

USO DE IA GENERATIVA

no aperfeiçoamento da avaliação
educacional no Brasil



Dados internacionais de Catalogação na Publicação (CIP)
Fundação Itaú

Uso da IA Generativa no Aperfeiçoamento da Avaliação Educacional no Brasil
[recurso eletrônico]. — São Paulo : Fundação Itaú, 2025.
70 p. : il. color. ; PDF.

Inclui bibliografia.

ISBN: 978-85-7979-211-3

1. Inteligência Artificial - Educação. 2. Avaliação Educacional. 3. Inteligência Artificial
Generativa - Educação. 4. Educação básica - Brasil. 5. Políticas Públicas de Educação
I. Fundação Itaú. II. Observatório. III. Título.

CDD 371.26

Bibliotecário Fernando Galante Silva CRB-8/10536

Expediente

FUNDAÇÃO ITAÚ

PRESIDENTE

Eduardo Saron

GERENTE OBSERVATÓRIO

Carla Christine Chiamareli

OBSERVATÓRIO

Alan Pessoa Valadares

Ligia Dona de Souza

GERENTE DE COMUNICAÇÃO INSTITUCIONAL E ESTRATÉGICA

Ana de Fátima Sousa

COMUNICAÇÃO INSTITUCIONAL E ESTRATÉGICA

Alan Albuquerque

Renato Corch

Fabiana Sousa

Ailson Taveira

CRÉDITOS DA PUBLICAÇÃO

COORDENAÇÃO

Maria Helena Guimarães de Castro | Cátedra Instituto Ayrton Senna/IEA/USP Polo Ribeirão Preto

Priscilla de Albuquerque Tavares | FGV/SP

EQUIPE TÉCNICA

Barbara Alves Medeiros

Mariah Morikawa

APOIO ADMINISTRATIVO

Silvia Véspoli Godoy

REVISÃO

Alex Criado

DIAGRAMAÇÃO

Boitatá Design

CAPA

Mimosas Design

Imagem gerada por Inteligência Artificial: Sora, OpenAI



Sumário Executivo

As transformações trazidas pela Inteligência Artificial (IA) já alcançam as salas de aula, os sistemas de gestão e, cada vez mais, as avaliações educacionais. Diante desse cenário, a pesquisa “Uso de IA Generativa no Aperfeiçoamento da Avaliação Educacional no Brasil” teve como objetivo compreender como as novas tecnologias podem apoiar o aprimoramento das avaliações na educação básica — tanto as avaliações formativas, realizadas no cotidiano escolar, quanto as somativas. Ou ainda as avaliações em larga escala, como o Saeb e os exames internacionais, a exemplo do Pisa e do PIRLS.

O estudo envolveu:

- A revisão de 66 documentos nacionais e internacionais, entre artigos, relatórios e estudos técnicos sobre IA e avaliação educacional;
- 13 entrevistas com especialistas e gestores ligados a universidades, órgãos públicos, centros de avaliação e organizações internacionais;
- A realização de quatro eventos de debate e disseminação, com participação de pesquisadores, representantes de redes de ensino e de instituições brasileiras e estrangeiras.

A pesquisa mostra que a IA está transformando a avaliação educacional, ao abrir a possibilidade de unir o que a educação sempre tratou separadamente: aprender e avaliar. As novas tecnologias permitem que a avaliação se torne um processo contínuo, adaptativo e integrado à aprendizagem, oferecendo diagnósticos instantâneos e feedbacks personalizados a estudantes e professores.

Já é possível observar grandes avanços nas experiências de países como Austrália, França, Estados Unidos (Naep) e OCDE (Pisa, PIRLS, TIMSS). Tais experiências mostram que a IA vem sendo usada para:

Gerar e corrigir automaticamente itens e redações, com alto grau de precisão e menor custo;

Analisar dados de processo, como tempo de resposta e estratégias de resolução, revelando como o estudante pensa e aprende;

Avaliar competências complexas, como raciocínio matemático, criatividade e habilidades socioemocionais;

Oferecer feedback imediato e personalizado para alunos e professores;

Reduzir vieses humanos e aumentar a confiabilidade e comparabilidade entre países e contextos linguísticos.

Essas experiências têm em comum três pilares:

- Desenvolvimento de provas digitais, aplicadas em plataformas seguras e adaptativas;
- Clareza dos objetivos pedagógicos, com a tecnologia a serviço da aprendizagem e não da simples automatização;
- Formação contínua de professores e investimento em infraestrutura digital.

Na educação básica, a IA pode integrar as avaliações formativas e somativas, fornecendo informações que ajudam docentes e gestores a tomar decisões pedagógicas em tempo real. Tecnologias já utilizadas no Brasil — como o reconhecimento automático de fala para avaliar fluência em leitura, testes adaptativos e sistemas de correção automatizada de escrita — demonstram ganhos de precisão, agilidade e personalização.

Além disso, a IA permite uma nova geração de avaliações capazes de observar o processo de aprendizagem e não apenas o resultado final, promovendo o desenvolvimento de competências cognitivas e socioemocionais mais amplas.

Apesar do potencial, a pesquisa mostra que o Brasil e a América Latina ainda estão em fase inicial de incorporação da IA nas avaliações educacionais. As principais barreiras são:

- Infraestrutura digital desigual, especialmente nas redes públicas;
- Falta de formação e letramento em IA entre professores e gestores;
- Risco de ampliação das desigualdades, caso o acesso às tecnologias permaneça restrito;
- Desafios éticos e regulatórios, como privacidade dos dados, vieses algorítmicos e transparência das decisões automatizadas.

Há também resistências culturais à adoção de processos automatizados de correção e interpretação de resultados, além da necessidade de validação psicométrica rigorosa dos instrumentos gerados por IA.

Com base nas evidências analisadas e nas boas práticas internacionais, o estudo apresenta cinco caminhos estratégicos para o Brasil:

- 1 Modernizar o Saeb e as avaliações estaduais, incorporando plataformas digitais adaptativas e análises de processo que unam diagnóstico, acompanhamento e políticas públicas.
- 2 Criar uma estratégia nacional do uso da IA na avaliação educacional, com princípios éticos, protocolos de governança e transparência de algoritmos.
- 3 Formar professores e gestores para o uso pedagógico da IA, com foco na interpretação dos dados e na construção de feedbacks formativos.
- 4 Garantir infraestrutura tecnológica equitativa, assegurando conectividade e equipamentos nas escolas públicas.
- 5 Fomentar centros de pesquisa e inovação, capazes de desenvolver soluções nacionais contextualizadas e sustentáveis.

A pesquisa evidencia que a IA é uma aliada estratégica para tornar a avaliação mais justa, inteligente e voltada à aprendizagem. Quando usada de forma responsável, ela amplia a capacidade do sistema educacional de reconhecer as diferenças, acompanhar trajetórias e apoiar o desenvolvimento integral dos estudantes.

O desafio brasileiro não é apenas adotar novas tecnologias, mas redefinir o papel da avaliação: de um instrumento de classificação para um instrumento de aprendizagem e equidade.

Com visão pedagógica, infraestrutura adequada e formação docente, a Inteligência Artificial pode transformar as avaliações em poderosas ferramentas de inclusão e de melhoria da qualidade da educação, colocando o Brasil em sintonia com as inovações educacionais mais avançadas do mundo.



Índice

Sumário Executivo	3
Lista de Siglas	8
1. Introdução	12
2. Sistematização e Revisão da Literatura	15
Estatísticas e classificação dos documentos	16
Análise Transversal da Literatura	18
Avaliação com Tecnologias/IA na Educação Básica	19
Avaliação com Tecnologias/IA no Ensino Superior	22
Tecnologias/IA em Avaliação em Larga Escala - Educação Básica	25
Outros Enfoques Analíticos sobre IA na Avaliação	27
Contexto da América Latina e Caribe (LAC)	27
Frameworks conceituais para integração da IA	28
Pesquisa exploratória sobre LLMs e avaliação	28
Políticas e diretrizes para a educação básica	29
Guias globais e princípios éticos	29
3. Entrevistas com Atores-chave	34
Metodologia de condução e sistematização das entrevistas	34
Achados e Aprendizados das Entrevistas	35
4. Eventos de debate e disseminação	37
Evento 01 – Workshop “Como utilizar de modo adequado e ético a IA generativa para o desenvolvimento das avaliações educacionais”	37
Evento 02 – Workshop “Inteligência Artificial e Avaliação Educacional”	43
Evento 03 – Workshop “Inteligência Artificial e Inovações nas Avaliações do ETS: Mudanças no NAEP”	49
Evento 04 - Mesa-redonda ABAVE: “Inteligência Artificial e Avaliação Educacional”	57
5. Considerações finais	61
6. Referências bibliográficas	66

Lista de Siglas

Abave	Associação Brasileira de Avaliação Educacional
AERA	American Educational Research Association (Associação Americana de Pesquisa em Educação)
AES	Automated Essay Scoring (Pontuação Automática de Ensaios)
AIA	Aplicações de Inteligência Artificial
AIG	Automatic Item Generation (Geração Automatizada de Itens)
AIAS	Artificial Intelligence Assessment Scale (Escala de Avaliação da Inteligência Artificial)
ALC	Latin America and the Caribbean (América Latina e Caribe)
Anebhi	Associação Nacional de Educação Básica Híbrida
ANN	Artificial Neural Networks (Redes Neurais Artificiais)
APA	American Psychological Association (Associação Americana de Psicologia)
APP-7	Assessment Prompt Protocol – Version 7 (Conjunto experimental de questões utiliza das para avaliar o desempenho da IA)
ASL	Assessment as Learning (Avaliação como Aprendizagem)
ASR	Automatic Speech Recognition (Reconhecimento Automático de Fala)
AITC	Adjusted Item-Total Correlation (Correlação Ajustada Item-Total)
AWE	Automated Writing Evaluation (Avaliação Automatizada da Escrita)
BoW	Bag-of-Words (Saco de Palavras)
BPP-4	Benchmark Prompt Protocol – Version 4 (Conjunto experimental de benchmark usado para validação do modelo em questões de escolha única)
CAEd	Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (MG)
CAT	Computerized Adaptive Testing (Teste Adaptativo Computadorizado)
CNN	Convolutional Neural Networks (Redes Neurais Convolucionais)
COPPA	Children’s Online Privacy Protection Act (Lei de Proteção à Privacidade Online de Crianças)
CR	Constructed Response (Resposta Construída)
CPU	Central Processing Unit (Unidade Central de Processamento)
DINA	Deterministic Inputs, Noisy “And” Gate Model (Modelo Diagnóstico Psicométrico)
EA	Exact Agreement (Acordo Exato)
EAES	Automated Writing Evaluation (Avaliação Automatizada da Escrita)
EFL	English as a Foreign Language (Ensino e Aprendizagem de Inglês como Língua Estrangeira)
EL	English Learners (Aprendizes de Inglês)
Enem	Exame Nacional do Ensino Médio

ESEA	Elementary and Secondary Education Act (Lei de Educação Elementar e Secundária – EUA)
ET	Extended Time (Tempo Estendido)
ETS	Educational Testing Service
EWS	Early Warning System (Sistema de Alerta Precoce)
FERPA	Family Educational Rights and Privacy Act (Lei Federal de Direitos Educacionais e Privacidade – EUA)
FNN	Fully Connected Feedforward Neural Networks (Redes Neurais Feed-forward totalmente conectadas)
G/A	Grading/Assessment (Avaliação/Correção)
GenAI / IAG	Generative Artificial Intelligence (Inteligência Artificial Generativa)
GPL	General Public License (Licença Pública Geral)
GPT-2	Generative Pre-trained Transformer 2
GPT-3	Generative Pre-trained Transformer 3
GPT-3.5	Generative Pre-trained Transformer 3.5
GPT-4	Generative Pre-trained Transformer 4
GPTx	Generative Pre-trained Transformer Family (Família de Modelos Generative Pre-trained Transformer – versões futuras)
GRE	Graduate Record Examination (Exame de Admissão para Pós-Graduação)
GPUs	Graphics Processing Units (Unidades de Processamento Gráfico)
HAI	Human-Centred AI (IA Centrada no Ser Humano)
HE	Higher Education (Ensino Superior)
HEXACO	Six-Factor Personality Model (Modelo de Personalidade de Seis Fatores: Honestidade -Humildade, Emocionalidade, Extroversão, Amabilidade, Conscienciosidade, Abertura)
IA	Inteligência Artificial
Iave	Instituto de Avaliação Educativa (Portugal)
IIC	Inter-Item Correlation (Correlação Interitens)
ILSA	International Large-Scale Assessments (Avaliações Internacionais em Larga Escala)
IEA	International Association for the Evaluation of Educational Achievement
IEP	Individualized Education Plans (Planos de Educação Individualizados)
IR	Item Reliability (Fidedignidade do Item)
IRC	Item Response Curve (Curva de Resposta ao Item)
IRT	Item Response Theory (Teoria de Resposta ao Item)
ITS	Intelligent Tutoring Systems (Sistemas de Tutoria Inteligente)
K-12	Kindergarten to 12th Grade (Educação Básica)
LA	Learning Analytics (Análise de Aprendizagem)
LCA	Latent Class Analysis (Análise de Classes Latentes)

LLM	Large Language Models (Grandes Modelos de Linguagem)
LOR	Log Odds Ratio (Razão de Log Odds)
LPA	Latent Profile Analysis (Análise de Perfis Latentes)
LRL	Low Resource Languages (Idiomas de Baixo Recurso)
LSA	Latent Semantic Analysis (Análise Semântica Latente)
LSAT	Law School Admission Test
LSI	Latent Semantic Indexing (Indexação Semântica Latente)
LSTM	Long Short-Term Memory (Arquitetura de Rede Neural Recorrente)
MB	Model-Based (Baseado em Modelo)
MDE	Mineração de Dados Educacionais
MFRM	Many-Facet Rasch Model (Modelo de Rasch de Múltiplas Facetas)
ML	Machine Learning (Aprendizado de Máquina)
MMT	Multiple Machine Translations (Múltiplas Traduções Automáticas)
MMT_ANNs	Multiple Machine Translations with Artificial Neural Networks (Múltiplas Traduções Automáticas combinadas com Redes Neurais Artificiais)
MOOC	Massive Open Online Courses (Cursos Online Abertos Massivos)
MT	Machine Translation (Tradução Automática)
NAEP	National Assessment of Educational Progress (Avaliação Nacional do Progresso Educacional – EUA)
NCME	National Council on Measurement in Education
NEES	Núcleo de Excelência em Tecnologias Sociais
NeuralCD	Neural Cognitive Diagnosis (Diagnóstico Cognitivo Neural)
NIST	National Institute of Standards and Technology
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
ODS 4	Objetivo de Desenvolvimento Sustentável 4 (Educação de Qualidade)
OEI	Organização dos Estados Ibero-Americanos
PE	Personalized Education (Educação Personalizada)
PIRLS	Progress in International Reading Literacy Study
Pisa	Programme for International Student Assessment
PLN	Natural Language Processing (Processamento de Linguagem Natural)
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Itens Preferenciais de Relatório para Revisões Sistemáticas e Meta-análises)
PSI	Problem Solving and Inquiry (Resolução de Problemas e Investigação)
QA	Question Answering (Sistema de Pergunta e Resposta)
RNA	Redes Neurais Artificiais

RR	Research Report (Relatório de Pesquisa)
RR-25-03	Research Report 25-03 (Relatório de Pesquisa ETS 2025)
RSL	Revisão Sistemática da Literatura
Saeb	Sistema de Avaliação da Educação Básica
SAT	Scholastic Assessment Test
SEA+	Adaptive Platform for Assessing Mathematics, Reading and Natural Sciences (Plataforma adaptativa para avaliar Matemática, Leitura e Ciências Naturais – Uruguai)
Seduc-PI	Secretaria de Educação do Estado do Piauí
Seei	Sistema Experto de Avaliação Inteligente
SEL	Socioemotional Learning (Aprendizagem Socioemocional)
SMD	Standardized Mean Difference (Diferença Média Padronizada)
SRL	Self-Regulated Learning (Autorregulação da Aprendizagem)
SR	Self-Regulation (Autorregulação)
SSR	Self-Supervised Representation (Representação Auto-supervisionada)
STI	Sistemas de Tutoria Inteligente
SWD	Students With Disabilities (Estudantes com Deficiência)
SW	Semantic Web (Web Semântica)
SYA	Study Year Abroad (Ano de Estudo no Exterior)
TA	Machine Translation (Tradução Automática)
TAO	Test Authoring and Output (Plataforma TAO de avaliação digital)
TEQSA	Tertiary Education Quality and Standards Agency (Agência de Qualidade e Padrões do Ensino Superior – Austrália)
TIMSS	Trends in International Mathematics and Science Study
TRI	Teoria de Resposta ao Item
Ufal	Universidade Federal de Alagoas
UFC	Universidade Federal do Ceará
UFJF	Universidade Federal de Juiz de Fora
UFS	Universidade Federal de Sergipe
UL	Unsupervised Learning (Aprendizado Não Supervisionado)
UDL	Universal Design for Learning (Desenho Universal para a Aprendizagem)
USP-RP	Universidade de São Paulo – Ribeirão Preto
VR	Virtual Reality (Realidade Virtual)
WE	Writing Evaluation (Avaliação da Escrita)

1. Introdução

Uma das dimensões importantes para garantir o direito de todo estudante a aprender refere-se aos resultados das avaliações em larga escala. Elas monitoram o processo de aprendizagem e subsidiam diagnósticos para a tomada de decisões. Ao tornarem visível se o direito à aprendizagem de qualidade para todos foi ou não alcançado, essas avaliações acabam produzindo evidências sólidas para a formulação ou reformulação de políticas públicas.

