

ELABORAÇÃO DE ITENS DE MATEMÁTICA PARA AVALIAÇÕES DE LARGA ESCALA VIA INTELIGÊNCIA ARTIFICIAL GENERATIVA

DEVELOPMENT OF MATHEMATICS ITEMS FOR LARGE-SCALE
ASSESSMENTS THROUGH GENERATIVE AI

Matheus Maciel Maia

mmm6@discente.ifpe.edu.br

Emersson Rodrigues de Souza

emersson.souza@pesqueira.ifpe.edu.br

RESUMO

A presente pesquisa investiga a viabilidade do uso da Inteligência Artificial Generativa, especificamente o modelo Gemini, como ferramenta auxiliar na elaboração de itens de matemática para avaliações de larga escala, alinhados à Matriz de Referência do Sistema de Avaliação Educacional de Pernambuco (SAEPE). Diante da escassez de tempo dos docentes e do alto rigor técnico exigido na formulação dessas questões, o estudo objetivou analisar se a máquina pode otimizar a criação de material avaliativo. Caracterizada como uma pesquisa aplicada, exploratória e de abordagem qualitativa, a metodologia englobou revisão bibliográfica e uma fase experimental. Nessa etapa empírica, foram elaborados comandos (*prompts*) restritivos para a geração de itens focados em descritores do 9º ano do Ensino Fundamental. Os resultados demonstraram que a ferramenta é capaz de simular a estrutura exigida pelos guias oficiais, formulando enunciados contextualizados e distratores com precisão ao simular erros reais dos estudantes. No entanto, constatou-se a ocorrência de falhas estruturais e alucinações algorítmicas em algumas respostas. Conclui-se que a Inteligência Artificial é uma ferramenta educacional potente e viável para otimizar o tempo de planejamento, desde que sua utilização seja condicionada à curadoria do docente. O papel do professor transforma-se, tornando seu papel imprescindível na validação e no refinamento técnico-pedagógico do conteúdo gerado pela máquina.

Palavras-chave: Elaboração de Itens de Matemática. Avaliação de Larga escala. IA generativa.

ABSTRACT

This research investigates the feasibility of using Generative Artificial Intelligence, specifically the Gemini model, as an auxiliary tool in the development of mathematics items for large-scale assessments, aligned with the Reference Matrix of the Educational Assessment System of Pernambuco (SAEPE). Faced with the scarcity of teachers' time and the high technical rigor required to formulate these questions, the

study aimed to analyze whether the machine can optimize the creation of evaluative material. Characterized as applied, exploratory, and qualitative research, the methodology included a literature review and an experimental phase. In this empirical stage, restrictive *prompts* were developed to generate items focused on 9th-grade elementary school descriptors. The results demonstrated that the tool is capable of mimicking the structure required by official guides, formulating contextualized stems and distractors with precision by simulating real student errors. However, structural flaws and algorithmic hallucinations were observed in some responses. It is concluded that Artificial Intelligence is a powerful and viable educational tool to optimize planning time, provided its use is conditioned to the teacher's curation. The teacher's role is transformed, making their role essential in the validation and technical-pedagogical refinement of the machine-generated content.

Keywords: Mathematics Item Development. Large-Scale Assessment. Generative AI.

1 INTRODUÇÃO

As avaliações de aprendizagem são pilares fundamentais no processo educativo, atuando como instrumentos diagnósticos que fornecem dados essenciais para o planejamento pedagógico e a tomada de decisões no ambiente escolar (Luckesi, 2011; Hoffmann, 2005). No contexto das avaliações externas de larga escala, como o Sistema de Avaliação Educacional de Pernambuco (SAEPE), essa dimensão avaliativa ganha ainda mais relevância, pois busca identificar o desempenho dos estudantes em relação a um conjunto de habilidades específicas e selecionadas na Matriz de Referência. Para que essa mensuração seja precisa é necessária a construção de itens de múltipla escolha com alto rigor técnico e pedagógico, compostos por enunciados contextualizados, comandos claros e distratores plausíveis que mostram caminhos possíveis na resolução dos itens pelos discentes (CAEd, 2008; Brasil, 2010).

