana individual, embora ainda inferiores a tutoria humana especializada. Steenbergen-Hu e Cooper (2014) encontraram efeito positivo moderado em matemática, com tamanho de efeito médio de 0.34.

No entanto, Bennett (2011) observa que desenvolvimento de ITS efetivos demanda investimentos substanciais em engenharia de conhecimento e design pedagógico, limitando sua viabilidade para muitos contextos educacionais. Além disso, sistemas tendem a ser mais efetivos em domínios bem estruturados com regras claras (matemática, física) do que em áreas que demandam pensamento crítico, criatividade ou julgamento contextual.

5.2 Análise de Aprendizagem por Aprendizagem de Máquina

Técnicas de aprendizagem de máquina estão sendo aplicadas à análise de dados educacionais em larga escala para identificar padrões de aprendizagem, prever desempenhos futuros e personalizar intervenções. Algoritmos podem, por exemplo, analisar logs de interação de estudantes com plataformas digitais para identificar aqueles em risco de evasão ou que necessitam suporte adicional em tópicos específicos (CHEUNG; SLAVIN, 2013).

Sistemas de recomendação educacional, inspirados em plataformas comerciais como Netflix, utilizam algoritmos de filtragem colaborativa para sugerir recursos de aprendizagem baseados em padrões de estudantes similares. A promessa é personalização em escala: cada estudante recebe recomendações adaptadas a seu perfil de aprendizagem, interesses e necessidades diagnosticadas (HATTIE; TIMPERLEY, 2007).

Entretanto, Sadler (1989) alerta para riscos de opacidade algorítmica: quando sistemas de IA fazem recomendações baseadas em modelos complexos que nem professores nem





estudantes compreendem, pode-se minar agência docente e desenvolvimento de autorregulação estudantil. Além disso, sistemas de IA treinados em dados históricos podem perpetuar vieses existentes, favorecendo ou desfavorecendo sistematicamente determinados grupos de estudantes.

5.3 Processamento de Linguagem Natural para Feedback em Escrita

Avanços em processamento de linguagem natural (PLN) têm possibilitado sistemas mais sofisticados de análise e feedback sobre escrita estudantil. Ferramentas contemporâneas vão além de verificação gramatical, oferecendo análises de coerência argumentativa, adequação de vocabulário, estrutura retórica e até mesmo dimensões estilísticas. Sistemas como Criterion e WriteToLearn utilizam PLN para fornecer feedback formativo sobre múltiplos rascunhos (NICOL, 2010).

Evidências sobre efetividade são moderadamente positivas. Wilson e Czik (2016) encontraram que feedback automatizado sobre escrita pode melhorar qualidade de textos, especialmente quando combinado a oportunidades de revisão e quando sistemas fornecem não apenas identificação de problemas, mas sugestões específicas de melhoria. No entanto, limitações importantes persistem: sistemas ainda lutam com compreensão de contexto, nuances e criatividade; e há preocupações de que estudantes possam desenvolver dependência excessiva, não desenvolvendo capacidades próprias de autocrítica e revisão.

Mais recentemente, a emergência dos Grandes Modelos de Linguagem (Large Language Models, LLMs), especialmente a partir do lançamento público do ChatGPT em novembro de 2022, alterou substancialmente o panorama do feedback automatizado em escrita. Diferentemente dos sistemas de PLN de gerações anteriores, que tipicamente operavam com regras predefinidas e classificadores estatísticos voltados a dimensões específicas (gramática, coesão lexical, similaridade textual), os LLMs contemporâneos, como GPT-4, Claude e Gemini, são capazes de gerar feedback dialógico, contextualmente sensível e adaptado ao nível do estudante, abordando simultaneamente aspectos superficiais e profundos





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

da produção escrita (KASNECI et al., 2023). Estudos exploratórios indicam que esses modelos podem fornecer comentários sobre estrutura argumentativa, originalidade de ideias, coerência discursiva e adequação ao gênero textual com qualidade que se aproxima, em determinadas dimensões, do feedback docente experiente (ZHAI, 2023).

Entretanto, a integração de LLMs a práticas de avaliação formativa suscita questões pedagógicas e éticas inéditas. Pesquisas recentes apontam que esses modelos podem produzir “alucinações” (informações plausíveis, mas factualmente incorretas) e reproduzir vieses presentes em seus dados de treinamento (KASNECI et al., 2023). Além disso, a facilidade com que estudantes podem utilizar LLMs para gerar trabalhos completos desafia distinções tradicionais entre autoria humana e produção assistida por máquina, exigindo redefinição de critérios de avaliação e desenvolvimento de novas competências de literacia em IA (UNESCO, 2023). Como argumenta Zhai (2023), o desafio para educadores não é resistir à adoção dessas tecnologias, mas integrá-las de forma crítica, transformando-as em parceiras de aprendizagem que estimulem reflexão metacognitiva, e não atalhos que substituam o pensamento estudantil. Nesse sentido, os LLMs simultaneamente potencializam e problematizam o feedback formativo em escrita, demandando pesquisas empíricas robustas que ainda estão em estágio inicial de consolidação.