O Brasil desenvolveu um sistema robusto de avaliação em larga escala, em especial o Saeb, que serve como bússola de orientação das avaliações locais e regionais do país, mas que não passou por mudanças conceituais, metodológicas e de formatos nos últimos 20 anos. Apesar da sua qualidade e abrangência, as avaliações nacionais e subnacionais da educação básica brasileira precisam avançar e incorporar gradativamente as inovações observadas em diferentes sistemas. Entre essas inovações, sobressai a introdução da Inteligência Artificial (IA) no desenvolvimento de plataformas interativas e novos modelos de avaliação das competências do nosso século. As escolas e o processo de ensino e aprendizagem são impactados cada vez mais pelas novas tecnologias, tendência acelerada com a Inteligência Artificial generativa.

Em que medida a IA pode apoiar a melhoria das aprendizagens e da equidade dos sistemas de ensino? Como a avaliação em larga escala pode se beneficiar da IA reduzindo custos, aumentando a eficiência e a confiabilidade para medir o desempenho dos estudantes? Como a IA pode apoiar o planejamento pedagógico e o trabalho dos professores por meio de avaliações formativas na condução do processo de ensino e aprendizagem? Quais as pesquisas e soluções para o aperfeiçoamento de um dos instrumentos essenciais da avaliação, ou seja, os itens que compõem os testes cada vez mais interativos? Como a IA Generativa contribui para desenvolver a avaliação das competências mais complexas, cada vez mais requeridas no mundo atual? De que forma as ferramentas de IA vêm sendo utilizadas como soluções para construir instrumentos de avaliação (processual, formativa, somativa)? Como vêm sendo utilizadas para desenvolver plataformas para elaborar relatórios e devolutivas dos resultados educacionais e para aperfeiçoar os instrumentos de gestão no contexto da escola?

A avaliação educacional no Brasil enfrenta desafios significativos, que incluem a necessidade de personalização, escalabilidade e melhoria contínua dos instrumentos avaliativos. A recente evolução da Inteligência Artificial, em especial das técnicas de IA generativa, apresenta oportunidades promissoras para aperfeiçoar esses processos. Modelos de linguagem avançados podem, por exemplo, auxiliar na elaboração de itens de prova, gerar feedback personalizado para estudantes e automatizar a correção de avaliações discursivas, contribuindo para maior eficiência e precisão. Diante desse cenário, justifica-se o presente projeto que investiga o uso de IA generativa como estratégia de inovação para aprimorar a qualidade e a efetividade da avaliação educacional no país.

Esta pesquisa tem dois objetivos gerais:

- Identificar iniciativas nacionais e internacionais de uso da Inteligência Artificial generativa para o desenvolvimento de sistemas de avaliação da educação básica.
- Sistematizar as iniciativas em andamento e, a partir delas, trazer recomendações para aperfeiçoar as avaliações nacionais em larga escala e oferecer subsídios para o uso da IA em avaliações processual, formativa e somativa de forma geral.

Estes desdobram-se nos seguintes objetivos específicos:

- Mapear e analisar iniciativas existentes, no Brasil e no exterior, que utilizam IA (especialmente IA generativa) em diferentes tipos de avaliação educacional (formativa, somativa e processual).
- Identificar de que forma a IA generativa pode apoiar a elaboração de instrumentos avaliativos, a análise de resultados e a correção automatizada.
- Avaliar os benefícios e desafios associados ao uso da IA na avaliação, com ênfase em questões éticas, validação psicométrica, segurança de dados e aceitação por parte de educadores e gestores.

- Propor recomendação e mapear boas práticas para a integração adequada e ética de ferramentas de IA generativa em políticas e programas de avaliação educacional no Brasil.
-

A coleta de informações para a realização desta pesquisa incluiu a sistematização e revisão de 66 documentos bibliográficos da literatura acadêmica e literatura cinza sobre o emprego da Inteligência Artificial na avaliação educacional, cujos resultados estão descritos na Seção 2. Realizaram-se também 13 entrevistas e quatro eventos de discussão/debates com atores-chave (pesquisadores, gestores e empreendedores), envolvidos com o desenvolvimento ou a aplicação de ferramentas de Inteligência Artificial para o avanço da avaliação nas escolas, redes ou em larga escala. As informações obtidas destas interações estão descritas nas seções 3 e 4, respectivamente.

2. Sistematização e Revisão da Literatura

Para sistematizar os achados da literatura sobre o tema da pesquisa, foi conduzido um extenso levantamento bibliográfico sobre aplicações de IA na avaliação educacional, com foco em IA generativa. A busca envolveu fontes acadêmicas e institucionais, abrangendo diferentes perspectivas e também recomendações de leitura dos entrevistados.

Em nível nacional, foram pesquisados portais e repositórios de universidades brasileiras que são referências em educação e tecnologia, visando identificar estudos e publicações recentes relacionados ao tema. Em paralelo, foram consultados documentos técnicos e relatórios de organismos internacionais que atuam com avaliações externas, incluindo publicações da OCDE, estudos do ETS, da IEA e do Boston College acerca de inovações no TIMMS e PIRLS, além de materiais do Instituto de Avaliação Educativa de Portugal (Iave), entre outros.

A seleção dos documentos considerou a relevância para as perguntas de pesquisa do projeto, priorizando trabalhos que tratam do uso de IA em contextos de avaliação, seja na concepção de instrumentos, processamento de resultados ou políticas de avaliação educacional. Cada documento foi catalogado e classificado segundo:

- A natureza da publicação, nas seguintes categorias: i. apresentação (material divulgado em evento ou congresso); ii. artigo acadêmico, iii. livro; iv. position paper; v. relatório institucional, guia ou texto para discussão.
- O tema principal do documento abordar o uso de IA em avaliação em larga escala ou o emprego de IA em avaliações de forma geral.

Estatísticas e classificação dos documentos

Foram identificados um conjunto de 66 documentos que compõem o corpus de análise bibliográfica deste projeto. Desse total, 42 documentos (64%) são pesquisas e artigos publicados em periódicos acadêmicos; 16 (24%) são relatórios de disseminação produzidos por organismos internacionais ou instituições brasileiras; seis (9%) são apresentações realizadas em congressos; além de um livro (2%) e um position paper (2%) (Gráfico 01).

GRÁFICO 01 NATUREZA DO DOCUMENTO



Fonte: Elaboração própria

GRÁFICO 02 REFERÊNCIA INCLUÍDA NA SISTEMATIZAÇÃO



Fonte: Elaboração própria

Dos 66 documentos levantados, 12 (18%) têm como tema principal o uso da IA em avaliações em larga escala. Outras 34 referências (52%) abordam o emprego de ferramentas de IA em avaliações de forma mais abrangente. A classificação de cada documento definiu a sua inclusão para a leitura e sistematização em etapa posterior da pesquisa. Foram selecionados os documentos que tratam do emprego de ferramentas de IA em soluções de avaliação, sejam elas em larga escala, formativas ou em contexto mais abrangente, totalizando 46 referências bibliográficas, 70% do total de trabalhos levantados (Gráfico 02).

Foi realizada uma primeira análise desses documentos, segundo o tipo de avaliação educacional abordado (avaliação formativa, somativa ou processual/contínua) e segundo o foco do uso da IA reportado (por exemplo, se tratava de plataformas adaptativas de aprendizagem, bancos de itens gerados por IA, sistemas de correção automatizada, análises preditivas de desempenho, etc.). Essa primeira categorização permite identificar tendências e lacunas.

- **Por tipo de avaliação abordada:** Verificou-se uma distribuição equilibrada: aproximadamente 40% dos documentos tratam predominantemente de avaliação formativa, explorando como a IA pode apoiar o acompanhamento do aprendizado em sala de aula e a personalização do ensino. Cerca de 30% enfocam avaliações somativas ou de larga escala, discutindo especialmente a geração de itens de teste e correção automatizada em exames padronizados. Os 30% restantes abrangem avaliação processual ou diagnóstica, frequentemente ligada a sistemas inteligentes de tutoria e monitoramento contínuo do aluno ao longo do ano letivo. Alguns documentos abordam mais de um tipo de avaliação, mas foram categorizados pelo enfoque principal.
- **Por foco do uso da IA:** Em torno de um terço dos trabalhos levantados investiga plataformas adaptativas ou tutorias inteligentes, nas quais a IA é parte integral do ambiente de aprendizagem, que também realiza avaliações. Outro grande conjunto, aproximadamente 25%, dedica-se a questões de bancos de itens e geração automatizada de questões, incluindo estudos sobre a qualidade de itens gerados por IA e algoritmos para a seleção de questões conforme a proficiência do aluno. Cerca de 20% dos documentos concentram-se em correção automatizada de respostas abertas (especialmente redações e questões discursivas), cobrindo desde métodos de processamento de linguagem natural até análises de validade desses sistemas. Os demais 20-25% englobam temas diversos, como análise preditiva de desempenho educacional com machine learning, detecção de fraude em exames usando IA, e reflexões ético-regulatórias sobre IA na educação.

Esse panorama bibliográfico confirma que o interesse pelo tema é recente e em rápida expansão, refletindo a evolução tecnológica dos últimos anos. Muitas publicações são de 2020 em diante, indicando um campo em desenvolvimento. O anexo ‘Sistematização da literatura’, que compõe este relatório, apresenta a lista de documentos levantados, bem como a classificação segundo sua natureza, tema abordado (avaliação em larga escala, avaliação em geral), breve descrição do estudo e indicação sobre a inclusão em sistematização e análise posterior. Já as Referências bibliográficas incluem as indicações de todos os documentos levantados.

A seguir, é apresentada a análise transversal das referências bibliográficas incluídas no escopo da pesquisa. A sistematização e a análise individual de cada estudo encontram-se no Anexo I.

Análise Transversal da Literatura

A crescente integração da Inteligência Artificial (IA) no campo da educação representa um fenômeno transformador, mas complexo, que redefine práticas pedagógicas e de gestão. Esta análise bibliográfica tem como propósito sistematizar e discutir criticamente a produção acadêmica recente sobre a aplicação de tecnologias e IA nos processos de avaliação educacional, abrangendo desde a educação básica até o ensino superior, e também nas avaliações em larga escala. O objetivo é mapear tendências, inovações e desafios emergentes.

Para desvendar essa complexidade, a análise adota uma estrutura categorizada, partindo do eixo central da “Avaliação com tecnologias/IA”. A partir deste eixo, a análise se aprofunda em manifestações específicas do fenômeno em diferentes níveis de ensino — educação básica, ensino superior e avaliação em larga escala — e em enfoques transversais que oferecem perspectivas complementares, como contextos regionais, frameworks conceituais, pesquisas exploratórias, diretrizes políticas e princípios éticos.

A escolha da avaliação como eixo central justifica-se por seu papel fundamental nos sistemas educacionais. A IA surge com a promessa de revolucionar este campo, oferecendo um potencial disruptivo que se manifesta em oportunidades de personalização e eficiência e em desafios incontornáveis

de natureza ética e pedagógica. Por essa razão, a avaliação serve como um prisma ideal para analisar o impacto da IA na educação, revelando tanto seu potencial para otimizar processos quanto os riscos associados à sua implementação.

Esta revisão privilegia tanto a identificação de tendências e contribuições recorrentes na literatura (como a automação do feedback e a análise de dados em larga escala) quanto a análise dos achados específicos e das metodologias empregadas em cada estudo referenciado. A análise inicia-se, a seguir, com a exploração das aplicações da IA no contexto da educação básica.

Avaliação com Tecnologias/IA na Educação Básica

A educação básica constitui um campo fértil e estratégico para a aplicação de Inteligência Artificial na avaliação, pois abrange desde o desenvolvimento de competências fundamentais, como a alfabetização, até a preparação para exames sistêmicos que definem trajetórias educacionais. As tecnologias de IA prometem personalizar o acompanhamento do aluno, otimizar o tempo do professor e gerar diagnósticos mais precisos. Esta seção analisará como a literatura recente documenta o uso de plataformas, metodologias e pesquisas concretas para transformar a avaliação nesse nível de ensino.

A literatura aponta para uma diversidade de aplicações da IA na educação básica, que vão desde ferramentas de apoio à avaliação formativa em sala de aula até sistemas nacionais de avaliação em larga escala. No contexto da sala de aula, estudos destacam o potencial da IA para fornecer feedback e promover a autorregulação do aluno (Hopfenbeck et al., 2023). Um exemplo concreto é uma ferramenta de relatórios visuais na China que, habilitada por IA, analisa dados de exames mensais para gerar devolutivas personalizadas, demonstrando um impacto positivo no desempenho e na autoeficácia dos estudantes (Liao et al., 2024). Essa diversidade se manifesta em diferentes modelos de implementação: a América Latina, por exemplo, exhibe um cenário caracterizado por uma abordagem de mercado, com plataformas do setor privado como as brasileiras Letrus, Pontue e Geekie, e a uruguaia SEA+, que utilizam IA para avaliações adaptativas e feedback detalhado em escrita e outras áreas (Rivas, 2025). Em contraste direto, a França adota uma estra-

tégia centralizada e sistêmica, com um sistema de e-assessment em larga escala liderado pelo Estado, que abrange milhões de estudantes no início de cada ano letivo (Bret, 2024). Este contraste evidencia um trade-off entre a agilidade e a inovação pontual de um modelo orientado ao mercado e a coerência sistêmica e potencial de equidade de uma abordagem estatal. A pesquisa também explora o uso de robôs educacionais para analisar o processo de aprendizagem e de análise preditiva, a fim de antecipar o desempenho dos alunos (Martínez-Comesaña et al., 2023).

As tecnologias de IA têm sido aplicadas para desenvolver competências específicas e para monitorar a trajetória dos estudantes. No Brasil, o CAEd/UFJF desenvolve tecnologias de reconhecimento automático de fala (ASR) para automatizar a avaliação da fluência em leitura, uma habilidade crucial nos anos iniciais (Silva; Scortegagna, 2023). Plataformas como a Letrus focam no aprimoramento da escrita, fornecendo correções detalhadas (Rivas, 2025). Em outra frente, modelos de mineração de dados educacionais são utilizados para prever o risco de evasão no ensino médio com altas taxas de precisão (Alkan, 2024). Além disso, a literatura aponta para o potencial de tecnologias imersivas, como a realidade virtual (VR), como uma ferramenta aplicável para enriquecer a experiência de aprendizagem e, implicitamente, de avaliação nos ensinos fundamental e médio (Ahmad et al., 2022).