Contudo, a elaboração de itens com tal nível de excelência técnico-pedagógica constitui um desafio diário para os docentes da Educação Básica, que frequentemente dispõem de tempo escasso para conciliar o planejamento de aulas com a redação de questões inéditas e bem estruturadas. Diante desse cenário, a integração de novas tecnologias educacionais, em especial a Inteligência Artificial Generativa, ascende como uma alternativa promissora para otimizar o trabalho do professor (Silva, 2023; Lima, 2025).

Nesse contexto, a presente pesquisa tem como objetivo central investigar a viabilidade do uso da Inteligência Artificial Generativa, especificamente o modelo Gemini, como ferramenta auxiliar na elaboração de itens de matemática alinhados à Matriz de Referência do SAEPE para o 9º ano do Ensino Fundamental. O estudo busca compreender se, ao ser delimitada por um comando (*prompt*) restritivo, a máquina é capaz de simular a estrutura oficial dos testes de larga escala e criar distratores possíveis e viáveis.

A relevância e a justificativa deste trabalho residem na premissa de que a tecnologia, por mais avançada que seja, não anula a figura do educador. Pelo contrário, a pesquisa defende que a inserção da Inteligência Artificial no ensino da matemática transforma o papel do professor de um redator solitário para um validador técnico indispensável. Através de intervenções e refinamentos, demonstra-se que o

olhar humano é estritamente necessário para corrigir eventuais falhas de simetria e “alucinações” algorítmicas cometidas pela máquina.

Para alcançar os objetivos propostos, este Trabalho de Conclusão de Curso está estruturado em cinco capítulos principais. Após esta Introdução, o Capítulo 2 apresenta a Fundamentação Teórica, percorrendo sobre a estrutura dos itens de múltipla escolha, o funcionamento do SAEPE e os conceitos basilares da Inteligência Artificial. O Capítulo 3 detalha os Percursos Metodológicos, caracterizando a pesquisa como aplicada e exploratória (Gil, 2023), e definindo o fluxograma de interação com a IA. O Capítulo 4 expõe os Resultados e Discussões da fase empírica, trazendo a análise técnica e a intervenção docente nos itens gerados para os descritores selecionados. Por fim, o Capítulo 5 apresenta as Considerações Finais, sintetizando as descobertas e apontando propostas para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo é destinado a dar o embasamento teórico e conceitual necessário para sustentar a presente pesquisa. De modo a compreender a viabilidade de novas tecnologias no campo da educação, faz-se necessário debater, primeiramente, o contexto multidimensional da avaliação. Assim, o texto inicia-se dando uma abordagem geral acerca da avaliação e afunila-se, adentrando nas avaliações contínuas e, por conseguinte, nas avaliações de larga escala. Nesse contexto, é feita a análise da estrutura da menor engrenagem que compõe uma avaliação: o item de múltipla escolha.

Em um segundo momento, a fim de propor uma solução para os desafios enfrentados pelos docentes na elaboração de itens de matemática com alto rigor técnico, o capítulo adentra o campo da Inteligência Artificial (IA). Nele, são exploradas as suas estruturas mais básicas, o advento do Processamento de Linguagem Natural (PLN) e o funcionamento das IAs generativas. Por fim, o referencial teórico converge para a intersecção desses dois grandes eixos, apresentando e fundamentando como a IA pode servir como uma ferramenta viável para a criação de itens de matemática.

2.1 AVALIAÇÃO

Uma das práticas educacionais mais comuns, presenciadas desde a escola básica até cursos superiores, é a tão conhecida avaliação. Nos dias atuais, de modo geral, poderíamos diferenciar avaliações que ocorrem internamente das que ocorrem externamente à escola. Nesse contexto, faz-se necessário definir o conceito de avaliação sob a ótica da literatura especializada.

Para Luckesi (2011, p. 104), a avaliação é “[...] como um juízo de qualidade sobre dados relevantes, tendo em vista uma tomada de decisão”. Por outro lado, como afirma Hoffmann (2005, p. 17), a avaliação é a “[...] Reflexão permanente do educador sobre sua realidade, e acompanhamento, passo a passo, do educador, na sua trajetória de construção”. Ainda conforme Haydt (1988, p. 10), “Avaliar é julgar ou fazer a apreciação de alguém ou alguma coisa, tendo como base uma escala de valores”.