A análise das três frentes em que a inteligência artificial atualmente intervém na avaliação formativa (tutoria inteligente, análise de aprendizagem e processamento de linguagem natural) revela um deslocamento conceitual relevante. Se os sistemas das gerações anteriores ampliavam, sobretudo, a escala e a velocidade da coleta de evidências, os sistemas contemporâneos baseados em LLMs deslocam o eixo da promessa tecnológica para a geração de feedback de natureza dialógica, função historicamente atribuída ao docente humano. Esse deslocamento não é trivial: ele aproxima as ferramentas de uma das três operações pedagógicas indissociáveis identificadas anteriormente, a mediação entre desempenho atual e desejado, mas o faz sem incorporar a relação intersubjetiva, o conhecimento contextual do estudante e a responsabilidade ética do julgamento pedagógico, que constituem condições de possibilidade do feedback efetivo. Daí a inadequação tanto do entusiasmo acrítico quanto da

Revista *OWL Journal*, Campina Grande - PB, v.4 n.5 (2026) - ISSN 2965-2634

A Revista *OWL Journal* está licenciada com uma Licença Creative Commons Atribuição (CC BY)





rejeição reativa: o que se impõe é a formulação explícita de princípios pedagógicos que governem a incorporação dessas tecnologias, conforme se discute a seguir.

6. ARQUITETURA PEDAGÓGICA FORMATIVA TECNOLOGICAMENTE MEDIADA: UM FRAMEWORK ANALÍTICO

As seções precedentes percorreram os fundamentos teóricos da avaliação formativa, sistematizaram criticamente o ecossistema de ferramentas disponíveis e examinaram as transformações introduzidas pela inteligência artificial generativa. A travessia conduziu a uma constatação central: a tecnologia, por si, não replica a potência da avaliação formativa bem implementada; ela amplia o alcance operacional de uma prática cuja efetividade depende de variáveis pedagógicas que precedem e excedem o instrumental empregado. Essa constatação, no entanto, permanece abstrata sem um dispositivo conceitual que permita avaliar, em cada caso concreto, se determinada adoção tecnológica configura ou não uma prática genuinamente formativa. Cabe, portanto, formalizar tal dispositivo.

Esta seção propõe e formaliza o conceito de **Arquitetura Pedagógica Formativa Tecnicamente Mediada** (APFTM), framework analítico que articula cinco dimensões interdependentes por meio das quais é possível avaliar se determinada configuração tecnológica sustenta ou descaracteriza a operação formativa. A APFTM não constitui prescrição de boas práticas nem catálogo de funcionalidades desejáveis; é dispositivo conceitual que permite a pesquisadores, gestores e docentes examinarem criticamente tecnologias avaliativas existentes e orientarem decisões de adoção, desenvolvimento ou descontinuidade. As cinco dimensões propostas são: (i) diagnóstica; (ii) dialógico-temporal; (iii) de agência e autorregulação; (iv) ético-algorítmica; e (v) de reversibilidade pedagógica.

Cada dimensão é apresentada nas subseções seguintes mediante estrutura tripartite: definição conceitual, fundamentação teórica e indicadores empíricos operacionalizáveis. Esse último elemento atende à exigência de que categorias conceituais analíticas, em pesquisa educacional, sejam suficientemente operacionalizáveis para sustentar investigação empírica





subsequente, evitando que o framework opere apenas em registro retórico-interpretativo. A subseção 6.6 articula as cinco dimensões em uma proposta integrada e discute condições de implementação do framework em contextos institucionais concretos.

6.1 Dimensão Diagnóstica

A dimensão diagnóstica refere-se à capacidade do sistema tecnológico de gerar evidências de aprendizagem em granularidade pedagogicamente útil. Não se trata apenas de quantidade de dados produzidos, mas de sua qualidade interpretativa: evidências formativamente relevantes são aquelas que permitem identificar não apenas o que o estudante sabe ou desconhece, mas onde, no processo de aprendizagem, encontra-se a lacuna entre desempenho atual e desejado (SADLER, 1989). Sistemas que produzem grandes volumes de dados sobre desempenho, mas insensíveis a essa distinção, oferecem ilusão diagnóstica sem substância pedagógica.

Pesquisas recentes em learning analytics têm avançado nessa direção. Baker e Inventado (2014) demonstram que sistemas analíticos avançados são capazes de inferir, a partir de padrões comportamentais finos, não apenas o desempenho do estudante, mas também seu estado motivacional e cognitivo. Gašević, Dawson e Siemens (2015) argumentam, contudo, que a sofisticação técnica desses sistemas não se traduz automaticamente em utilidade pedagógica: dados só se convertem em diagnóstico formativo quando articulados a teorias da aprendizagem que permitam interpretá-los, e quando são apresentados a docentes em formato que sustente decisão pedagógica em tempo útil. Nesse sentido, a dimensão diagnóstica do framework não é técnica, mas hermenêutica: avalia-se um sistema pela qualidade pedagógica das inferências que sustenta, não pela quantidade de variáveis que processa.

Indicadores operacionais desta dimensão, observáveis em estudos empíricos sobre sistemas avaliativos concretos, incluem: (a) presença de tipologias de erro pedagogicamente significativas, distinguindo, por exemplo, erros conceituais de erros procedurais ou de





inatenção; (b) capacidade de localizar a lacuna de aprendizagem em níveis específicos da progressão curricular, e não apenas registrar acertos e erros agregados; (c) integração entre dados de desempenho e contextualização pedagógica (objetivos da aula, sequência didática em curso); e (d) acessibilidade dos dados ao docente em interface que sustente interpretação acionável em tempo formativamente útil.