Nível/País	Tecnologia IA	Tipo de Avaliação	Principais Achados	Desafios	Autor(es)/Ano
1º ano/EUA	Deep Learning (RNNs)	Formativa, diagnóstica	Previsão de necessidades de testes com ~90% de precisão; redução de viés docente.	Validação restrita (Matemática); dificuldade de interpretação pelos professores.	Shin et al. (2022)
Básica/Brasil	ASR, Gamificação, CAT	Diagnóstica, formativa, somativa	Alta concordância com corretores humanos (94–98%); relatórios diagnósticos.	Pouca pesquisa em ASR infantil; necessidade de infraestrutura digital.	CAEd/UFJF (2023)
4º ano PIRLS / Global	GPT-3 (AIG)	Somativa, diagnóstica	Itens gerados semelhantes aos humanos; redução de custos.	Exige verificação factual e revisão humana.	Bezirhan & von Davier (2023)
Básica/AL	Correção de redações, adaptatividade, alerta precoce	Diagnóstica, formativa, somativa	Feedback detalhado e alerta de risco de evasão.	Risco de plágio, divisão digital, desumanização.	Rivas (2025)
4º ano TIMSS / Global	ANNs, tradução automática, PLN	Somativa	Correção automatizada comparável à humana (Kappa 0,86)	Dependência da qualidade do treino; vieses linguísticos	Jung et al. (2024, 2025)
9º ano/China	NeuralCD, relatórios visuais	Formativa	Melhora desempenho e auto-eficácia; reduz carga docente	Ansiedade de teste; estudo restrito a Biologia	Liao et al. (2023)
6º–8º ano/ Alemanha	LLMs (GPT-3.5, GPT-4)	Formativa	Identificação de erros lógicos com alta acurácia	Baixa performance em erros específicos; não lê dados multi-modais	Bewersdorff et al. (2023)
8ª série NAEP / EUA	HAI, ML, dados de processo	Processual, Diagnóstica	Perfis holísticos de aprendizagem; feedback explicativo	Anotação trabalhosa; dilemas éticos	Guo et al. (2024)
EF e EM/Global (revisão)	Redes neurais, chatbots, robôs, preditiva	Preditiva, Somativa, Formativa	Previsão precisa de desempenho; avaliações automatizadas	Necessidade de grandes bases de dados; limitações pedagógicas	Martinez-Comesaña et al. (2023)
Básica/AL	Tutores (ALEKS), feedback (Teach-FX), geração de conteúdo	Formativa, Diagnóstica, Processual	ALEKS melhorou notas em Matemática; feedback contínuo a professores	Escalabilidade limitada; divisão digital	Molina et al. (2024) / World Bank
K-12 e Superior/Global	GenAI (AIAS)	Formativa, Somativa, Processual	Estrutura ética para uso da GenAI; apoio à escrita e linguagem	Ferramentas de detecção ineficazes; equidade de acesso	Perkins et al. (2024, 2025)

Fonte: Elaboração própria

Enquanto a educação básica explora essas ferramentas para a personalização e o diagnóstico precoce, o ensino superior apresenta um panorama distinto, com outras prioridades e desafios, que serão abordados a seguir.

Avaliação com Tecnologias/IA no Ensino Superior

No ensino superior, a aplicação da Inteligência Artificial na avaliação é explorada com um duplo foco: por um lado, busca-se otimizar processos em larga escala, como a correção de tarefas em Massive Open Online Courses (MOOCs); por outro, visa-se engajar os estudantes em tarefas complexas e autênticas que espelhem as exigências do mercado de trabalho. Esta seção examina como a literatura aborda o uso de IA e tecnologias digitais em universidades e na formação continuada, com uma atenção crescente ao impacto disruptivo da IA generativa (GenAI).

O advento da IA generativa representou uma inflexão paradigmática, deslocando o foco da pesquisa de sistemas preditivos e adaptativos — dominantes até 2018 (Zawacki-Richter et al., 2019) — para a compreensão do uso de ferramentas generativas e colaborativas pelos próprios estudantes. Se antes a pesquisa se concentrava em modelos para predição de desempenho e evasão ou avaliação automatizada, estudos recentes investigam os padrões de uso de ferramentas como o ChatGPT em avaliações de escrita e formativas, revelando que os alunos demonstram discernimento ao escolher diferentes tipos de IA para propósitos específicos, como planejamento e verificação de respostas (Foung et al., 2024; Huang et al., 2024).

Exemplos concretos de plataformas e metodologias ilustram essa tendência. Um sistema baseado em IA para cursos de ciência de dados, por exemplo, combina análise de código e Processamento de Linguagem Natural (PLN) para oferecer correção automatizada e feedback detalhado em avaliações formativas e somativas (Vittorini et al., 2021). Da mesma forma, modelos de mineração de dados educacionais são aplicados para prever notas finais e o risco de evasão em cursos universitários, alcançando altos níveis de precisão (Alkan, 2024).

Para guiar a integração dessas novas tecnologias, a literatura também propõe frameworks teóricos. Um exemplo notável é o arcabouço que busca in-

tegrar a GenAI ao design de avaliações alinhadas à Taxonomia de Bloom. E, assim, fornecer um guia prático para que educadores possam criar tarefas que promovam habilidades de ordem superior, em vez de simplesmente tentar criar avaliações “à prova de IA” (Page et al., 2024).

Nível/País	Tecnologia IA	Tipo de Avaliação	Principais Achados	Desafios	Autor(es)/Ano
Superior/Global	ITS, SR, SL	Formativa, adaptativa, somativa	IA inevitável e benéfica: automatiza tarefas, personaliza aprendizado, amplia acessibilidade.	Alto custo, falta de experiência física, docentes desinformados.	Ahmad et al. (2021)
Superior/Global	ITS, VR, LA	Formativa, somativa	IA reduz carga docente, oferece tutoria personalizada e adaptativa.	Setor educacional pouco orientado por dados.	Ahmad et al. (2022)
Superior/Global	Mineração de Dados Educacionais (MDE)	Diagnóstica, formativa, somativa	Correção automatizada rápida; modelos preditivos alcançaram 57–80% de acurácia.	Éticos: privacidade, desumanização; baixa confiança docente.	Alkan (2024)
Superior/EUA	GenAI (Chat-GPT)	Formativa	Uso crescente para explicações e checagem de respostas.	Dependência excessiva; risco de fraude acadêmica.	Huang et al. (2024)
Superior/Hong Kong	GenAI + Grammarly	Somativa (45%), formativa	GenAI apoiou planejamento e brainstorming; Grammarly nivelou escrita.	Riscos: informações falsas, viés, equidade free vs. premium.	Foung et al. (2024)
Superior/Global	GenAI	Formativa, somativa	Framework integra GenAI às dimensões de Bloom; feedback imediato.	GenAI falha em habilidades superiores; inconsistência de detecção.	Page et al. (2024)
Superior/Global	GenAI (LLMs)	Formativa, somativa (alto risco)	LLMs tiveram bom desempenho em GRE, SAT e LSAT.	Problemas de validade/reliabilidade; fraude em exames.	Hao et al. (2024)
Superior/MOOCs	PLN, ML, SVM	Formativa, somativa	Correção automatizada de tarefas complexas; bom alinhamento com notas humanas.	Feedback ainda limitado; risco em casos limítrofes.	Vittorini et al. (2021)
Superior/Global (Revisão Sist.)	Adaptativo, AES	Formativa (predom.), diagnóstica, somativa	Melhora de até 17% em desempenho escrito.	Falta de modelos pedagógicos; opacidade dos algoritmos.	Gonzalez-Catalayud et al. (2021)

Superior/Global (Revisão Sist.)	AES, ITS	Formativa, somativa, diagnóstica	“Renascimento” da avaliação com feedback just-in-time.	Predomínio da computação, pouca base pedagógica; riscos éticos pouco tratados.	Zawacki-Richter et al. (2019)
Superior/Global	Framework Conceitual	Formativa, somativa	Abordagem centrada no aluno, ética e integrada.	Abordagem instrumental prevalente; risco de impactos da GenAI.	Madland et al. (2024)
Superior/Global	GenAI (AIAS Framework)	Formativa, somativa	AIAS adota GenAI de forma ética e transparente.	Ineficácia de detectores, equidade de acesso, necessidade de redesign avaliativo.	Perkins et al. (2024)
Superior/Global (Revisão de Escopo)	GenAI	Formativa, somativa	Feedback imediato, autoavaliação, percepção de imparcialidade.	Integridade acadêmica (54% dos artigos), dependência excessiva.	Xia et al. (2024)
Superior/Global	LLMs (Chat-GPT)	Formativa (QA de exames)	Geração de distratores e revisão de itens.	Opacidade (“caixa-preta”), inconsistência em questões complexas.	Willert & Wurz (2025)
Superior/Global	ITS, AES, LLMs	Formativa (feedback), diagnóstica	Mede competências complexas, reduz carga docente.	IA frágil, alucinações, vigilância ética.	Cardona et al. (2023)
Superior/ALC	Tutores, alerta precoce, gestão	Diagnóstica, formativa, somativa	Geekie e OE Saber personalizam rotas; Uplanner agiliza gestão.	Risco de desumanização; reforço de vieses; divisão digital.	Rivas (2025); Molina et al. (2024)
Superior/Global	IA Responsável (princípios)	Formativa, somativa, diagnóstica	Feedback perspicaz, inferências válidas.	Vieses algorítmicos, risco de sobrecarga tecnológica, isolamento social.	Johnson (2025)

Fonte: Elaboração própria

A aplicação da IA, contudo, não se restringe à sala de aula, tendo um papel igualmente transformador nas avaliações sistêmicas, como será explorado na próxima seção.

Tecnologias/IA em Avaliação em Larga Escala - Educação Básica

As avaliações em larga escala, como o Sistema de Avaliação da Educação Básica (Saeb) no Brasil e exames internacionais como o Pisa e o TIMSS, são instrumentos cruciais de política educacional. A Inteligência Artificial oferece um potencial sem precedentes para aumentar a eficiência, a profundidade e a equidade desses exames. A transição para plataformas digitais permite coletar novos tipos de dados e aplicar métodos analíticos avançados que prometem diagnósticos mais ricos sobre os sistemas de ensino. Esta seção explorará como os estudos relacionam IA e tecnologias com esses exames, discutindo as vantagens, os limites e os desafios emergentes.

A literatura recente destaca inovações metodológicas significativas. Uma edição especial do *International Electronic Journal of Elementary Education* aponta para o uso de dados de processo (como cliques e tempo de resposta) para analisar o comportamento dos examinandos e identificar padrões de engajamento ou desengajamento. Outro avanço é o emprego de modelos de machine learning, como o XGBoost, para prever com alta precisão quais estudantes necessitam de acomodações, como tempo estendido, promovendo maior equidade (Bezirhan; Şahin, 2025). Pesquisas com avaliações como o TIMSS demonstram como a combinação de tradução automática e correção automatizada de respostas abertas pode garantir consistência entre diferentes países e idiomas, além de reduzir custos operacionais (Jung et al., 2024; Jung et al., 2025).

No Brasil, a experiência do Centro de Políticas Públicas e Avaliação da Educação (CAEd/UFJF) é um exemplo de aplicação concreta dessas inovações. O centro tem desenvolvido tecnologias como o reconhecimento automático de fala para avaliar a fluência em leitura, o teste adaptativo computadorizado, que ajusta a dificuldade dos itens ao desempenho do aluno, e a geração automática de itens (Silva; Scortegagna, 2023). A experiência da França na transição para avaliações digitais em massa, por sua vez, levanta um desafio crucial: a comparabilidade dos resultados entre os formatos digital e impresso, já que os mesmos itens tendem a ser mais difíceis na versão digital (Bret, 2024). Além disso, o uso de grandes modelos de linguagem (LLMs) para gerar itens de leitura para exames como o PIRLS é explorado como uma promessa de eficiência, mas também é visto como uma fonte de ris-

cos significativos relacionados à validade, confiabilidade e segurança dos testes (Hao et al., 2024).

Nível/País	Tecnologia IA	Tipo de Avaliação	Principais Achados	Desafios	Autor(es)/Ano
Global (IL-SAs – TIMSS, PIRLS, Pisa, NAEP)	LLMs, ANNs, CNNs, AIG, ML, PLN, TA	Somativa (larga escala), processual, diagnóstica, adaptativa	Alta precisão em correção multilíngue ($r \approx 0,92$), redações gráficas (92%); AIG gera itens/distratores; análise de dados de processo melhora validade e equidade.	Fraude e uso indevido; complexidade em respostas intermediárias; dependência da qualidade de dados; altos custos e necessidade de literacia em IA.	Bezirhan von Davier (2023), Guo et al. (2024), Hao et al. (2024), Jung et al. (2024, 2025), IEJEE SI (2025), von Davier (2025)
Global (K-12 e Secundário)	GenAI (LLMs), AES, ITS, adaptativo	Formativa (predom.), diagnóstica, somativa, preditiva	Feedback imediato em larga escala; AES comparável a humanos; predição de evasão (até 95%); IA mede competências não cognitivas; frameworks (AIAS) apoiam uso ético.	Fragilidade (alucinação), declínio de escrita/pensamento crítico, limitação de formas de aprendizagem; pesquisa dominada por computação, pouca base pedagógica.	Alkan (2024), Cardona et al. (2023), Chen et al. (2020), Gonzalez-Cala-tayud et al. (2021), Hopfenbeck et al. (2023), Martínez-Comesaña et al. (2023), Owan et al. (2023), Perkins et al. (2024), Roe et al. (2025), Saputra et al. (2024)
América Latina e Caribe (Brasil, Chile, Uruguai, Colômbia, Equador)	Plataformas (Geekie, Letrus, Pontue, OE Saber, ALEKS), EWS	Diagnóstica, formativa, somativa	Correção automatizada de textos (Letrus, Pontue); feedback adaptativo (Geekie, OE Saber); SEA+; ALEKS melhora notas em Matemática; EWS identifica risco de abandono.	Divisão digital e infraestrutura precária; riscos à privacidade (Big Data); necessidade de letramento digital.	CAEd (2023), Molina et al. (2024), Rivas (2025), World Bank (2024)
França (básica – larga escala)	IA para geração de itens (AIG), E-assessment	Diagnóstica (nacional), adaptativa	Avaliação digital para 8 milhões de alunos; AIG gera itens matemáticos com controle de dificuldade; exploração de testes adaptativos.	Comparabilidade papel vs. digital; dificuldades na geração de distratores/textos originais.	Flip (2024 – França)
EUA (1º ano)	Deep Learning (LSTM), K-shape clustering	Formativa, diagnóstica, adaptativa	Previsão de necessidade de testes com ~90% de acurácia; otimização do valor diagnóstico.	Dificuldade docente na interpretação; risco de resistência dos alunos ao excesso de testes.	Shin et al. (2022)
Portugal (básica – larga escala)	Automação digital (IA na correção)	Formativa (low stakes), somativa (high stakes)	Aplicação digital para 80 mil alunos simultaneamente; correção automática das respostas.	Falta de computadores e formação digital; dificuldade de digitação em séries iniciais.	Flip (2024 – Portugal)

Fonte: Elaboração própria

Para além dos níveis de ensino e da escala de aplicação, a literatura apresenta outros eixos de análise fundamentais para uma compreensão mais ampla do fenômeno. Esses eixos serão abordados a seguir.

Outros Enfoques Analíticos sobre IA na Avaliação

Uma análise completa do impacto da IA na avaliação educacional requer a exploração de enfoques transversais que contextualizam, orientam e questionam sua aplicação. A tecnologia não opera no vácuo; ela é moldada por realidades regionais, guiada por modelos teóricos, validada por pesquisas empíricas e regulada por políticas e princípios éticos. Esta seção se aprofundará em cinco dimensões cruciais: o cenário específico da América Latina, os frameworks conceituais propostos para a integração da IA, as pesquisas exploratórias sobre LLMs, as diretrizes políticas emergentes e os princípios éticos que devem nortear todo o processo.