Diante dessas compreensões, este trabalho adota o entendimento de que avaliar transcende a mera verificação estática. Nesse sentido, a avaliação configura-se como um processo contínuo, dinâmico e analítico, funcionando como uma verdadeira ferramenta pedagógica para a coleta de dados, visando à tomada de decisões que promovam a melhoria educacional.

Os autores acima citados abordam várias dimensões da avaliação levando em conta o seu propósito, no entanto, Sousa (2000, p. 101) afirma que “cada escolha

define os métodos e a definição dos métodos que melhor possam atender ao tipo de decisão a ser tomada”.

Inclusive, Sousa (2000) classifica as dimensões da avaliação educacional de acordo com o espaço pedagógico definido pela sua atuação, conforme descrito a seguir.

Assim, a **avaliação de sala de aula** tem como foco o processo de ensino-aprendizagem e visa a subsidiar o aperfeiçoamento da prática docente; **a avaliação institucional** permite a análise da instituição educativa e indica a efetividade da instituição educativa no cumprimento de sua função social; **a avaliação de programa e projetos educativos** focaliza sua atenção nos propósitos e estratégias concebidos por determinado programa previsto para aperfeiçoar ou corrigir desvios de um sistema de ensino; **a avaliação de currículo** tem seu centro de atenção voltado para a análise do valor psicossocial dos objetivos e conteúdos propostos de um curso organizado para formar o aluno e para o estudo da efetividade dos processos previstos em sua implementação; **a avaliação de sistema** focaliza sistemas de ensino visando subsidiar políticas públicas na área educacional. (Sousa, 2000, p. 101-102).

Diante desse contexto, fica evidente a necessidade de clarificarmos as nuances entre a avaliação contínua e a de larga escala. A avaliação de sala de aula tem um propósito muito mais formativo, centrado principalmente em “permitir o acompanhamento escolar do aluno, possibilitando assim que se tome decisões de reajuste e de revisão durante o processo de ensino” (Sousa, 2000, p. 102). Todavia, a avaliação de sistema, por sua vez, abrange um conjunto significativamente maior, uma vez que é feita uma análise na organização educacional com o objetivo de coletar os dados e informações essenciais para criação de políticas públicas por parte do governo (Sousa, 2000).

Compreendendo o sistema de ensino como um todo, as avaliações de larga escala surgem como forma fundamental de identificar os entraves estruturais possíveis. É a partir dessa análise que o governo encontra uma forma de intervir. Conforme adverte Sousa (2000):

A desigualdade de educação oferecida nas escolas dos diferentes sistemas de ensino coloca para o estado o dever de compreender as dimensões dessas disparidades e tomar decisões no sentido de corrigir injustiças. Compete ao Estado, então, seja em nível nacional, estadual ou municipal, controlar a qualidade da educação oferecida em suas escolas e implantar medidas que permitam garantir a equidade no processo educacional. (Sousa, 2000, p. 113).

No Brasil, as avaliações externas remontam a um legado de décadas que teve início com o Sistema de Avaliação da Educação Básica (Saeb), gerido pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). O Saeb serve

de espelho para embasar de forma metodológica os sistemas estaduais (Silva, 2023), como é o caso do Sistema de Avaliação Educacional de Pernambuco (SAEPE¹).

Com o objetivo de compor esses testes cognitivos, faz-se uso da Matriz de Referência, que é um documento basilar para a construção dos itens. Diferente do que se imagina, a Matriz de Referência não se confunde com o currículo escolar (Silva, 2023), mas sim faz uso de um “recorte” assertivo (Pernambuco, 2025), no que tange às habilidades e competências específicas esperadas para o discente na etapa de escolarização específica, as quais podem ser mensuradas por meio de testes padronizados.

De modo que esses testes obtenham a excelência necessária para mensurar as habilidades exigidas pela matriz, fica evidente a demanda por um instrumento construído sob forte rigor técnico. E, tendo em vista essa necessidade, é imprescindível a análise da decomposição estrutural da menor engrenagem que compõe uma avaliação.

2.2 ESTRUTURA E ELABORAÇÃO DE ITENS DE MÚLTIPLA ESCOLHA

Ao analisarmos a fundo as avaliações educacionais, adentramos num campo essencial para sua estrutura: a unidade mais elementar, o item. Rotineiramente chamado de “questão”, ele é utilizado pelo professor com o objetivo de colher informações acerca do discente, embasar atitudes e nortear planejamentos (Silva, 2023).