6.2 Dimensão Dialógico-Temporal

A dimensão dialógico-temporal articula duas propriedades historicamente reconhecidas como condições de efetividade do feedback formativo: seu caráter relacional, dirigido ao estudante particular em sua trajetória de aprendizagem, e sua oportunidade temporal, entregue em momento que permita ação subsequente. Hattie e Timperley (2007) formalizaram essas propriedades nas três perguntas que estruturam o feedback efetivo (para onde vou, onde estou, como prossigo), e Wiliam e Thompson (2007) enfatizam que feedback que estudantes não têm oportunidade de utilizar torna-se essencialmente avaliação somativa disfarçada.

A emergência dos grandes modelos de linguagem desloca essa discussão para terreno conceitualmente novo. Bearman e Ajjawi (2023) argumentam que a IA generativa transforma o feedback de evento pontual em fluxo contínuo, conferindo aos estudantes acesso permanente a respostas personalizadas e potencialmente reconfigurando a economia temporal da prática avaliativa. Mollick (2024) propõe que essa transformação não substitui a relação dialógica humana, mas a recompõe sob nova configuração: o feedback passa a operar em camadas, com a IA assumindo respostas rotineiras e o docente reservando-se a intervenções de maior densidade pedagógica. A potencialidade dessa recomposição, entretanto, depende de configurações de uso que preservem a centralidade do docente como mediador qualificado, evitando que o estudante converta a interação com a máquina em substituto integral da relação pedagógica.





Indicadores operacionais incluem: (a) latência entre coleta de evidência e devolutiva ao estudante, mensurável em estudos empíricos; (b) grau de personalização contextual do feedback, distinguindo respostas genéricas de intervenções ajustadas à trajetória específica do estudante; (c) existência de oportunidades estruturadas para revisão e retomada do trabalho após recepção do feedback (ciclos de rascunho-feedback-revisão, múltiplas tentativas com remediação intermediária); e (d) preservação de espaço relacional docente-estudante mesmo em arquiteturas mediadas por IA, observável na frequência e qualidade de interações pedagógicas humanas não substituídas pelo sistema automatizado.

6.3 Dimensão de Agência e Autorregulação

A dimensão de agência e autorregulação refere-se ao grau em que o sistema avaliativo mobiliza o estudante como sujeito do próprio processo de aprendizagem, e não como objeto de mensuração externa. Wiliam e Thompson (2007), em sua quinta estratégia-chave, enfatizam a ativação dos estudantes como donos de sua própria aprendizagem; Nicol (2010) argumenta que ferramentas tecnológicas só sustentam autorregulação quando fornecem aos estudantes visualizações de progresso ao longo do tempo, facilitam estabelecimento de objetivos pessoais e promovem reflexão metacognitiva. A questão central é se a mediação tecnológica amplia ou estreita o espaço de protagonismo discente.

Luckin (2018) tem desenvolvido formulações relevantes sobre o papel da IA como apoio à metacognição, sustentando que sistemas inteligentes podem ser desenhados não para substituir a cognição estudantil, mas para torná-la visível ao próprio estudante, ampliando sua capacidade de monitorar e regular o próprio aprendizado. Essa formulação inverte a lógica frequentemente assumida nos discursos comerciais sobre IA educacional, segundo a qual a personalização algorítmica seria, por si, sinônimo de respeito à individualidade do estudante. Personalização sem agência, adverte a autora, é apenas adaptação técnica a perfis estatísticos, podendo, paradoxalmente, reduzir a autonomia ao invés de ampliá-la, quando o estudante torna-se receptor passivo de trajetórias algorítmicas pré-definidas.





Indicadores operacionais desta dimensão incluem: (a) existência de interfaces estudantis que tornem visíveis ao próprio aluno seus padrões de aprendizagem e suas estratégias mais e menos efetivas; (b) participação estudantil na definição de critérios de sucesso e na interpretação das evidências geradas pelo sistema; (c) oportunidades para autoavaliação e avaliação por pares mediadas pela plataforma; e (d) ausência de mecanismos de gamificação ou personalização que substituam decisões pedagógicas do estudante por trajetórias algorítmicas determinísticas.

6.4 Dimensão Ético-Algorítmica

A dimensão ético-algorítmica trata da governança dos dados produzidos pelo sistema avaliativo, da transparência dos critérios algorítmicos que sustentam suas inferências e da responsabilidade pelo julgamento pedagógico que delas decorre. Distintamente de dimensões anteriores, cujo objeto é a operação formativa em si, esta dimensão examina as condições institucionais, jurídicas e éticas em que a operação ocorre. Sua centralidade no framework decorre da constatação de que arquiteturas avaliativas tecnologicamente sofisticadas, mas eticamente opacas, podem produzir efeitos pedagogicamente regressivos mesmo quando bem-sucedidas em outras dimensões.