Contexto da América Latina e Caribe (LAC)

O cenário da Inteligência Artificial na educação na América Latina e Caribe é descrito como “em construção” (Rivas, 2025). Um mapeamento exploratório das iniciativas de avaliação na região revela algumas características recorrentes. Entre as principais plataformas identificadas estão as brasileiras (Letrus, Pontue, Geekie), a colombiana (OE Saber) e a uruguaia (SEA+). Uma análise dessas iniciativas mostra o predomínio de soluções desenvolvidas pelo setor privado, muitas das quais têm origem em tecnologias internacionais. O informe aponta que a região ainda está atrasada em sua capacidade de inovação autônoma em IA. Os desafios centrais para a expansão e apropriação dessas tecnologias incluem a persistente brecha digital, que limita o acesso equitativo, e a descontinuidade das políticas públicas, que dificulta a implementação de projetos de longo prazo (Rivas, 2025).

Frameworks conceituais para integração da IA

A literatura propõe frameworks com finalidades distintas: orientar a prática pedagógica, classificar o nível de automação, ou garantir a validação técnica. Para a prática em sala de aula, destacam-se a AI Assessment Scale (AIAS) (Perkins et al., 2024), que oferece um modelo de cinco níveis de integração da GenAI, desde a proibição até a colaboração plena, e a proposta de Page, Meyers & Billings (2024), que conecta o uso da GenAI à Taxonomia de Bloom para fomentar habilidades de ordem superior. Em contraste, modelos como o de Terpstra e Kruis (2024) focam em uma taxonomia técnica, classificando os níveis de autonomia da IA na correção, enquanto a abordagem de Willert e Würz (2025) se concentra na validação a priori, sugerindo o uso de testes bayesianos como método de garantia de qualidade para as tarefas avaliativas antes de sua aplicação.

Pesquisa exploratória sobre LLMs e avaliação

A ascensão dos grandes modelos de linguagem (LLMs) impulsionou uma onda de pesquisas exploratórias para avaliar suas capacidades e limitações. Um estudo que testou a viabilidade de usar LLMs para tarefas de avaliação de personalidade revelou inconsistências significativas e imprecisões computacionais, concluindo que esses modelos ainda não são confiáveis para avaliações formais (Lu; Fadel, 2025). Essa visão cautelosa é compartilhada por outros pesquisadores que, embora reconheçam o potencial transformador dos LLMs na geração de itens e na personalização da avaliação, alertam para riscos como “alucinações” (geração de informações falsas), vieses nos dados de treinamento e questões de segurança (Hao et al., 2024). Em uma perspectiva diferente, há quem veja a própria incerteza das respostas da IA como um indicador útil, sugerindo que a variabilidade nas classificações de um LLM pode, na verdade, sinalizar ambiguidades na própria formulação das questões de avaliação, servindo como um feedback para os elaboradores de itens (Willert; Würz, 2025).

Políticas e diretrizes para a educação básica

A rápida disseminação da IA nas escolas tornou urgente a criação de políticas e diretrizes para orientar seu uso. O relatório do Departamento de Educação dos Estados Unidos, *Artificial Intelligence and the Future of Teaching and Learning*, estabelece como princípio central a necessidade de manter humanos no circuito (*human in the loop*). A recomendação é que a tecnologia de IA deve sempre apoiar e aperfeiçoar, mas nunca substituir, o julgamento e a interação humana. O documento também enfatiza a importância de alinhar a IA a uma visão pedagógica clara e de promover pesquisas sensíveis ao contexto local (Cardona et al., 2023). Na mesma linha, o *Australian Framework for Generative Artificial Intelligence in Schools* é citado como um exemplo de iniciativa nacional que busca fornecer uma orientação estruturada para o uso responsável da IA nas escolas da educação básica (citado em Hopfenbeck et al., 2023).

Guias globais e princípios éticos

Há um forte consenso na literatura global sobre a necessidade de uma governança ética para a IA na educação. As diretrizes da Unesco fazem um chamado explícito para um uso da IA que seja inclusivo, equitativo e ético, em alinhamento com o Objetivo de Desenvolvimento Sustentável 4 (ODS 4) (Miao et al., 2021). Essa visão converge com os princípios de IA responsável propostos por organizações como a ETS, que detalham compromissos com a imparcialidade e a mitigação de vieses, a privacidade e a segurança dos dados, a transparência e a explicabilidade dos algoritmos, e com a responsabilidade pelas decisões tomadas com base em IA (Johnson, 2025). Esses princípios ecoam os alertas sobre riscos éticos — como viés algorítmico, vigilância de estudantes e aprofundamento de desigualdades — levantados em diversas outras fontes (Cardona et al., 2023; Hao et al., 2024; Rivas, 2025). E consolidam a visão de que a supervisão humana e a governança ética são pilares inegociáveis para a implementação da IA na educação.

A análise desses múltiplos enfoques permite, agora, traçar um balanço geral da literatura, identificando suas contribuições, lacunas e o caminho a seguir.

Esta revisão sistemática da literatura revelou que a aplicação de tecnologias e Inteligência Artificial na avaliação educacional é um campo dinâmico, multifacetado e em rápida expansão. A análise das diferentes categorias de estudo permitiu mapear um panorama complexo, repleto de oportunidades promissoras e desafios críticos. Esta seção final elaborará um balanço comparativo das categorias analisadas, destacará as tendências, lacunas e recomendações emergentes, e concluirá com uma reflexão sobre o futuro da avaliação educacional na era da IA.

Um balanço comparativo das áreas estudadas revela diferentes prioridades e focos, que refletem as missões distintas de cada nível. As avaliações em larga escala priorizam a eficiência e a comparabilidade para a formulação de políticas, utilizando a IA para automatizar a correção, analisar dados de processo e prever a necessidade de acomodações (Jung et al., 2025; Bezirhan; Şahin, 2025). O ensino superior, por sua vez, enfatiza a autenticidade e o alinhamento com o mercado de trabalho, integrando a GenAI em avaliações que espelham desafios profissionais e desenvolvem habilidades críticas (Page et al., 2024; Founng et al., 2024). Já a educação básica busca um delicado equilíbrio entre o apoio formativo individual e a preparação sistêmica, explorando ferramentas para diagnóstico precoce e feedback personalizado (Silva; Scortegagna, 2023; Liao et al., 2024).

Oportunidade ou benefício	Descrição e impacto estratégico
1. Personalização da aprendizagem e instrução adaptativa	A IA adapta métodos, conteúdos e avaliações às características singulares dos estudantes. Sistemas de tutoria inteligente oferecem suporte individualizado, ajustando-se ao nível de conhecimento e ao ritmo de cada aluno.
2. Aumento da eficiência administrativa e redução da carga docente	A automação de tarefas como registro de alunos, controle de frequência e relatórios libera tempo dos professores, permitindo foco em atividades pedagógicas complexas e atenção direta aos estudantes.
3. Melhoria da qualidade da avaliação e consistência (correção automatizada)	A correção automatizada de testes e produções escritas aumenta a precisão e a objetividade da pontuação, reduzindo vieses subjetivos e garantindo confiabilidade em contextos multilíngues.
4. Feedback imediato e personalizado	A IA gera devolutivas detalhadas sobre desempenho e estilo de escrita em tempo real, favorecendo o aprendizado contínuo e a autonomia dos alunos.
5. Análise de dados de processo e insights preditivos	Com base em grandes volumes de dados (tempo de resposta, cliques, uso de ferramentas), a IA identifica padrões de engajamento, contextualiza o desempenho e prevê riscos de abandono ou defasagem.
6. Avaliação de habilidades complexas e não cognitivas	A análise multimodal possibilita medir atributos como pensamento crítico, criatividade, colaboração e competências socioemocionais, antes difíceis de avaliar de forma sistemática.
7. Geração automatizada de itens e conteúdo	Ferramentas de IA generativa, em colaboração com especialistas, aceleram a criação de itens avaliativos robustos, textos de apoio e distratores plausíveis, além de apoiar a elaboração de planos de aula e tarefas.
8. Suporte à formação docente	A IA contribui para o desenvolvimento profissional de professores ao fornecer simulações, recomendações de práticas pedagógicas e sugestões de sequências e planos de aula adaptados às necessidades dos alunos.
9. Promoção da equidade e acessibilidade	Recursos como leitores de tela e síntese de voz ampliam a inclusão de alunos com necessidades especiais. Além disso, algoritmos podem apoiar avaliações mais justas e prever adaptações necessárias.
10. Inovação em avaliações e engajamento	A IA viabiliza experiências avaliativas dinâmicas e interativas, promovendo maior motivação e envolvimento dos estudantes em ambientes imersivos de aprendizagem.

Da análise transversal da literatura, emergem contribuições recorrentes, lacunas importantes e recomendações para o futuro, que podem ser sintetizadas da seguinte forma:

Contribuições recorrentes

- Fornecimento de feedback personalizado e imediato, superando as limitações de escala do ensino tradicional (Vittorini et al., 2021; Rivas, 2025).
- Automação de tarefas de correção, liberando tempo para que os professores se dediquem a atividades pedagógicas de maior valor, como a mediação e o acompanhamento individualizado (Ahmad et al., 2022; Chen et al., 2020).
- Análise de grandes volumes de dados para identificar padrões de aprendizagem, prever riscos de evasão e fundamentar políticas educacionais com base em evidências (Zawacki-Richter et al., 2019; Bezirhan; Şahin, 2025).

Lacunas identificadas

- Falta de estudos longitudinais que avaliem o impacto de longo prazo do uso da IA nos resultados de aprendizagem e no desenvolvimento dos estudantes (Zawacki-Richter et al., 2019).
- Carência de pesquisas fora do eixo do Norte Global e da Ásia, com o contexto da América Latina, por exemplo, ainda sendo descrito como “em construção” e com poucas iniciativas de pesquisa autônoma (Rivas, 2025).
- Insuficiente conexão entre o desenvolvimento tecnológico e as teorias pedagógicas consolidadas, com muitas aplicações focando em otimização técnica sem uma reflexão aprofundada sobre seus fundamentos e implicações educacionais (Zawacki-Richter et al., 2019; Hopfenbeck et al., 2023).

Recomendações futuras:

- Investimento urgente na formação de professores e no letramento em IA, capacitando os educadores para usar as novas ferramentas de forma crítica, ética e eficaz (Rivas, 2025; Cardona et al., 2023).
- Desenvolvimento de políticas e diretrizes claras em esferas institucionais, nacionais e global para governar o uso da IA, garantindo segurança, equidade e transparência (Miao et al., 2021; Johnson, 2025).
- Redesenho das avaliações para que sejam mais autênticas e focadas em habilidades de ordem superior, como pensamento crítico e resolução de problemas, em vez de meramente tentar torná-las “à prova de IA” (Page et al., 2024; Perkins et al., 2024).

Em conclusão, a literatura aponta para o papel estratégico e inevitável da Inteligência Artificial na reconfiguração da avaliação educacional. No entanto, essa transformação deve ser conduzida com rigorosos cuidados éticos e pedagógicos. Os alertas sobre viés algorítmico, privacidade de dados, equidade e, acima de tudo, a importância da supervisão humana são um consenso (Johnson, 2025; Cardona et al., 2023). A promessa da IA na educação só será plenamente realizada se a tecnologia for concebida e implementada para servir como uma ferramenta que amplia — e não substitui — o julgamento, a empatia e a interação humana que estão no cerne do processo educacional.

3. Entrevistas com Atores-chave

Metodologia de condução e sistematização das entrevistas

As entrevistas foram conduzidas de forma semiestruturada, seguindo um roteiro padronizado elaborado pela equipe do projeto. Cada entrevista teve duração média de 60 minutos e ocorreu por videoconferência, com consentimento dos participantes. O roteiro abrangeu perguntas sobre a descrição da iniciativa de avaliação educacional em que o entrevistado está envolvido, o tipo de avaliação realizada (p. ex., avaliação formativa em sala de aula, exames somativos de larga escala, avaliações processuais contínuas), o papel específico da IA (sobretudo IA generativa) nessas iniciativas, os resultados e evidências obtidos até o momento e os principais desafios enfrentados. Essa abordagem flexível permitiu explorar em profundidade as perspectivas dos entrevistados, mantendo comparabilidade nos tópicos centrais discutidos.

Para o exame dos depoimentos, adotou-se a técnica de análise de conteúdo qualitativa. Cada entrevista foi sintetizada em uma ficha analítica padronizada, contendo campos correspondentes aos principais eixos investigados (Ficha de análise da entrevista, disponível no Anexo I). Esses eixos incluíram: Iniciativa (descrição do projeto ou programa de avaliação mencionado), Tipo de avaliação (formativa, somativa, processual, etc.), Papel da IA (como e em quais etapas a IA generativa é empregada), Resultados (benefícios, evidências ou impactos observados) e Desafios (questões técnicas, éticas, operacionais relatadas), Recomendações e aprendizados (para políticas públicas, para educadores e para o desenvolvimento de tecnologias). As fichas analíticas facilitaram a comparação transversal entre entrevistas, permitindo identificar padrões e divergências. A partir das fichas, a equipe consolidou os achados em categorias temáticas gerais.

Achados e Aprendizados das Entrevistas

De modo geral, as entrevistas revelaram um panorama rico de iniciativas e perspectivas sobre o uso de IA na avaliação educacional. Identificou-se que grande parte das iniciativas mapeadas se concentra em avaliação formativa, ou seja, no apoio ao professor em sala de aula por meio de ferramentas inteligentes que geram feedback imediato aos alunos e adaptam atividades conforme o desempenho.

A maioria das experiências relatadas utiliza IA generativa para criar itens de avaliação (p. ex., questões de múltipla escolha, problemas abertos ou mesmo estruturas de simulados completos), visando enriquecer bancos de itens com maior rapidez. Nas iniciativas voltadas à avaliação somativa, destacam-se projetos em fase-piloto que empregam IA para correção automatizada de respostas discursivas, especialmente redações e perguntas abertas, buscando alinhar as notas automáticas às dos avaliadores humanos.

Em termos de resultados, os entrevistados apontaram benefícios como ganho de eficiência (redução do tempo de elaboração de provas e correção), personalização do processo avaliativo (itens ajustados ao nível do aluno) e obtenção de dados mais granulares sobre a aprendizagem. Alguns relataram evidências preliminares de melhora no engajamento dos estudantes e na qualidade do feedback ofertado aos docentes. Por outro lado, desafios importantes foram enfatizados: preocupações com a validade e confiabilidade das avaliações geradas por IA, a necessidade de treinamento de professores para utilizar essas ferramentas, a resistência cultural em aceitar decisões automatizadas e as questões éticas relativas à privacidade de dados e vieses algorítmicos.

Também foi mencionada a infraestrutura tecnológica desigual nas redes de ensino como fator que pode limitar a implementação ampla dessas inovações. Em síntese, as entrevistas reforçam o grande potencial transformador da IA generativa na avaliação educacional, ao mesmo tempo em que alertam para a necessidade de cuidados e estudos adicionais antes de sua adoção em larga escala, assim como adaptações ao sistema brasileiro, tendo em vista as desigualdades sociais.

No Anexo II, encontra-se a Ficha de análise individual das entrevistas. No Anexo III, encontram-se as sínteses individuais das entrevistas, que incluem o perfil do entrevistado e a identificação da iniciativa, seguida pelos principais pontos abordados nas escutas: papel da IA na iniciativa, resultados e desafios identificados.

Também foi construído um quadro-resumo, que apresenta uma visão panorâmica das entrevistas realizadas. O quadro inclui, além da identificação da iniciativa abordada por cada entrevistado, o tipo de avaliação enfocado, o uso específico da IA generativa, os resultados/benefícios reportados e os desafios mencionados.