Holmes e Tuomi (2022) sistematizam um conjunto de princípios éticos aplicáveis a sistemas de IA em educação, articulando exigências de transparência algorítmica, proteção de dados, mitigação de vieses e preservação da agência humana sobre decisões pedagógicas. Selwyn (2022) e Williamson (2017), em registro mais crítico, alertam para os riscos da datificação educacional, fenômeno pelo qual a aprendizagem é progressivamente reduzida àquilo que sistemas algorítmicos conseguem mensurar, processar e relatar. Para Williamson, a governança algorítmica em educação não é problema marginal, mas reconfiguração estrutural das relações de poder no campo educacional, exigindo escrutínio público sobre quem define os critérios incorporados aos sistemas, quem se beneficia da datificação e quem é





responsabilizado quando julgamentos algorítmicos produzem efeitos adversos sobre estudantes.

No contexto brasileiro, essa dimensão articula-se diretamente à Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018), cuja aplicação ao tratamento de dados de estudantes ainda demanda consolidação jurisprudencial e protocolos institucionais robustos. Indicadores operacionais desta dimensão incluem: (a) transparência sobre quais dados são coletados, como são processados e por quem podem ser acessados; (b) explicabilidade dos critérios algorítmicos que sustentam diagnósticos e recomendações automáticas; (c) preservação da autoridade pedagógica docente sobre julgamentos finais, evitando automação plena de decisões com consequência educacional relevante; e (d) mecanismos de contestação acessíveis a estudantes e responsáveis sobre inferências produzidas pelo sistema.

6.5 Dimensão de Reversibilidade Pedagógica

A dimensão de reversibilidade pedagógica, a mais original do framework proposto, refere-se à possibilidade epistemológica de revisar diagnósticos, classificações e trajetórias atribuídas pelo sistema avaliativo à luz de nova evidência ou nova compreensão sobre o estudante. Esta dimensão responde a uma característica historicamente reconhecida da avaliação formativa, frequentemente diluída em arquiteturas algorítmicas: a provisoriedade de seus juízos. Avaliação formativa pressupõe, por definição, que o que se afirma sobre o estudante hoje é hipótese de trabalho, não veredito; sua função é orientar a ação pedagógica subsequente, e não fixar trajetórias.

Sistemas algorítmicos, contudo, tendem ao oposto: classificações produzidas em determinado momento são incorporadas ao histórico do estudante e passam a condicionar processamentos subsequentes, gerando o que Perrotta (2021) caracteriza como path dependence algorítmica. Estudantes classificados precocemente como tendo desempenho fraco em determinada competência tendem a receber, dali em diante, materiais e trajetórias recomendadas pelo sistema que reforçam essa classificação, mesmo quando ela tenha se





baseado em evidência insuficiente ou contextualmente enviesada. Hoffmann (2017), em registro distinto, sustenta posição convergente ao argumentar contra o caráter classificatório da avaliação tradicional, defendendo a centralidade da provisoriedade dos juízos avaliativos como condição da mediação pedagógica.

A dimensão de reversibilidade pedagógica, portanto, examina se o sistema avaliativo preserva ou compromete essa provisoriedade. Indicadores operacionais incluem: (a) possibilidade institucional e técnica de revisão de diagnósticos com base em nova evidência, mediante intervenção docente; (b) ausência de mecanismos algorítmicos que cristalizem classificações em rótulos persistentes não revisáveis; (c) transparência sobre o histórico de classificações atribuídas ao estudante, permitindo identificar momentos de inflexão e revisão; e (d) existência de protocolos institucionais para descontinuidade de uso do sistema sem prejuízo pedagógico, condição que protege a instituição contra dependências tecnológicas que comprometam sua autonomia pedagógica.

6.6 Articulação Sistêmica e Condições de Implementação

As cinco dimensões apresentadas não operam isoladamente. A APFTM postula que arquiteturas tecnologicamente mediadas só configuram prática formativa genuína quando sustentam, conjuntamente, as cinco dimensões em grau pedagogicamente suficiente. Sistemas que avançam fortemente em uma ou duas dimensões, mas falham significativamente em outras, produzem configurações híbridas que mimetizam a avaliação formativa em sua superfície, sem reproduzi-la em sua estrutura. Por exemplo, uma plataforma adaptativa pode apresentar excelente desempenho diagnóstico (dimensão 6.1) e oferecer feedback temporalmente oportuno (dimensão 6.2), mas, se opera por classificações algorítmicas não revisáveis (falha em 6.5) ou não preserva a autoridade pedagógica docente sobre julgamentos finais (falha em 6.4), descaracteriza-se como dispositivo formativo.

A implementação concreta da APFTM em contextos institucionais demanda três condições articuladas. A primeira é a subordinação da decisão tecnológica à decisão





pedagógica: como adverte Bennett (2011), a adoção de ferramentas avaliativas deve decorrer de prévia clarificação de objetivos formativos, e não da disponibilidade comercial ou da pressão institucional pela adoção de IA. A segunda é o desenvolvimento de competências docentes específicas para a operação do framework, incluindo capacidade de interpretar dados em registros analíticos, identificar lacunas nas cinco dimensões em sistemas concretos e configurar arranjos pedagógicos que compensem fragilidades dimensionais detectadas. Kingston e Nash (2011) identificam que implementações bem-sucedidas de avaliação formativa tecnologicamente mediada incluem programas robustos de desenvolvimento profissional articulando aspectos técnicos e pedagógicos. A terceira é a existência de governança institucional capaz de sustentar a dimensão ético-algorítmica em arranjos contratuais, técnicos e jurídicos concretos, particularmente em contextos de adoção de plataformas comerciais cuja operação interna nem sempre é transparente à instituição usuária.

Cabe aqui explicitar com precisão o status epistemológico do framework proposto. A APFTM é, neste artigo, construto teórico de natureza analítica, derivado por inferência conceitual da literatura sobre avaliação formativa, feedback efetivo e crítica à mediação algorítmica da educação. Os indicadores operacionais associados a cada dimensão, apresentados ao longo das subseções 6.1 a 6.5, constituem proposições operacionalizáveis para pesquisa empírica futura, e não instrumentos psicométricos validados. Sua função, no presente trabalho, é dupla: por um lado, demonstrar a operacionalizabilidade do construto, atendendo à exigência de que categorias analíticas em pesquisa educacional sejam suficientemente concretas para sustentar investigação empírica; por outro, sinalizar caminhos pelos quais investigações subsequentes podem testar, refinar e eventualmente refutar elementos do framework. A validação dos indicadores como instrumentos de pesquisa demanda etapas que excedem os limites desta revisão, incluindo, por exemplo, aplicação a estudos de caso, análise de confiabilidade entre observadores e eventual desenvolvimento de protocolos de mensuração. O reconhecimento explícito desse status preserva o framework do risco oposto: por um lado, ser desautorizado por sua falta de validação psicométrica; por





outro, ser apropriado precocemente como instrumento pronto, descontextualizado de seu fundamento teórico.

Uma segunda explicitação se impõe, de natureza epistemológica mais profunda. Pode parecer haver tensão entre a tese central do artigo, segundo a qual tecnologias não replicam a potência da avaliação formativa, e a proposição de um framework para avaliar a adequação formativa de configurações tecnológicas. Essa tensão é, na verdade, aparente, e sua dissolução requer distinção precisa entre dois tipos de juízo avaliativo. O framework APFTM não opera como atestado positivo de que determinada configuração tecnológica *é* formativa; opera como dispositivo que identifica condições de possibilidade para que ela não *descaracterize* a operação formativa quando incorporada a uma arquitetura pedagógica mais ampla. A distinção é decisiva: as cinco dimensões funcionam como filtros negativos, e não como certificados positivos. Uma configuração tecnológica que falhe gravemente em uma ou mais dimensões descaracteriza-se como dispositivo formativo, mas o fato de uma configuração não falhar nas cinco dimensões não a converte, por si, em prática formativa – converte-a apenas em ferramenta cuja adoção é compatível com prática formativa conduzida por mediação humana qualificada. A efetividade formativa permanece, em última instância, função da arquitetura pedagógica humana em que a tecnologia se inscreve, não propriedade emergente da própria tecnologia. Essa precisão articula coerentemente a dimensão crítica e a dimensão construtiva do artigo: o framework é construído precisamente porque a tecnologia, sozinha, não basta – sua utilidade está em identificar quais configurações não comprometem a prática formativa, deixando à mediação humana qualificada a tarefa de efetivá-la.

A APFTM, em síntese, oferece-se à pesquisa empírica subsequente como instrumento de análise comparativa de sistemas, de avaliação institucional de adoções e de orientação ao desenvolvimento de tecnologias. Sua validação plena, contudo, depende de aplicação empírica em estudos de caso concretos, etapa de pesquisa que excede os limites deste artigo de natureza analítico-crítica e que se sugere como agenda investigativa imediata.





7. DESAFIOS E CONSIDERAÇÕES CRÍTICAS

Apesar de potencialidades significativas, a integração de tecnologias a práticas de avaliação formativa enfrenta desafios importantes que devem ser reconhecidos e endereçados. Esta seção examina algumas questões críticas que persistem no campo.

7.1 Equidade e Acesso Digital

A dependência de tecnologias para avaliação formativa pode exacerbar desigualdades educacionais quando estudantes têm acesso diferenciado a dispositivos, conectividade e suporte técnico. Bennett (2011) observa que enquanto ferramentas baseadas em escola podem mitigar algumas dessas disparidades, muitas plataformas contemporâneas assumem acesso doméstico a tecnologias, potencialmente desfavorecendo estudantes de contextos socioeconômicos menos privilegiados.

No Brasil, esse desafio é particularmente pronunciado. Dados da pesquisa TIC Educação 2024 indicam que, embora 96% das escolas brasileiras possuam acesso à internet, com avanços expressivos em redes municipais e rurais, persistem desigualdades significativas no uso pedagógico desses recursos: enquanto 67% dos alunos de escolas estaduais utilizam a internet para atividades solicitadas pelos professores, na rede municipal essa proporção é de apenas 27% (CETIC.BR, 2025). Tais disparidades, somadas à heterogeneidade de infraestrutura entre regiões e à desigualdade no acesso doméstico a dispositivos, impõem cautela na adoção de avaliações tecnologicamente mediadas como prática universal, especialmente em contextos de escolas públicas com alunado em situação de vulnerabilidade socioeconômica.

Além de acesso material, questões de literacia digital são relevantes. Estudantes variam em familiaridade e conforto com tecnologias digitais; aqueles com menor experiência podem enfrentar curvas de aprendizagem que obscurecem avaliação de conhecimento de conteúdo. Shute e Rahimi (2017) recomendam que sistemas avaliativos tecnológicos sejam





desenhados com interfaces intuitivas, forneçam suporte técnico acessível, e quando possível, ofereçam modalidades alternativas de participação.