Esse quadro permite visualizar comparativamente os diferentes contextos investigados. De forma geral, observa-se que cerca de metade dos entrevistados discute plataformas adaptativas ou sistemas de avaliação formativa impulsionados por IA; aproximadamente um terço aborda iniciativas de avaliação somativa automatizada (notadamente correção de provas escritas); e os demais se referem a usos mistos ou avaliação processual contínua assistida por ferramentas inteligentes.

Praticamente todas as entrevistas mencionam resultados positivos, ainda que iniciais. Os mais citados são a possibilidade de avaliar competências complexas, o aumento da celeridade e a redução de custos das avaliações. Também há ponderações sobre limitações e precauções necessárias, relacionadas a questões de segurança e ética e, para o caso brasileiro, desafios relacionados à infraestrutura, acessibilidade e letramento digital de professores e alunos.

4. Eventos de debate e disseminação

Evento 01

Workshop

“Como utilizar de modo adequado e ético a IA generativa para o desenvolvimento das avaliações educacionais”



Coordenação: Maria Helena Guimarães de Castro e Priscilla Tavares

Organização: Silvia Véspoli e Barbara Alves Medeiros

Apresentação: Leon Furze, da Deakin University, Austrália

Debatedores:

Alisson Santos (Fundação Itaú)

Guilherme Cintra (Fundação Lemann)

Ivan Claudio Pereira Siqueira (UFBA)

Priscilla Tavares (FGV/EESP)

Como parte das atividades de disseminação e debate, foi realizado em 12 de maio de 2025, das 19h às 20h30, o primeiro workshop do estudo, com o tema “Como utilizar de modo adequado e ético a IA generativa para o desenvolvimento das avaliações educacionais?”.

O evento foi coordenado por Maria Helena Guimarães de Castro e contou com a apresentação principal de Leon Furze, pesquisador da Deakin University, Austrália. Atuaram como debatedores convidados Alisson Santos, Guilherme Cintra, Ivan Siqueira e Priscilla Tavares, profissionais com atuação relevante nas áreas de educação, avaliação e tecnologia.

Os participantes do workshop abordaram boas práticas e diretrizes para a incorporação da IA generativa na avaliação educacional, enfatizando a importância de um uso ético e transparente dessas ferramentas. Destacaram-se os riscos potenciais, como vieses nos algoritmos, questões de privacidade dos dados de estudantes e a possibilidade de uso indevido da IA (p. ex., para fraudar avaliações), bem como estratégias para mitigar esses problemas. Por outro lado, foram discutidas diversas oportunidades proporcionadas por essa tecnologia, incluindo a personalização em larga escala do processo avaliativo, a redução da carga de trabalho dos professores em tarefas repetitivas de correção e a criação de itens e testes mais contextualizados às realidades dos alunos. Os debatedores concordaram que a IA generativa pode se tornar uma aliada valiosa na educação básica, desde que haja capacitação docente contínua, marcos regulatórios atualizados e critérios claros para sua aplicação responsável. O público presente pôde interagir via chat, trazendo questões que reforçaram pontos sobre a necessidade de validação científica das ferramentas de IA antes de sua adoção massiva e sobre o papel das instituições (como Inep e conselhos de educação) na orientação e supervisão dessas inovações. Ao final, a coordenadora Maria Helena Castro sintetizou os encaminhamentos, destacando a elaboração de um documento de princípios éticos para IA na avaliação, a ser discutido em futuros fóruns.

Resumo das discussões

O workshop teve início com uma introdução à temática central, a adoção de IA em avaliações educacionais, por Maria Helena Castro. Leon Furze proferiu a palestra principal, focando na integração ética da Inteligência Artificial nas práticas de avaliação. O professor relatou experiências bem-sucedidas na Austrália por meio da escala de uso de Inteligência Artificial proposta por ele.

Os participantes discutiram as diversas aplicações práticas da IA para otimização de avaliações, incluindo avaliações formativas e somativas. Guilherme Cintra enfatizou o conceito de “fricção ótima”, que visa utilizar IA para liberar recursos cognitivos de alunos para tarefas realmente significativas. Priscilla Tavares abordou o desafio brasileiro de incorporar IA nas salas de aula devido às disparidades de acesso e conectividade.

Ivan Siqueira levantou preocupações sobre questões éticas e de acessibilidade, principalmente para estudantes com necessidades especiais, um ponto essencial para a implantação segura e inclusiva de tecnologias de IA.

Alisson Santos apresentou as iniciativas da Fundação Itaú no uso de IA para avaliações, destacando a necessidade de desenvolver ferramentas adaptativas à realidade brasileira.

Sumário da palestra de Leon Furze

O professor Leon Furze, da Deakin University, Austrália, focou no uso estratégico e ético da Inteligência Artificial nas avaliações educacionais. Ele apresentou a AI Assessment Scale (AIAS)¹, desenvolvida em 2023 e atualizada em 2024, como um guia estruturado para a incorporação de IA em contextos educacionais. A escala, disponível na Tabela 01 a seguir, foi construída para ajudar professores e instituições na reflexão e planejamento sobre como a IA pode ser inserida nas avaliações de forma que realmente potencialize o aprendizado, ao invés de apenas automatizar os processos ou trapacear nas etapas cognitivas.

1 <https://leonfurze.com/2024/08/28/updating-the-ai-assessment-scale/>

Tabela 01 – Escala do uso de Inteligência Artificial em avaliação

1. Sem uso de IA	<p>A tarefa ou avaliação é realizada inteiramente sem o uso de IA, em ambiente controlado, garantindo que o produto gerado pelos estudantes se baseia exclusivamente em seu conhecimento, entendimento e habilidades.</p> <p>O aluno não deveria usar ferramentas de IA em nenhum momento nesta tarefa ou avaliação. Ele deveria demonstrar suas habilidades e conhecimentos.</p>
2. Uso de IA para planejamento	<p>A IA poderia ser usada para atividades pré-tarefa, tais como brainstorming, desenho ou pesquisa inicial. Este nível foca no uso efetivo da IA para planejamento, síntese e ideação, mas a avaliação da tarefa deve enfatizar a habilidade do estudante de desenvolver e refinar essas ideias de maneira independente.</p> <p>O aluno pode usar a IA para planejamento, desenvolvimento de ideias e pesquisa. O produto final deve refletir o quanto o estudante desenvolveu e refinou essas ideias.</p>
3. Uso de IA como colaboradora	<p>A IA pode ser usada para ajudar a completar a tarefa, incluindo a geração de ideias, rascunho, feedback e refinamento. Os estudantes devem avaliar criticamente e modificar os resultados sugeridos pela IA demonstrando seu entendimento.</p> <p>O aluno deve usar a IA para ajudá-lo com tarefas específicas, como o rascunho de um texto, o refinamento ou a avaliação de um trabalho. Ele deve avaliar criticamente e modificar o conteúdo gerado pela IA para obter seu produto final.</p>
4. Uso de IA como coautora	<p>A IA pode ser usada para completar qualquer elemento da tarefa, com os alunos direcionando a ferramenta para atingir os objetivos finais. Avaliações neste nível devem requerer engajamento dos alunos com a IA para resolver problemas e atingir os objetivos da tarefa.</p> <p>O aluno deve usar extensivamente a IA em seu trabalho como desejar, ou com direcionamento específico na sua avaliação. Foca no direcionamento das ferramentas de IA para atingir seus objetivos enquanto demonstra sua capacidade de pensamento crítico.</p>
5. Uso de IA para exploração	<p>A IA é usada de maneira criativa para aperfeiçoar o processo de resolução de um problema, gerar novos insights ou desenvolver soluções inovadoras para resolver problemas. Estudantes e educadores desenham em conjunto as avaliações para explorar de maneira única as aplicações da IA dentro deste campo de estudo.</p> <p>O aluno deve usar a IA de forma criativa para resolver a tarefa, potencialmente desenhando, em conjunto com o professor, novas abordagens para ela.</p>

Leon demonstrou a aplicação prática da Inteligência Artificial em escolas da Austrália, ressaltando seu uso crescente para fornecer feedbacks formativos rápidos e personalizados que aprimoram significativamente o processo de ensino-aprendizagem. Ele destacou a capacidade da IA em desenvolver competências de ordem superior, como pensamento crítico e criatividade, quando utilizada de forma equilibrada e crítica.

Outro ponto abordado por Leon foi a preocupação com a “atrofia cognitiva”, na qual os estudantes poderiam deixar de desenvolver habilidades essenciais se a tecnologia fizesse todo o trabalho. Ele propôs o conceito de optimal friction (fricção cognitiva/fricção ótima, em tradução aproximada) para evitar essa dependência da tecnologia. Furze sugere que a IA deve apoiar, mas nunca substituir o esforço cognitivo e o pensamento crítico necessários para um aprendizado profundo.

Ele destacou que no contexto educacional e tecnológico, optimal friction é usada para descrever o nível ideal de dificuldade ou desafio que promove aprendizado significativo. Ou seja, é quando a tarefa não é tão fácil que se torna trivial, nem tão difícil que se torna desmotivadora, justamente o ponto de equilíbrio que estimula o esforço cognitivo e o desenvolvimento do aluno.

Além disso, Furze enfatizou a importância de um design universal para garantir que a IA em contextos educacionais seja acessível e ética. E descreveu como a Austrália integra a tecnologia para beneficiar alunos com necessidades especiais. Ele sublinhou questões de ética e privacidade de dados, indicando que a IA deve ser aplicada de maneira a personalizar o aprendizado sem introduzir vieses ou comprometer a segurança dos dados dos alunos.

Ao lançar luz sobre como o uso significativo da IA deveria partir de objetivos pedagógicos claramente definidos, Leon convenceu os participantes da centralidade do planejamento educativo no uso de tecnologias. Ele delineou como a IA pressiona uma reformulação curricular, exemplificando mudanças necessárias para acompanhar os novos desafios do ensino contemporâneo.

Por fim, Furze destacou a necessidade de uma responsabilidade compartilhada entre professores, gestores, pesquisadores e formuladores de políticas públicas para criar um ecossistema educacional com IA responsável e eficaz. Ele incentivou o uso da escala apresentada como um ponto de partida para discussões estratégicas sobre a implementação da IA nas escolas, não

apenas enquanto ferramenta técnica, mas como um promotor de inovação ética e crítica dentro da educação.

Análise e relevância para a educação brasileira

O evento evidenciou que o uso da IA em avaliações educacionais pode revolucionar o modelo de ensino atual ao facilitar a personalização do aprendizado e otimizar recursos educacionais. No entanto, desafios significativos permanecem, particularmente em relação à equidade no acesso à tecnologia, formação de professores e infraestrutura digital na rede pública.

A discussão sinaliza que a IA pode ajudar a atenuar desigualdades educacionais se implementada de maneira inclusiva e ética. Para o Brasil, onde as diferenças regionais e socioeconômicas são marcantes, a adaptação de modelos e práticas de sucesso internacionais, como os australianos, deve ser feita cuidadosamente, levando em conta as especificidades locais.

O evento sublinhou a necessidade de integração das avaliações formativas e somativas através de tecnologias, permitindo feedbacks rápidos e precisos, essenciais para o acompanhamento eficiente do progresso educacional dos alunos.

O workshop “Como utilizar de modo adequado e ético a IA generativa para o desenvolvimento das avaliações educacionais” foi um evento importante no debate sobre a modernização das práticas de avaliação no Brasil. Destacou desafios e oportunidades que a IA traz ao campo educacional, oferecendo um caminho promissor para aprimorar a qualidade e a equidade do ensino no país. Reforça-se a importância de políticas públicas que acompanhem essas transformações, garantindo que todos os estudantes brasileiros possam se beneficiar do potencial revolucionário da Inteligência Artificial na educação.

Evento 02

Workshop

“Inteligência Artificial e avaliação educacional”



Coordenação: Maria Helena Guimarães de Castro e Priscilla Tavares

Organização: Silvia Véspoli e Barbara Alves Medeiros

Apresentação: Andreas Schleicher (diretor do Departamento de Educação e Competências da OCDE) e Tiago Caliço (psicometrista e analista da OCDE)

Debatedores:

Joaquim Neto (Abave e CAEd)

Lúcia Dellagnelo (OCDE)

Seiji Isotani (USP/São Carlos e NEES/Ufal)

No dia 24 de junho de 2025, das 10h às 12h, aconteceu o segundo workshop do presente estudo, intitulado “Inteligência Artificial e avaliação educacional: a experiência do Pisa/OCDE”. O evento, coordenado por Maria Helena Guimarães de Castro e Priscilla Tavares, teve como palestrantes principais, Andreas Schleicher, diretor do Departamento de Educação e Competências da OCDE e Tiago Caliço, psicometrista e analista da OCDE. Os debatedores convidados foram Joaquim Neto, Lúcia Dellagnelo e Seiji Isotani, especialistas que atuam na área da avaliação, educação e tecnologia.

Resumo das discussões

Durante o workshop, várias questões foram abordadas, especialmente em relação à transformação da avaliação educacional. Schleicher defendeu que a IA oferece uma oportunidade para integrar aprendizagem e avaliação,

promovendo um modelo mais contínuo, adaptativo e centrado no aluno. Caliço apresentou experiências práticas da OCDE no uso da IA em avaliações internacionais, como o Pisa, destacando avanços como a classificação automatizada de respostas, a criação de tarefas interativas e a promoção do letramento em IA. O debate entre os palestrantes Lucia Dellagnelo, Joaquim Neto e Seiji Isotani abordou os desafios pedagógicos e técnicos do uso de IA na educação e nos processos de avaliação educacional.

Sumário da palestra de Andreas Schleicher e Tiago Caliço

Andreas Schleicher destacou que um dos maiores equívocos históricos da educação foi a separação entre aprendizagem e avaliação. Segundo ele, a chegada da Inteligência Artificial (IA) representa uma oportunidade inédita de reunir esses dois processos, promovendo um modelo educacional mais integrado, contínuo e adaptativo. Schleicher defende que, com o uso da IA, é possível construir avaliações que não apenas verifiquem resultados, mas acompanhem e apoiem o processo de aprendizagem em tempo real.

Durante sua fala, ele ressaltou que a pandemia revelou a vulnerabilidade dos sistemas educacionais e trouxe à tona a necessidade de repensar o papel da educação frente aos desafios contemporâneos. Entre estes, as mudanças climáticas, a robótica e a própria IA, que se tornará cada vez mais presente em nossas vidas. Diante disso, o foco da educação deve se deslocar da simples memorização de conteúdos para o desenvolvimento de competências como pensamento crítico, resolução de problemas, criatividade e colaboração.

Para Schleicher, é essencial compreender o verdadeiro potencial da IA, que vai além de ferramentas específicas como o ChatGPT. A IA pode analisar padrões de comportamento, oferecer feedback imediato, monitorar progresso e adaptar tarefas ao nível do estudante. Isso permite uma avaliação mais rica, que observa não apenas se o aluno acertou ou errou uma resposta, mas como ele chegou àquela resposta, quais estratégias utilizou e como lidou com desafios.

Nesse contexto, surge o conceito de letramento em IA, desenvolvido pela OCDE. Trata-se da capacidade de utilizar, analisar, criar e refletir sobre a IA. A avaliação, segundo Schleicher, também precisa evoluir. Os testes tradicio-

nais, que exigem respostas fechadas e isoladas, estão sendo substituídos por tarefas mais autênticas, que colocam o aluno diante de problemas reais e complexos. Nessas atividades, os estudantes devem decompor problemas, usar dados, pensar computacionalmente e aplicar conhecimentos de diversas disciplinas. A IA atua como aliada nesse processo, ajustando o nível de dificuldade das tarefas conforme o desempenho do aluno e oferecendo caminhos alternativos quando ele encontra dificuldades.

Um ponto central de sua fala é a defesa de uma avaliação que se torne uma experiência de aprendizagem. Para isso, é preciso que a avaliação seja natural, contínua e integrada ao cotidiano escolar, permitindo que os alunos aprendam com seus erros, reflitam sobre suas estratégias e evoluam de forma individual.