7.2 Privacidade e Uso Ético de Dados

Plataformas de avaliação formativa coletam frequentemente dados granulares e contínuos sobre atividades de aprendizagem de estudantes. Questões importantes emergem sobre propriedade desses dados, usos permissíveis, proteção de privacidade e consentimento informado. Wiliam e Thompson (2007) enfatizam necessidade de políticas claras que especifiquem que dados são coletados, como são armazenados e protegidos, quem tem acesso, e que direitos estudantes possuem sobre seus próprios dados.

Além disso, há preocupações sobre usos secundários de dados educacionais para fins comerciais ou de vigilância. Hattie e Timperley (2007) argumentam que dados de avaliação formativa devem servir primariamente a propósitos de apoio à aprendizagem, não de accountability punitiva ou marketing comercial. Transparência sobre práticas de dados e envolvimento de comunidades educacionais em decisões sobre governança de dados são essenciais.

7.3 Risco de Estreitamento Curricular

Ferramentas tecnológicas de avaliação tendem a ser mais facilmente desenvolvidas para domínios bem estruturados com respostas claramente corretas ou incorretas. Há risco de que ênfase em avaliação tecnologicamente mediada possa resultar em estreitamento curricular, privilegiando aspectos facilmente mensuráveis em detrimento de objetivos educacionais mais complexos como pensamento crítico, criatividade, colaboração e desenvolvimento socioemocional (BLACK; WILIAM, 1998).

Sadler (1989) alerta que educadores devem resistir à tentação de permitir que capacidades de ferramentas disponíveis determinem que aprendizagens são valorizadas.





Avaliação formativa abrangente deve incorporar múltiplos métodos, incluindo observação, conversação, análise de trabalhos complexos e avaliação por portfólios, não se limitando a dimensões capturáveis por sistemas tecnológicos.

7.4 Sustentabilidade e Custos

Muitas plataformas tecnológicas envolvem custos de licenciamento, exigem infraestrutura de TI robusta e demandam suporte técnico contínuo. Kingston e Nash (2011) observam que entusiasmo inicial com ferramentas inovadoras pode não ser sustentável quando custos recorrentes se tornam evidentes ou quando mudanças em modelos de negócios de fornecedores alteram termos de acesso. Instituições devem considerar sustentabilidade de longo prazo ao selecionar ferramentas, preferindo quando viável soluções de código aberto ou com modelos de licenciamento estáveis e transparentes.

8. CONSIDERAÇÕES FINAIS

A travessia analítica empreendida ao longo deste artigo permite, neste fechamento, retornar à tensão estrutural enunciada na introdução, entre a racionalidade instrumental que organiza a oferta de tecnologias avaliativas e a racionalidade pedagógica dialógica que define a efetividade da avaliação formativa, agora com elementos para reorganizá-la em chave propositiva. O percurso evidenciou que essa tensão não é meramente conceitual: ela se manifesta empiricamente na discrepância sistemática entre os tamanhos de efeito reportados para tecnologias avaliativas isoladamente consideradas (geralmente pequenos a moderados) e os tamanhos de efeito atribuídos ao feedback de qualidade quando entendido como prática pedagógica integral (substancialmente maiores). Essa discrepância não desautoriza as tecnologias; redefine sua função: elas operam como amplificadores condicionais, cuja potência depende da arquitetura pedagógica em que se inscrevem.





Sustenta-se, com base nessa travessia, a tese de que a contribuição das tecnologias digitais para a avaliação formativa é predominantemente operacional, ampliando frequência, escala, imediaticidade e granularidade da coleta e devolução de evidências, e não constitutiva: as tecnologias não substituem, e em geral nem sequer aproximam-se de substituir, as três operações pedagógicas que conferem efetividade à avaliação formativa, a saber, a clarificação compartilhada de critérios, a identificação contextualizada da lacuna entre desempenho atual e desejado e a mobilização do estudante para fechá-la. Essa formulação tem consequência prática direta: a pergunta pedagogicamente pertinente não é “que ferramentas adotar?”, mas “que arquitetura pedagógica torna a adoção de uma ferramenta efetivamente formativa?”. A formalização do framework da Arquitetura Pedagógica Formativa Tecnicamente Mediada (APFTM), apresentada na seção 6, constitui a tradução conceitual e operacional dessa inversão de prioridade: ao identificar cinco dimensões pelas quais determinada configuração tecnológica pode sustentar ou descaracterizar a operação formativa, o framework oferece dispositivo analítico passível de aplicação empírica a sistemas concretos, instrumentalizando o que, sem ele, permaneceria como crítica genérica à adoção acrítica de tecnologias avaliativas.

A emergência dos grandes modelos de linguagem complica, sem inverter, essa formulação. Pela primeira vez, ferramentas tecnológicas reivindicam capacidade de gerar feedback dialógico, dimensão historicamente tida como insubstituível. Essa reivindicação merece ser examinada empiricamente, e não rejeitada por princípio; mas o exame, realizado sem ingenuidade, revela que mesmo o feedback gerado por LLMs permanece desprovido das condições intersubjetivas que sustentam o feedback docente efetivo, a saber, o conhecimento longitudinal do estudante, a responsabilidade ética pelo julgamento e a ancoragem em uma relação pedagógica concreta. A IA generativa, portanto, não dissolve a tensão central: redesenha-a em termos mais agudos, exigindo da pesquisa educacional refinamento conceitual sobre o que se entende por “feedback” e quais de suas dimensões são, e quais não são, delegáveis a sistemas computacionais.