Ao final, Schleicher reforça que o objetivo da educação deve ser preparar os jovens para o futuro, e não para o passado. Em um mundo em constante transformação, a competência mais importante será a capacidade de aprender rapidamente, adaptar-se a novas situações e aplicar o conhecimento de forma crítica e criativa.

O segundo palestrante, Tiago Caliço, apresentou um panorama dos projetos atuais e em desenvolvimento que exploram o uso da Inteligência Artificial no contexto do Pisa e em avaliações relacionadas conduzidas pela OCDE. Segundo ele, há três principais usos práticos da IA já em andamento: tradução e adaptação de itens (o que é chamado de transadaptação), classificação automática de respostas abertas, e criação de materiais de suporte em larga escala.

A transadaptação é fundamental devido à aplicação do Pisa em mais de 80 países e em 50 idiomas. A IA tem ajudado não apenas na tradução, mas na adaptação de textos e questões a diferentes contextos linguísticos e culturais, mantendo o mesmo nível linguístico e cognitivo.

Outro uso central da IA está na classificação automática de tarefas abertas, como ocorreu na avaliação de pensamento criativo no Pisa 2022. Nessa edição, estudantes precisavam propor títulos para uma imagem, uma tarefa difícil de classificar manualmente. A equipe liderada por Ricardo Primi conseguiu, com uso de IA, alcançar níveis de correlação entre classificadores humanos e máquinas semelhantes aos encontrados entre avaliadores hu-

manos treinados, demonstrando o potencial da IA para avaliações em larga escala.

Além disso, empresas como a Cambium Assessment, que trabalham com avaliações padronizadas nos Estados Unidos, já utilizam IA para classificar textos, detectar uso de IA por alunos, identificar sinais de sofrimento emocional (como relatos de violência ou bullying) e gerar feedback automatizado. Essas aplicações são possíveis graças a modelos treinados com dados de alta qualidade, sob protocolos rigorosos de privacidade, segurança e reprodutibilidade — evitando o uso de modelos comerciais hospedados na nuvem.

Tiago também destacou os avanços no uso de IA para a criação de materiais de aprendizagem e avaliação em larga escala. No Pisa 2029, um novo domínio será avaliado: literacia em mídia e IA. Para isso, estão sendo desenvolvidos protótipos com motores de busca simulados e conteúdos gerados por IA, controlados em termos de qualidade e relevância. Isso permite fornecer aos alunos fontes variadas de informação para que realizem tarefas como pesquisar e resumir conteúdos — habilidades centrais no mundo digital.

Outro projeto em desenvolvimento é o domínio “Aprender no mundo digital”, no qual os alunos deverão criar programas simples e modelar sistemas causais. Esses sistemas são interativos e fornecem feedback imediato, algo essencial para promover a autorregulação da aprendizagem. Um exemplo é o projeto “O poder do feedback”, que será implementado a partir de 2025. Nele, alunos do 4º e 5º ano realizarão tarefas de matemática com suporte de um agente de IA que fornece feedback em tempo real, comparando um grupo de controle com outro que recebe esse suporte interativo.

A colaboração também é um foco importante. A partir da experiência do Pisa 2018, que envolvia interação com agentes autônomos, a equipe desenvolve agora tarefas mais dinâmicas. Por exemplo, em um novo protótipo, o aluno interage com personagens IA em um ambiente virtual (como uma praia), buscando informações para resolver um problema real (redução da poluição), por meio de comunicação escrita ou oral.

QUADRO DE REFERÊNCIA

LITERACIA EM MÍDIA E INTELIGÊNCIA ARTIFICIAL

Refletir e agir de forma ética e responsável: interagir com conteúdo digital, plataformas de mídia e sistemas de IA de maneira responsável, justa e respeitando os valores humanos.

Acessar e usar: pesquisar e selecionar informações digitais de forma eficaz em um ambiente de realidade aumentada pela IA.

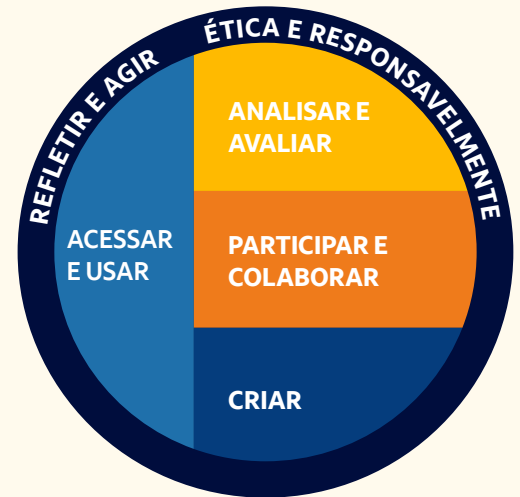
Analisar e avaliar: avaliar criticamente a fiabilidade, o viés e os objetivos de conteúdos digitais gerados por humanos e por IA.

Participar e colaborar: participar ativamente e colaborar com outras pessoas em conteúdos digitais, com consciência do papel da IA na mediação, amplificação e distorção de interações.

Criar: produzir conteúdos digitais com compromisso e de forma ética, complementando a criatividade humana com o uso aprimorado da IA.

Fonte: OCDE

VERSÃO ATUALIZADA



Por fim, Tiago enfatizou o desenvolvimento do quadro de referência para a literacia em mídia e IA, construído em parceria com a Comissão Europeia e a Code.org. Esse quadro orienta o desenvolvimento curricular e das avaliações, com base em cinco grandes eixos: agir de forma ética e responsável; acessar e usar IA; analisar e avaliar; participar e colaborar; e criar IA. O objetivo é preparar os alunos não apenas para compreender o funcionamento da IA, mas para usá-la de forma crítica, criativa e consciente — considerando também o papel que a IA exerce na produção de mídias e na forma como consumimos informação atualmente.

Após a apresentação dos palestrantes, os debatedores foram convidados a comentar os temas discutidos. A debatedora Lúcia Dellagnelo destacou a importância de diferenciar dois níveis de avaliação:

- 1 Avaliações internacionais em larga escala, como o Pisa, que visam comparar sistemas e gerar rankings.
- 2 A avaliação como parte do processo de aprendizagem, integrada à prática pedagógica cotidiana, em que o professor é o agente central.

Ela enfatizou que a avaliação deve funcionar como indutora da aprendizagem, especialmente quando acompanhada de feedback formativo — elemento que pode ser significativamente potencializado pela IA. No entanto, Lúcia alertou que, para isso, é preciso discutir quais competências e conhecimentos são necessários para que professores usem a IA de forma eficiente, e para que alunos desenvolvam letramento digital e compreensão crítica sobre a IA.

Ela também pontuou a necessidade de clareza sobre o que a IA pode e não pode fazer. Embora a IA avance rapidamente em várias tarefas, ainda há dimensões humanas que ela não alcança, o que reforça a ideia de que a IA não substituirá os professores. O desafio atual está em avaliar não apenas resultados, mas os processos e capacidades de aprendizagem, o que torna a avaliação mais complexa.

Já Joaquim Neto provocou uma discussão técnica com Tiago Caliço sobre como atribuir proficiência ao aluno ao final de tarefas complexas a partir desse novo tipo de avaliação desenvolvido pelo Pisa. Caliço então respondeu que existe a intenção de inovar (avaliar competências complexas por meio de tarefas abertas e interativas) e a necessidade prática de manter métricas comparáveis e escalas de proficiência, como a tradicional média 500 e desvio padrão 100 do Pisa. Ressaltou também que, embora haja vontade de avançar, essa mudança exige investimento político, científico e financeiro, e que muitos países ainda preferem manter os modelos tradicionais baseados na Teoria da Resposta ao Item (TRI).

Por fim, Seiji Isotoni comentou a frase de Andreas sobre “normalizar o fracasso” e destacou a importância de utilizar a IA para entender melhor os dados educacionais e, a partir disso, oferecer apoio personalizado ao aluno, ajudando-o a transformar dificuldades em motivação para continuar aprendendo, em vez de sentir frustração. E, lançou uma pergunta provocativa: ao apostar no uso massivo da IA para fins educacionais, quais capacidades cognitivas estamos dispostos a perder em troca de ganhos em eficiência? A provocação toca no equilíbrio entre o que se ganha com tecnologia e o que se pode comprometer do ponto de vista do desenvolvimento humano.

Evento 03

Workshop

“Inteligência Artificial e inovações nas avaliações do ETS: mudanças no NAEP”



Coordenação: Maria Helena Guimarães de Castro e Priscilla Tavares

Organização: Silvia Véspoli Godoy e Barbara Alves Medeiros

Apresentação: Luis Saldivia (ETS) e Matthew Johnson (ETS)

Debatedores:

Francisco Soares (UFMG)

Maria Inês Fini (Anebhi)

Ruben Klein (Abave)

No dia 21 de agosto de 2025, das 15h às 17h, o terceiro workshop “Inteligência Artificial e inovações nas avaliações do ETS – mudanças no NAEP” ocorreu de forma síncrona pelo canal do YouTube da Fundação Itaú. O evento, coordenado por Maria Helena Guimarães de Castro e Priscilla Tavares, contou com os palestrantes Luis Saldivia, especialista no desenvolvimento de avaliações, incluindo o NAPP (Avaliação Nacional do Progresso Educacional) dos Estados Unidos e Matthew Johnson, atual diretor do Centro de Pesquisa em Psicometria Fundamental e Estatística na Divisão de Pesquisa e Desenvolvimento do ETS. Os debatedores convidados foram Francisco Soares, Maria Inês Fini e Ruben Klein, especialistas nas áreas da avaliação e educação.

Resumo das discussões

Durante o workshop, os participantes discutiram de forma aprofundada o uso da Inteligência Artificial na avaliação educacional. Matthew Johnson apresentou como a IA generativa pode acelerar a criação de itens, automa-

tizar correções e fornecer feedback personalizado, destacando tanto os benefícios — como maior escala, rapidez e consistência — quanto os desafios, incluindo viés, calibração de itens e diversidade de dialetos. Maria Inês e Francisco Soares compartilharam experiências práticas no Brasil, ressaltando a importância de alinhar as ferramentas aos objetivos de aprendizagem, garantir que os alunos e professores estejam familiarizados com os recursos digitais e usar a IA de forma responsável para apoiar o aprendizado. Ruben reforçou a necessidade de integrar a análise de dados de processo e observação do comportamento dos estudantes com a avaliação digital, destacando que melhorias exigem equipes qualificadas e investimento contínuo. O debate final enfatizou que, embora a IA ofereça oportunidades significativas, sua implementação deve equilibrar inovação tecnológica com supervisão humana e rigor psicométrico.

Sumário da palestra de Luis Saldivia e Matthew Johnson

O primeiro palestrante, Luis Saldivia, começou explicando que sua exposição teria como foco a experiência da ETS no uso da Inteligência Artificial generativa e as transformações recentes no National Assessment of Educational Progress (NAEP), considerado o “boletim da nação” dos Estados Unidos.

Ele apresentou o NAEP como uma avaliação de caráter independente, conduzida pelo Centro Nacional de Estatística, ligado ao Ministério da Educação dos EUA. Desde 1990, o Centro oferece relatórios com detalhamento em nível estadual.

Saldivia destacou que a mudança mais significativa dos últimos anos foi a transição do NAEP do papel para o digital. Essa mudança trouxe grandes oportunidades, mas também desafios. Um dos principais objetivos era manter a comparabilidade histórica das tendências já consolidadas desde 1990. Para isso, foram conduzidos diversos estudos sobre o impacto da digitalização e sobre como adaptar itens tradicionais do impresso para o ambiente digital, sem comprometer a validade dos resultados.

Paralelamente, o processo também abriu espaço para a introdução de novos recursos tecnológicos, capazes de tornar a avaliação mais representativa e interativa. Entre os exemplos apresentados estão:

- Ferramentas de apoio ao estudante: calculadoras embutidas, recursos para destacar trechos na tela, paletas para símbolos matemáticos, funções de marcação e organização.
- Itens interativos: imagens dinâmicas, animações e simulações em vídeo que permitem explorar conceitos científicos em diferentes fases, como no exemplo da reação nuclear.
- Itens de zona: tarefas em que os estudantes clicam em regiões específicas de uma imagem para indicar a resposta correta, especialmente úteis em ciências e biologia.
- Simulações abertas: ambientes digitais nos quais os alunos podem manipular variáveis, coletar dados próprios e, a partir deles, construir respostas com base em evidências.

Segundo Saldivia, esses recursos não apenas ampliam as possibilidades pedagógicas, mas também permitem uma nova forma de coleta de dados. Ao invés de registrar apenas respostas finais, o sistema digital recolhe dados processuais (process data), como cliques, tempo de execução, sequências de ações e uso de ferramentas. Isso gera informações detalhadas sobre o engajamento dos alunos e suas estratégias de resolução de problemas. Ele mostrou, por exemplo, como estudantes de diferentes níveis de proficiência interagiram com uma mesma tarefa de representação gráfica em formato de pizza, destacando que os menos proficientes dependiam mais da calculadora e tinham maior dificuldade em completar a tarefa, enquanto os mais avançados realizavam a resolução de forma direta.

Essas análises possibilitam identificar problemas de design nos itens (como telas que exigem scroll excessivo, levando o aluno a sair e retornar) e construir perfis qualitativos do desempenho. Permitem, assim, compreender onde os alunos falham, como resolvem problemas e quais estratégias utilizam. O palestrante ressaltou que tais avanços seriam inviáveis em avaliações apenas em papel.

Outro ponto enfatizado foi a renovação do marco conceitual do NAEP, sobretudo na área de Matemática. Hoje, o foco vai além do conteúdo: avalia-se também os processos de raciocínio e as práticas matemáticas, como reco-

nhecer e traduzir representações, justificar resultados, modelar situações do mundo real e descontextualizar problemas. A avaliação também passou a incluir o conceito de matemática colaborativa, entendida como um processo social, em que os estudantes precisam interpretar contribuições de colegas, avaliar ideias, modificá-las e trabalhar conjuntamente para resolver problemas complexos.

Esse novo paradigma exige tarefas mais autênticas, baseadas em cenários interativos que simulam situações reais. Tais tarefas são estruturadas em múltiplas etapas, permitindo avaliar desde a compreensão inicial até a modelagem matemática e a solução final. A integração de dados processuais possibilita capturar não apenas se o aluno acertou ou errou, mas como ele pensou, interagiu e se engajou no percurso da resolução.

Luis Saldivia concluiu reforçando que a avaliação digital transforma a própria natureza da mensuração educacional: em vez de apenas verificar resultados, busca-se capturar a aprendizagem em movimento, valorizando processos cognitivos, afetivos e sociais. Essa abordagem requer equipes multidisciplinares (designers, especialistas em acessibilidade, analistas de dados, psicometristas) e tem como objetivo último compreender o modo como os estudantes aprendem, interagem e enfrentam desafios. Segundo ele, é nesse ponto que a IA tem potencial de ampliar ainda mais a capacidade das avaliações, oferecendo análises mais rápidas e detalhadas sobre o comportamento dos estudantes.

Johnson iniciou sua apresentação destacando a definição de Inteligência Artificial proposta pela OCDE, lembrando que muitos dos conceitos envolvidos já vinham sendo explorados em práticas de avaliação educacional há décadas. No entanto, sublinhou que os avanços recentes da IA generativa, como os modelos de linguagem, geração de imagens, áudios e vídeos, vêm trazendo transformações profundas para o campo da avaliação.

A partir desse enquadramento, apresentou os quatro principais benefícios que a IA pode oferecer às avaliações, sintetizados nos chamados “4S” (das iniciais em inglês):

- velocidade na produção e correção de instrumentos;
- escala, ao permitir atender um número muito maior de estudantes;
- síntese, na medida em que integra diferentes fontes de dados sobre respostas e processos; e
- surpresa, pois abre a possibilidade de novas descobertas, ao mesmo tempo em que impõe o desafio de lidar com vieses e alucinações dos modelos.