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

Os desafios discutidos - equidade de acesso digital, privacidade de dados, riscos de estreitamento curricular e questões de sustentabilidade - não devem ser minimizados. Requerem atenção cuidadosa de educadores, gestores, desenvolvedores de tecnologia e formuladores de políticas. Como argumenta Bennett (2011), transformação produtiva de avaliação formativa mediante tecnologias depende não apenas de inovação técnica, mas de design pedagógico rigoroso, implementação contextualmente sensível e governança ética.

Olhando para o futuro, avanços em inteligência artificial, processamento de linguagem natural e análise de aprendizagem prometem capacidades ainda mais sofisticadas de diagnóstico personalizado e feedback adaptativo. No entanto, como observam Sadler (1989) e Nicol (2010), tecnologias mais avançadas não eliminam a necessidade de julgamento pedagógico humano, expertise de conteúdo e relações autênticas entre professores e estudantes. O desafio é desenvolver ecologias híbridas que aproveitem potencialidades de sistemas tecnológicos enquanto preservam e amplificam dimensões insubstituíveis de interação humana em educação.

Finalmente, cabe enfatizar que a finalidade última de avaliação formativa - com ou sem mediação tecnológica - é o apoio ao aprendizado de todos os estudantes. Práticas avaliativas devem ser orientadas primariamente por esse compromisso, utilizando tecnologias como meios para fins educacionais, nunca como fins em si mesmos. Quando implementadas com essa clareza de propósito, tecnologias podem efetivamente contribuir para a construção de ambientes educacionais mais responsivos, personalizados e promotores de aprendizagem profunda e significativa.

Pesquisas futuras poderão aprofundar a compreensão sobre as condições específicas sob as quais diferentes tecnologias são mais ou menos efetivas, explorar formas de integrar sistemas de IA a práticas formativas preservando a agência humana e investigar estratégias para mitigar riscos de inequidade e estreitamento curricular associados à avaliação tecnologicamente mediada. Como agenda investigativa imediata, sugere-se a aplicação empírica do framework APFTM a estudos de caso de sistemas avaliativos concretos, particularmente em ambientes brasileiros de educação básica e superior, viabilizando sua

Revista *OWL Journal*, Campina Grande - PB, v.4 n.5 (2026) - ISSN 2965-2634

A Revista *OWL Journal* está licenciada com uma Licença Creative Commons Atribuição (CC BY)





validação, refinamento e eventual expansão. Estudos comparativos entre plataformas comerciais internacionais, plataformas brasileiras e soluções de código aberto, organizados em torno das cinco dimensões do framework, poderiam produzir evidências sistemáticas sobre quais configurações tecnológicas efetivamente sustentam a operação formativa em contextos institucionais reais. No contexto brasileiro, uma agenda particularmente promissora compreende ainda estudos empíricos sobre a articulação entre avaliação formativa tecnologicamente mediada e a BNCC, análises do impacto da LGPD na governança de dados educacionais e investigações sobre soluções de baixo custo e código aberto adequadas a redes públicas de ensino, contribuindo para um desenvolvimento de tecnologias avaliativas mais sensível às especificidades e aos desafios do sistema educacional nacional.

REFERÊNCIAS

BAKER, R. S.; INVENTADO, P. S. Educational data mining and learning analytics. In: LARUSSON, J. A.; WHITE, B. (Ed.). **Learning analytics: from research to practice**. New York: Springer, 2014. p. 61-75.

BEARMAN, M.; AJJAWI, R. Learning to work with the black box: pedagogy for a world with artificial intelligence. **British Journal of Educational Technology**, v. 54, n. 5, p. 1160-1173, 2023.

BENNETT, R. E. Formative assessment: a critical review. **Assessment in Education: Principles, Policy & Practice**, v. 18, n. 1, p. 5-25, 2011.

BLACK, P.; WILIAM, D. Assessment and classroom learning. **Assessment in Education: Principles, Policy & Practice**, v. 5, n. 1, p. 7-74, 1998.

BLOOM, B. S. Learning for mastery. **Evaluation Comment**, v. 1, n. 2, p. 1-12, 1968.

BRASIL. Ministério da Educação. **Base Nacional Comum Curricular**. Brasília: MEC, 2018.

CALDWELL, J. E. Clickers in the large classroom: current research and best-practice tips. **CBE Life Sciences Education**, v. 6, n. 1, p. 9-20, 2007.





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

CETIC.BR. **Pesquisa sobre o uso das tecnologias de informação e comunicação nas escolas brasileiras: TIC Educação 2024**. São Paulo: Comitê Gestor da Internet no Brasil, 2025.

CHEUNG, A. C. K.; SLAVIN, R. E. The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: a meta-analysis. **Educational Research Review**, v. 9, p. 88-113, 2013.

GAŠEVIĆ, D.; DAWSON, S.; SIEMENS, G. Let's not forget: learning analytics are about learning. **TechTrends**, v. 59, n. 1, p. 64-71, 2015.

HATTIE, J. **Visible learning**: a synthesis of over 800 meta-analyses relating to achievement. London: Routledge, 2009.

HATTIE, J.; TIMPERLEY, H. The power of feedback. **Review of Educational Research**, v. 77, n. 1, p. 81-112, 2007.

HOFFMANN, J. **Avaliação: mito e desafio: uma perspectiva construtivista**. 45. ed. Porto Alegre: Mediação, 2017.

HOLMES, W.; TUOMI, I. State of the art and practice in AI in education. **European Journal of Education**, v. 57, n. 4, p. 542-570, 2022.

HUNSU, N. J.; ADESOPE, O.; BAYLY, D. J. A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. **Computers & Education**, v. 94, p. 102-119, 2016.

KASNECI, E.; SEßLER, K.; KÜCHEMANN, S.; BANNERT, M.; DEMENTIEVA, D.; FISCHER, F.; GASSER, U.; GROH, G.; GÜNNEMANN, S.; HÜLLERMEIER, E.; KRUSCHE, S.; KUTYNIOK, G.; MICHAELI, T.; NERDEL, C.; PFEFFER, J.; POQUET, O.; SAILER, M.; SCHMIDT, A.; SEIDEL, T.; STADLER, M.; WELLER, J.; KUHN, J.; KASNECI, G. ChatGPT for good? On opportunities and challenges of large language models for education. **Learning and Individual Differences**, v. 103, art. 102274, 2023.

KINGSTON, N.; NASH, B. Formative assessment: a meta-analysis and a call for research. **Educational Measurement: Issues and Practice**, v. 30, n. 4, p. 28-37, 2011.

Revista *OWL Journal*, Campina Grande - PB, v.4 n.5 (2026) - ISSN 2965-2634

A Revista *OWL Journal* está licenciada com uma Licença Creative Commons Atribuição (CC BY)





REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634

LIMA, T. C. S.; MIOTO, R. C. T. Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica. **Revista Katálisis**, Florianópolis, v. 10, n. esp., p. 37-45, 2007.

LUCKESI, C. C. **Avaliação da aprendizagem escolar: estudos e proposições**. 22. ed. São Paulo: Cortez, 2011.

LUCKIN, R. **Machine learning and human intelligence: the future of education for the 21st century**. London: UCL Institute of Education Press, 2018.

MINAYO, M. C. S. **O desafio do conhecimento: pesquisa qualitativa em saúde**. 14. ed. São Paulo: Hucitec, 2014.

MOLLICK, E. **Co-intelligence: living and working with AI**. New York: Portfolio/Penguin, 2024.

NICOL, D. J. From monologue to dialogue: improving written feedback processes in mass higher education. **Assessment & Evaluation in Higher Education**, v. 35, n. 5, p. 501-517, 2010.

PERROTTA, C. Programming the platform university: learning analytics and predictive infrastructures in higher education. **Research in Education**, v. 109, n. 1, p. 53-71, 2021.

SADLER, D. R. Formative assessment and the design of instructional systems. **Instructional Science**, v. 18, n. 2, p. 119-144, 1989.

SAILER, M.; HOMNER, L. The gamification of learning: a meta-analysis. **Educational Psychology Review**, v. 32, n. 1, p. 77-112, 2020.

SCRIVEN, M. The methodology of evaluation. In: TYLER, R. W.; GAGNÉ, R. M.; SCRIVEN, M. **Perspectives of curriculum evaluation**. Chicago: Rand McNally, 1967. p. 39-83.

SELWYN, N. **Education and technology: key issues and debates**. 3. ed. London: Bloomsbury, 2022.

SHUTE, V. J.; RAHIMI, S. Review of computer-based assessment for learning in elementary and secondary education. **Journal of Computer Assisted Learning**, v. 33, n. 1, p. 1-19, 2017.

Revista *OWL Journal*, Campina Grande - PB, v.4 n.5 (2026) - ISSN 2965-2634

A Revista *OWL Journal* está licenciada com uma Licença Creative Commons Atribuição (CC BY)





STEENBERGEN-HU, S.; COOPER, H. A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. **Journal of Educational Psychology**, v. 106, n. 2, p. 331-347, 2014.

UNESCO. **Guidance for generative AI in education and research**. Paris: UNESCO Publishing, 2023.

VANLEHN, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. **Educational Psychologist**, v. 46, n. 4, p. 197-221, 2011.

WHITTEMORE, R.; KNAFL, K. The integrative review: updated methodology. **Journal of Advanced Nursing**, v. 52, n. 5, p. 546-553, 2005.

WILIAM, D.; THOMPSON, M. Integrating assessment with instruction: what will it take to make it work? In: DWYER, C. A. (Ed.). **The future of assessment: shaping teaching and learning**. Mahwah: Lawrence Erlbaum Associates, 2007. p. 53-82.

WILLIAMSON, B. **Big data in education: the digital future of learning, policy and practice**. London: SAGE, 2017.

WILSON, J.; CZIK, A. Automated essay evaluation software in English Language Arts classrooms: effects on teacher feedback, student motivation, and writing quality. **Computers & Education**, v. 100, p. 94-109, 2016.

ZHAI, X. ChatGPT for next generation science learning. **XRDS: Crossroads, The ACM Magazine for Students**, v. 29, n. 3, p. 42-46, 2023.

Recebido em: 12/04/2026

Aprovado em: 30/04/2026

Publicado em: 24/05/2026