Com base nesse arcabouço, descreveu quatro frentes de aplicação da IA que já estão em curso. A primeira refere-se à pontuação automática de respostas escritas e orais, área em que a ETS acumula mais de 25 anos de pesquisa. Essa abordagem garante consistência e rapidez na correção, mas traz o risco de enviesamento por padrões linguísticos ou demográficos e a dificuldade de transparência na decisão dos algoritmos.

A segunda frente é a geração automática de itens, potencializada pela IA generativa, capaz de criar milhares de questões em curto prazo. Apesar do ganho de diversidade, essa estratégia demanda rigoroso controle de qualidade e monitoramento da validade dos itens. A terceira aplicação diz respeito ao monitoramento da integridade acadêmica, em que ferramentas de IA podem detectar plágio, padrões de resposta suspeitos e até comportamentos inadequados por meio de áudio e vídeo. Esse recurso contribui para a segurança dos exames, mas levanta preocupações éticas quanto à privacidade, à possibilidade de falsos positivos e ao impacto desigual em diferentes grupos de estudantes. A quarta área envolve a análise dos processos de resposta, na qual a IA viabiliza o exame das estratégias cognitivas e do engajamento dos alunos durante a realização da prova, permitindo gerar feedbacks formativos mais ricos e personalizados.

Na parte final, Matthew destacou novas tendências em desenvolvimento, como a avaliação baseada em conversação com agentes de IA, que pode ser usada tanto para medir competências de alunos quanto para apoiar o trabalho de professores; as avaliações gamificadas, que exploram ambientes digitais como o Roblox para observar habilidades socioemocionais em contextos mais autênticos; e, sobretudo, a necessidade de assegurar o uso responsável

da IA em avaliações, de modo a garantir que sua adoção esteja alinhada ao construto que se pretende medir, e não apenas guiada por razões de eficiência ou moda tecnológica.

Encerrando sua fala, reforçou que a questão central não é apenas o que a IA pode fazer, mas sim como pode ser usada de maneira transparente, justa e educacionalmente significativa, preservando o rigor científico e a equidade que devem sustentar qualquer sistema de avaliação.

Francisco Soares iniciou sua fala reconhecendo a relevância das apresentações de Luis Saldivia e de Matthew Johnson, destacando a referência do ETS na pesquisa e desenvolvimento de avaliações educacionais. Ele contextualizou sua observação a partir de uma experiência no Brasil, conduzida na Fundação Getúlio Vargas em parceria com o professor Jorge Lira, voltada à avaliação formativa, especialmente na leitura.

Soares explicou que, embora seja possível avaliar leitura com métodos tradicionais de papel e lápis, o grande desafio está na devolutiva personalizada para cada aluno. Para atender a milhares de estudantes, a Inteligência Artificial se torna essencial, permitindo gerar orientações específicas com base nas respostas dos alunos, algo inviável em larga escala apenas com professores. Esse uso da IA visa ampliar o alcance da avaliação formativa, evitando que ela se reduza apenas a testes de múltipla escolha.

Francisco também destacou que, para a aplicação da IA, ainda é necessário transformar as respostas escritas em arquivos processáveis por algoritmos de linguagem natural, enfrentando desafios como o “dialetto próprio” dos estudantes. A IA não substitui o professor; pelo contrário, ela potencializa a atuação docente, permitindo que os professores interpretem e utilizem os resultados para apoiar a aprendizagem.

Outro ponto importante de sua fala foi sobre a produção de itens de avaliação. A IA generativa permite criar conjuntos de perguntas mais consistentes e alinhadas à compreensão leitora, focando não apenas em habilidades isoladas, mas na evidência de aprendizado real. Apesar das possibilidades tecnológicas, Francisco enfatizou a importância de utilizar textos autênticos e manter o foco no aprendizado do aluno, evitando medir apenas a posição inadequada de grandes grupos.

Por fim, ele reforçou que avaliações bem planejadas e suportadas por IA têm impacto direto na sala de aula: o que se mede e testa deve necessariamente influenciar o ensino, mudando práticas pedagógicas e apoiando efetivamente a aprendizagem.

Maria Inês começou destacando o prazer em ouvir a apresentação de Matthew Johnson e aproveitou para falar sobre sua experiência com itens de múltipla escolha. Ela comentou sobre seu entusiasmo com as ferramentas de Inteligência Artificial, como Gemini e Perplexity, que têm demonstrado grande capacidade de gerar itens de avaliação contextualizados, mesmo para estudantes de extratos socioeconômicos desfavorecidos. Ressaltou que, apesar da tecnologia ainda não permitir simulações animadas nos testes, os itens gerados pelas IAs são muito bem estruturados e coerentes.

Maria Inês explicou que utiliza essas ferramentas não apenas para criar itens, mas também para analisar erros dos alunos, seguindo sua perspectiva piagetiana de que o erro é um ponto de partida para o aprendizado. Destacou que a IA consegue produzir rapidamente itens de diferentes níveis de dificuldade, e que os especialistas fazem curadoria para garantir a qualidade. Por fim, ela levantou uma questão para os colegas Matthew e Luiz: se, ao aprimorar-se o uso da IA incluindo itens de resposta construída pelo aluno, será possível substituir o teste empírico de dificuldade dos itens, mantendo a confiabilidade e adequação das avaliações.

Ruben Klein enfatizou as oportunidades trazidas pela avaliação digital, especialmente em termos de análise de dados, produção de itens e estrutura das provas. Mas também ressaltou os desafios do Brasil, onde muitos alunos não têm familiaridade com computadores e com as tecnologias necessárias para interagir com testes digitais. Ruben destacou a importância de preparar os estudantes para essa experiência e trazer os recursos tecnológicos para a sala de aula, de forma a melhorar o aprendizado.

Ele apontou também a relevância da análise de erros, especialmente em itens de múltipla escolha, como forma de compreender o raciocínio dos alunos e aperfeiçoar a produção de itens. No entanto, alertou para a necessidade de calibrar adequadamente esses itens, garantindo sua qualidade e confiabilidade para avaliações em larga escala. Ruben concluiu reforçando que a melhoria da avaliação educacional depende de equipes grandes e bem estruturadas, investimento adequado e, principalmente, que os avanços

cheguem à sala de aula, capacitando professores e beneficiando efetivamente o aprendizado dos estudantes.

Chegando ao fim do workshop, os palestrantes ressaltaram a complexidade e os desafios da implementação de IA nas avaliações, especialmente no que se refere à pontuação automatizada e à geração de itens. Enfatizaram que, embora modelos generativos como Gemini e ChatGPT tenham grande potencial para criar rapidamente itens de avaliação, é essencial identificar previamente os problemas que podem surgir, calibrar os itens e testar sua eficácia com estudantes reais para garantir a confiabilidade da avaliação. Além disso, foi destacada a importância de que as interações digitais sejam intuitivas e alinhadas aos construtos que se deseja avaliar, permitindo que os estudantes demonstrem efetivamente seu conhecimento.

O próximo passo, segundo eles, seria integrar estatística, psicometria e IA de maneira unificada, combinando modelos preditivos com dados empíricos para criar avaliações mais precisas, seguras e escaláveis. E, assim, evitar a dependência de processos fragmentados de criação e validação de itens.

Evento 04

Mesa-redonda Abave

“Inteligência Artificial e avaliação educacional”



Coordenação: Maria Helena Guimarães de Castro e Priscilla Tavares

Organização: Silvia Véspoli e equipe Abave

Apresentação: Maria Helena Guimarães de Castro e Priscilla Tavares

Debatedores:

Alan Valadares (Fundação Itaú)

Ig Ibert Bittencourt (NEES/Ufal)

Jairo de Souza (UFJF e CAEd)

Thomaz Veloso (NEES/Ufal e Educometrika)

A mesa-redonda foi realizada no âmbito das atividades da XIII Reunião da Associação Brasileira de Avaliação Educacional (Abave), em São Paulo, no dia 09 de setembro de 2025, das 13h às 15h. As coordenadoras da pesquisa realizaram a apresentação dos resultados obtidos até aquele momento e convidaram os debatedores para apresentarem seus projetos e ações na área.

Na abertura da mesa-redonda, a professora Maria Helena ressaltou a relevância do tema da Inteligência Artificial, destacou sua importância para a avaliação educacional e discutiu seus avanços e desafios em um contexto de rápidas transformações. Ela enfatizou o papel central da Abave na promoção desse debate e lembrou que a avaliação deve estar sempre articulada às políticas educacionais, contribuindo para a melhoria da aprendizagem e para a redução das desigualdades. Ao contextualizar o debate, ela apresentou os resultados da pesquisa “Inteligência Artificial e avaliação educacional”, financiada pela Fundação Itaú. Nesse sentido, chamou a atenção para a necessidade de integrar inovação tecnológica e compromisso com a rea-

lidade das escolas brasileiras. A professora Priscilla Tavares, antes de passar a palavra aos debatedores, ressaltou que a pesquisa tem como objetivo buscar insumos e inspiração em experiências do uso de Inteligência Artificial que possam servir de subsídios para o aprimoramento das avaliações em larga escala brasileiras. Neste contexto, a Abave seria o fórum mais importante para disseminar e debater os resultados da pesquisa.

O primeiro debatedor, Alan Valadares, coordenador do Observatório da Fundação Itaú, iniciou sua fala destacando a atuação da instituição nas áreas de cultura e educação, entendidas como dimensões indissociáveis para o desenvolvimento humano. Enfatizou a relevância da tecnologia, em especial da Inteligência Artificial, como uma transformação comparável a grandes revoluções históricas, como a eletricidade ou a máquina a vapor. Ressaltou, entretanto, que, assim como nessas revoluções, há sempre o risco de ampliação das desigualdades, sobretudo diante da rapidez com que a IA avança e da dificuldade de garantir acesso equitativo a escolas e estudantes. Nesse sentido, apresentou iniciativas apoiadas pela Fundação, como editais para projetos de IA voltados à redução de desigualdades, a criação de um Observatório de Inteligência Artificial na Educação e parcerias com instituições nacionais e internacionais. Por fim, destacou a importância de articular cultura e criatividade ao uso da tecnologia, lembrando que a criatividade é uma característica essencialmente humana e deve ser estimulada como diferencial no mundo digital.

Jairo de Souza, pesquisador do CAEd e professor da área de Ciência da Computação, foi o segundo debatedor a realizar suas considerações. Ele destacou sua trajetória desde 2018 com uso de IA em avaliações educacionais, dando ênfase na avaliação de fluência leitora. Relatou como a IA tem sido utilizada tanto na construção de itens, que exigem critérios complexos, quanto na análise de áudios de crianças do 2º ano, processo que em 2025 passou a ser integralmente avaliado por IA. Entre as principais vantagens desse uso, Jairo ressaltou a rapidez na entrega de resultados — essencial para avaliações formativas —, a padronização das correções, garantindo maior equidade entre os alunos, e a ampliação da quantidade e qualidade das informações disponíveis. Ao mesmo tempo, apontou desafios, como a necessidade de tornar comandos avaliativos mais claros e objetivos para que possam ser interpretados corretamente pela máquina, além de refletir sobre como traduzir o enorme volume de dados gerados em informações

úteis para os professores. Ele encerrou reforçando que o debate sobre uso, proteção e custos dessas tecnologias precisa avançar junto com seu desenvolvimento.

Em seguida, o terceiro debatedor, Ig Bittencourt, fundador do Núcleo de Excelência em Tecnologias Sociais (NEES) e professor da Universidade Federal de Alagoas (Ufal) destacou a necessidade de ir além do entusiasmo em torno da IA generativa, e pontuou que a Inteligência Artificial tem uma longa trajetória de contribuições aplicadas à educação desde os anos 1970. Trouxe exemplos de abordagens consolidadas, como o acompanhamento passo a passo da resolução dos estudantes (interloop) e a análise de comportamentos em tarefas (on task/off task), que permitem entender erros, estratégias e níveis de engajamento. Relatou experiências práticas em correção de redações e manuscritos, mostrando que a combinação entre máquina e professor — a chamada inteligência aumentada — gera melhores resultados, ao reduzir a carga de trabalho docente e enriquecer os feedbacks.

Ig defendeu uma tese provocativa: a de que a avaliação, tal como estruturada hoje, precisa ser repensada (utilizou a seguinte expressão: “morte da avaliação”), pois muitas vezes reforça desigualdades e se transforma em um negócio mais do que em um instrumento pedagógico. Para ele, é urgente levar a equidade a sério e repensar o papel da avaliação como um bem público digital, com abertura de códigos, itens e dados, de modo a estimular um ecossistema educacional mais justo e colaborativo.

Para fechar a mesa-redonda, o quarto e último debatedor, Thomaz Velloso, pesquisador sênior do (NEES/Ufal) destacou a importância de repensar a avaliação educacional diante da IA, enfatizando que tecnologias sozinhas não resolvem desigualdades. Ele ressaltou que muitos professores não têm formação técnica em avaliação, o que dificulta o uso eficaz dos dados em intervenções pedagógicas. Com base em experiências em Sobral e no Ceará, Thomaz apontou que soluções de IA podem fornecer feedback rápido e detalhado sobre o desempenho dos alunos, permitindo que professores intervenham em tempo real e evitem que dificuldades se acumulem. Para ser efetiva, a IA deve considerar contextos de baixa conectividade e baixa competência digital dos docentes. O foco central de sua fala foi que a IA deve ser aplicada como ferramenta para reduzir desigualdades, ouvindo as necessidades da escola e do professor, e não apenas como tecnologia. Ele concluiu

chamando a comunidade educacional a agir de forma criativa e prática para transformar a avaliação em um instrumento útil e equitativo.

Por fim, a professora Maria Helena encerrou a mesa reforçando que decretar a “morte da avaliação” não faz sentido, pois a avaliação é essencial para medir se os alunos aprenderam. Ela destaca que a IA não substitui o chão da escola no contexto atual, mas pode ser usada para tornar avaliações nacionais mais rápidas, menos caras e mais efetivas, com devolutivas que auxiliem diretamente na melhoria da aprendizagem.

Ela critica a lentidão de avaliações como o Saeb, cujos resultados só são divulgados quase um ano depois, tornando difícil que os professores utilizem esses dados pedagogicamente. Maria Helena enfatiza a necessidade de eficiência e relevância na avaliação, para que ela realmente contribua para a aprendizagem, e não apenas para rankings ou interesses políticos.

Além disso, ressalta que tecnologias, incluindo IA, devem ser usadas para agilizar processos, reduzir custos e melhorar a qualidade das avaliações, mas que seu impacto depende da formação e compreensão do professor. Ela defende um equilíbrio: nem otimismo exagerado nem pessimismo absoluto, mas um uso estratégico das avaliações e das tecnologias para apoiar melhorias reais no ensino e na alfabetização. Conclui, reforçando que a IA e outras tecnologias educacionais devem promover transformação cultural e digital para que as avaliações cumpram seu papel pedagógico efetivamente, em vez de reforçar desigualdades.

5. Considerações finais

A pesquisa “Uso de IA generativa no aperfeiçoamento da avaliação educacional no Brasil” demonstrou que estamos diante de um ponto de inflexão histórico para as políticas de avaliação na educação básica. A incorporação da Inteligência Artificial — especialmente a IA generativa e os modelos de linguagem de larga escala (LLMs) — oferece possibilidades inéditas de integrar as dimensões formativa e somativa da avaliação, de aprimorar a eficiência dos sistemas educacionais e de ampliar a capacidade diagnóstica sobre as aprendizagens. Ao mesmo tempo, impõe desafios éticos, pedagógicos e estruturais que exigem um reposicionamento das políticas públicas e das práticas escolares.

Os achados da pesquisa — provenientes da revisão de 66 estudos internacionais, de 13 entrevistas com especialistas e de quatro eventos de debate — mostram que a IA vem reconfigurando o conceito mesmo de avaliação. A tecnologia deixa de ser apenas um instrumento de mensuração para tornar-se parte constitutiva do processo de aprendizagem. Essa mudança é visível em experiências internacionais consolidadas, como a Austrália, que desenvolveu um framework nacional para o uso ético e pedagógico da IA nas escolas. Ou como a França, que aplica avaliações digitais em larga escala para mais de 8 milhões de estudantes; e a OCDE, que vem incorporando IA no Pisa para criar tarefas interativas e automatizar a classificação de respostas abertas.

Essas experiências indicam três avanços principais:

- Integração entre aprendizagem e avaliação — conceito defendido por Andreas Schleicher, da OCDE —, em que a IA permite que o processo avaliativo ocorra de modo contínuo, adaptativo e contextualizado. As ferramentas analisam não apenas o resultado, mas o processo cognitivo e metacognitivo do estudante: suas estratégias, erros, hesitações e formas de raciocínio.

- Ampliação da validade e da profundidade das avaliações, por meio da análise de dados de processo, da correção automatizada de respostas abertas e da geração de itens com controle de dificuldade. Esses recursos tornam possível medir competências mais complexas, como pensamento crítico, criatividade, colaboração e resolução de problemas — competências centrais do século XXI.
- Aumento da eficiência e da capacidade diagnóstica, com ganhos de tempo e precisão. Sistemas de IA já atingem correlação superior a 0,9 com corretores humanos em tarefas de escrita, e permitem devolutivas imediatas que fortalecem o papel formativo das avaliações, algo que antes era inviável em escala.

Na educação básica, os benefícios mais promissores da IA residem na capacidade de personalizar o acompanhamento das aprendizagens e integrar avaliações formativas e somativas. Ferramentas como as da Austrália ou as desenvolvidas pelo CAEd/UFJF — incluindo reconhecimento de fala para leitura, testes adaptativos e geração automática de itens — demonstram que é possível combinar dados de desempenho em tempo real com sistemas nacionais de monitoramento.

Essa integração cria um ecossistema em que o professor recebe feedbacks imediatos e acionáveis, os estudantes são desafiados em seu nível adequado de proficiência e os gestores educacionais obtêm evidências consolidadas para políticas públicas. A IA, nesse sentido, deixa de ser um recurso apenas tecnológico e passa a ser um novo paradigma de avaliação educativa, em que cada interação do aluno se torna uma oportunidade de diagnóstico e aprendizagem.

Além disso, a IA vem possibilitando a análise de erros e processos, auxiliando o desenvolvimento do raciocínio matemático, da criatividade e das competências socioemocionais — dimensões que os modelos tradicionais de avaliação dificilmente capturavam. Experimentos conduzidos no NAEP (EUA), no PIRLS e no Pisa 2022 evidenciam que a IA pode interpretar padrões de engajamento e estratégias de solução de problemas, oferecendo uma visão mais holística do aprendizado.

Apesar das promessas, a implementação da IA na avaliação ainda enfrenta importantes obstáculos. No Brasil e em grande parte da América Latina, há atraso estrutural na digitalização das avaliações e baixa capacidade de inovação autônoma. A região é fortemente dependente de soluções privadas, com desigualdade de acesso, descontinuidade de políticas e fragilidade nas bases de dados.

As entrevistas realizadas reforçam que a infraestrutura tecnológica limitada, o letramento digital insuficiente de professores e gestores, e a resistência cultural à automação são barreiras centrais. Persistem ainda riscos éticos, como vieses algorítmicos, privacidade dos dados de estudantes e a possibilidade de “atrofia cognitiva”, termo usado por Leon Furze (Deakin University) para descrever a dependência excessiva de sistemas automatizados. Para evitar isso, Furze defende o princípio da fricção ótima — isto é, um uso da IA que desafie o estudante sem eliminar o esforço cognitivo necessário para aprender.

Foram observadas experiências internacionais inspiradoras, que podem oferecer pistas valiosas para o Brasil:

- **Austrália:** adoção de provas digitais com IA integradas à prática pedagógica e formação docente alinhada ao framework AIAS (AI Assessment Scale), que orienta o uso responsável da IA nas escolas.

- **Pisa e OCDE:** uso de IA para tradução, adaptação e correção automática de respostas, além do desenvolvimento do novo domínio “literacia em mídia e IA”, que inclui competências éticas e críticas.

- **NAEP (EUA):** análise de dados de processo e uso de machine learning para compreender o raciocínio dos alunos, não apenas o resultado final.

- **PIRLS e TIMSS:** integração de IA na geração e tradução automática de itens, aumentando comparabilidade e reduzindo custos.

- **França:** sistema nacional de e-assessment digital, que articula diagnóstico individual e monitoramento sistêmico em larga escala.

Essas experiências têm características comuns: clareza dos objetivos pedagógicos, forte investimento em infraestrutura e formação de professores, e políticas nacionais de governança da IA educacional.

Com base nas evidências analisadas, a pesquisa propõe um conjunto de recomendações estratégicas para o avanço do Brasil no uso da IA em avaliações educacionais:

- Modernização do Saeb e dos sistemas estaduais de avaliação, incorporando plataformas digitais adaptativas e análises de processo que unam diagnóstico, acompanhamento e planejamento pedagógico.
- Criação de uma estratégia nacional de IA na avaliação, com princípios éticos, transparência de algoritmos e protocolos de proteção de dados.
- Formação continuada e letramento em IA para professores, com foco na interpretação pedagógica dos dados, no uso ético das ferramentas e na integração das avaliações ao currículo.
- Investimento em infraestrutura tecnológica e conectividade, assegurando condições equitativas para redes e escolas públicas.
- Adoção de pilotos e laboratórios de inovação, em parceria com universidades e centros de pesquisa, para testar modelos de avaliação digital contextualizados à realidade brasileira.

A Inteligência Artificial está redesenhando a fronteira entre ensino, aprendizagem e avaliação. Quando orientada por objetivos pedagógicos claros e princípios éticos sólidos, ela oferece uma via concreta para uma educação mais personalizada, inclusiva e baseada em evidências.

No ensino superior, a IA tem avançado em direção a outros propósitos — como automação de correções, análise de código, predição de desempenho e desenvolvimento de habilidades profissionais —, mas é na educação básica que seu impacto social é mais profundo, por atingir a base formativa e ampliar as oportunidades de aprendizagem com equidade.

O desafio do Brasil, portanto, não é apenas adotar a tecnologia, mas redefinir o propósito da avaliação: de instrumento de classificação para instrumento de aprendizagem e equidade. Isso exige visão sistêmica, políticas integradas e compromisso com a formação humana integral.

Se guiada por essa perspectiva, a IA pode se tornar não apenas uma ferramenta de eficiência, mas um vetor de justiça educacional, capaz de alinhar qualidade, inclusão e inovação no século XXI.

6. Referências bibliográficas

AHMAD, Sayed Fayaz et al. **Academic and Administrative Role of Artificial Intelligence in Education**. Sustainability, v. 14, n. 1101, 2022. Disponível em: <https://doi.org/10.3390/su14031101>.

AHMAD, S. F. et al. **The role and impact of Artificial Intelligence in transforming the educational sector**. [Título e Fonte não especificados], 2021.

ALKAN, Ayşe. **Artificial Intelligence: Its Role and Potential in Education**. İnsan ve Toplum Bilimleri Araştırmaları Dergisi, [s.l.], v. 13, n. 1, p. 483-497, 2024. DOI: 10.15869/itobiad.1331201.

BEWERSDORFF, Arne et al. **Assessing student errors in experimentation using artificial intelligence and large language models: a comparative study with human raters**. Computers and Education: Artificial Intelligence, v. 5, p. 100177, 2023. Disponível em: <https://doi.org/10.1016/j.caeai.2023.100177>.

BEZIRHAN, Ummugul; VON DAVIER, Matthias. **Automated reading passage generation with Open AI's large language model**. Computers and Education: Artificial Intelligence, v. 5, p. 100161, 2023. Disponível em: <https://doi.org/10.1016/j.caeai.2023.100161>.

BEZIRHAN, Ummugul; ŞAHİN, Murat Doğan (ed.). **Large Scale Assessment: Challenges and Innovations**. International Electronic Journal of Elementary Education, v. 17, n. 2, mar. 2025.

BRET, Anaïs. **Student Assessment in France: Recent Developments, Challenges and Opportunities**. [S.l.: s.n.], jun. 2024. Apresentação de slides.

CARDONA, Susan A. et al. **Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations**. Washington, D.C.: U.S. Department of Education, Office of Educational Technology, 2023.

CHEN, Lijia; CHEN, Pingping; LIN, Zhijian. **Artificial Intelligence in Education: A Review**. IEEE Access, 2020. DOI: 10.1109/ACCESS.2020.2988510.

FELIX, Juri; WEBB, Laura. **Use of artificial intelligence in education delivery and assessment**. [S.l.]: Parliamentary Office of Science and Technology, 23 jan. 2024. Disponível em: <https://doi.org/10.58248/PN712>.

FOUNG, Dennis; LIN, Linda; CHEN, Julia. **Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices.** Computers and Education: Artificial Intelligence, v. 6, p. 100250, 2024. Disponível em: <https://doi.org/10.1016/j.caeai.2024.100250>.

GONZÁLEZ-CALATAYUD, V. et al. **The application of Artificial Intelligence in student assessment: A systematic review.** International Journal of Educational Technology in Higher Education, v. 18, n. 1, 2021.

GUO, Hongwen et al. **Large-Scale Assessments for Learning: A Human-Centred AI Approach to Contextualizing Test Performance.** Journal of Learning Analytics, v. 11, n. 2, 2024.

HAO, Jiangang et al. **Transforming Assessment: The Impacts and Implications of Large Language Models and Generative AI.** Educational Measurement: Issues and Practice, v. 0, n. 0, p. 1–14, Mar. 2024.

HOPFENBECK, Therese N. et al. **Challenges and opportunities for classroom-based formative assessment and AI: a perspective article.** Frontiers in Education, v. 8, 23 nov. 2023. DOI: 10.3389/feduc.2023.1270700.

HUANG, T. et al. **Empirical study of undergraduate students' use of ChatGPT in individual formative e-assessments.** Australasian Journal of Educational Technology, 2024.

INSTITUTO DE AVALIAÇÃO EDUCACIONAL (IAVE). **Digital transition of large-scale assessment in Portugal: a long and winding road.** [S.l.: s.n.], 2024. Apresentação de slides.

JOHNSON, Matthew S. **Responsible AI for measurement and learning: Principles and practices.** Princeton, New Jersey, United States: ETS Research Institute, ETS, 2025. (Research Report, RR-25-03).

JUNG, Ji Yoon; TYACK, L.; VON DAVIER, M. **Towards the implementation of automated scoring in international large-scale assessments: Scalability and quality control.** Computers and Education: Artificial Intelligence, v. 8, p. 100375, 2025.

JUNG, J. Y.; TYACK, L.; VON DAVIER, M. **Integrating Automated Translation with Neural Networks for Automated Scoring of Constructed-Response Items in Multilingual ILSAs.** Large-scale Assessments in Education, 2024.

LIAO, W. et al. **AI-enabled visual report tool to enhance learning performance and self-regulated learning in high school biology.** Journal of Computer Education, 2023.

LU, B.; FADEL, C. **Large Language Models as a Computational Tool for Exploratory Research in Educational and Psychological Measurement: Opportunities and Challenges.** *Journal of Educational Measurement*, 2025.

MADLAND, Colin et al. **Developing the Technology-Integrated Assessment Framework.** *Open/Technology in Education, Society, and Scholarship Association Journal*, [S.l.], v. 4, n. 1, p. 1-19, 2024. DOI: <https://doi.org/10.18357/otessaj.2024.4.1.63>.

MARTÍNEZ-COMESAÑA, Miguel et al. **Impact of artificial intelligence on assessment methods in primary and secondary education: systematic literature review.** *Revista de Psicodidáctica*, v. 28, p. 93–103, 2023.

MIAO, Fengchun et al. **AI and education: guidance for policy-makers.** Paris: Unesco, 2021.

MOLINA, E. et al. **AI revolution in education: What you need to know.** In: *Digital Innovations in Education*. Washington DC: Banco Mundial, 2024.

MONTOYA RAMÍREZ, Nancy Maritza et al. **Avaliação da aprendizagem na era da inteligência artificial.** *Polo del Conocimiento*, (ed. 92), Riobamba, v. 9, n. 3, p. 1977-1998, mar. 2024. DOI: 10.23857/pc.v9i3.6770.

OWAN, Valentine Joseph et al. **Exploring the potential of artificial intelligence tools in educational measurement and assessment.** *Eurasia Journal of Mathematics, Science and Technology Education*, [S.l.], v. 19, n. 8, em2307, 2023. Disponível em: <https://doi.org/10.29333/ejmste/13428>.

PAGE, Eric; MEYERS, Gretchen; KRAHE BILLINGS, Eve. **Theory to practice: An assessment framework for generative AI.** *Intersection: A journal at the intersection of assessment and learning*, v. 5, n. 4, p. 114-126, 2024.

PERKINS, Mike et al. **The Artificial Intelligence Assessment Scale (AIAS): a Framework for Ethical Integration of Generative AI in Educational Assessment.** *Journal of University Teaching and Learning Practice*, v. 21, n. 6, 2024.

PERKINS, Mike; ROE, Jasper; FURZE, Leon. **The AI Assessment Scale Revisited: a Framework for Educational Assessment.** 2024. Preprint.

RIVAS, Axel. **La llegada de la IA a la educación en América Latina: en construcción.** [S.l.]: ProFuturo - OEI, 2025.

ROE, Jasper; PERKINS, Mike; FURZE, Leon. **From Assessment to Practice: Implementing the AIAS Framework in EFL Teaching and Learning.** [S.l.: s.n.], 2025. Preprint.

SAPUTRA, Indra et al. **The Evolution of Educational Assessment: How Artificial Intelligence is Shaping the Trends and Future of Learning Evaluation.** The Indonesian Journal of Computer Science, v. 13, n. 6, p. 9056-9074, dez. 2024. Disponível em: <https://doi.org/10.33022/ijcs.v13i6.4465>.

SHIN, B. et al. **Improving prediction of student performance in computerized formative assessments to optimize teachers' test administration decisions using deep learning frameworks.** Journal of Computer Education, v. 9, n. 1, p. 71–91, 2022. DOI: <https://doi.org/10.1007/s40692-021-00196-7>.

SILVA, Hilda Aparecida Linhares da; SCORTEGAGNA, Liamara (org.). **Pesquisa aplicada a instrumento de avaliação educacional: desenvolvimentos e perspectivas com base na experiência do CAEd/UFJF.** Juiz de Fora, MG: Editora UFJF/ Gestão e Avaliação Educacional, 2023. Arquivo: 6,82 mb. 171 p.

TERPSTRA, et al. **Framework for categorizing the autonomy of AI in final exam scoring.** [S.l.: s.n.], 2025. Apresentação de slides.

VITTORINI, Pierpaolo; MENINI, Stefano; TONELLI, Sara. **An AI-Based System for Formative and Summative Assessment in Data Science Courses.** International Journal of Artificial Intelligence in Education, v. 31, p. 159–185, 2021.

VON DAVIER, Matthias. **Artificial Intelligence and Process Data Applications in TIMSS and PIRLS.** In: IRC 2025 TIMSS & PIRLS INTERNATIONAL STUDY CENTER TECHNICAL SYMPOSIUM, 2025, Boston. Anais... Boston: Boston College, 2025.

WILLERT, Nico; WÜRZ, Phi Katharina. **Assisting quality assurance of examination tasks: Using a GPT model and Bayesian testing for formative assessment.** Computers and Education: Artificial Intelligence, v. 8, 2025, 100343.

XIA, Qi et al. **A scoping review on how generative artificial intelligence transforms assessment in higher education.** International Journal of Educational Technology in Higher Education, v. 21, n. 40, 2024. Disponível em: <https://doi.org/10.1186/s41239-024-00468-z>.

ZAWACKI-RICHTER, Olaf et al. **Systematic review of research on artificial intelligence applications in higher education – where are the educators?.** International Journal of Educational Technology in Higher Education, v. 16, n. 1, 2019.

USO DE IA GENERATIVA
no aperfeiçoamento da avaliação
educacional no Brasil



ISBN: 978-85-7979-211-3



9 788579 792113