Quando utilizado em avaliações de larga escala, como o SAEPE, o item adquire uma finalidade mais específica: a de ser um mecanismo de análise objetiva para verificar se o estudante conseguiu consolidar uma determinada habilidade, em conformidade com seu nível e com a metodologia da avaliação externa (Silva, 2023).

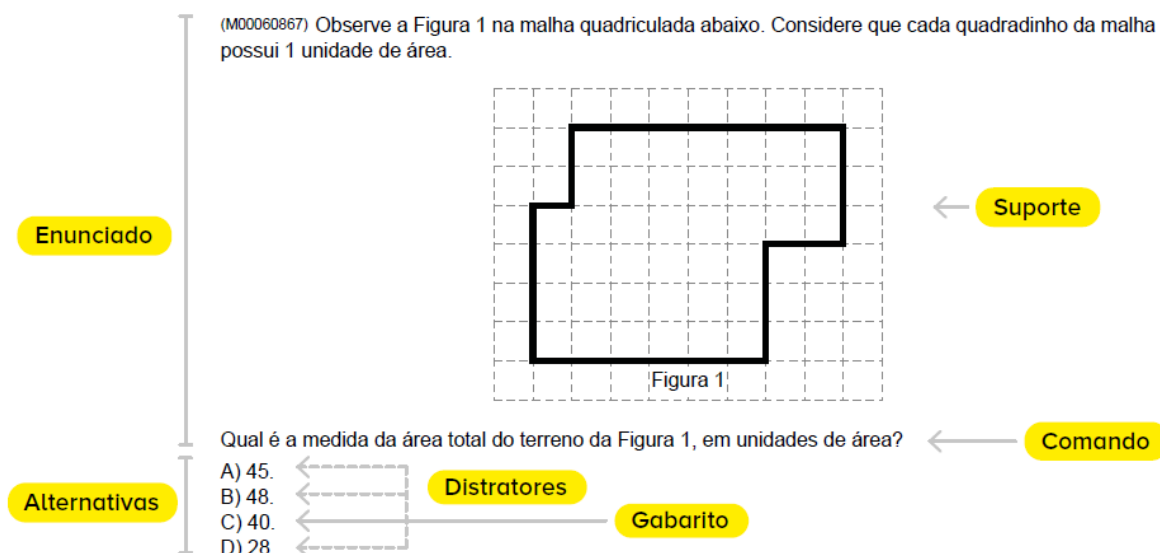
Ainda que as atividades avaliativas sejam utilizadas rotineiramente em sala de aula com uma estrutura mais flexível e foco no desenvolvimento individual, os itens de larga escala, em contrapartida, são mais rígidos e possuem enfoque no desempenho geral. Dentre as categorias de itens, o SAEPE e outras avaliações externas usam predominantemente os itens de resposta orientada, popularmente conhecidos como itens objetivos ou de múltipla escolha (Pernambuco, 2025).

O SAEPE, como avaliação externa, possui a unidimensionalidade como característica essencial de seus itens. Em outros termos, cada item deve avaliar uma única habilidade. Conforme detalhado no guia do programa (Pernambuco, 2025), se os itens apresentassem multidimensionalidade, os resultados se tornariam inconclusivos, visto que não haveria como determinar com precisão qual habilidade os discentes não desenvolveram.

Com o objetivo de validar os itens de múltipla escolha, sua construção deve seguir uma precisão técnica de excelência e estrutura bem definida. Em conformidade com o SAEPE (Pernambuco, 2025), sua configuração deve ser composta por:

¹ O Sistema de Avaliação Educacional de Pernambuco (SAEPE) é uma avaliação externa em larga escala criada com o propósito principal de atender aos estudantes do estado de Pernambuco, de forma que contemple todas as especificidades locais. Assim como outras avaliações externas, é uma ferramenta que auxilia na tomada de decisões a partir dos resultados de suas avaliações. Essas decisões contemplam os planejamentos escolares, o aperfeiçoamento do processo de ensino e aprendizagem, a identificação de desafios e a definição de estratégias que visam garantir a aprendizagem. (Pernambuco, 2025). Disponível em: https://prototipos.caeddigital.net/arquivos/pe/colecoes/2025/SAEPE_2025_RE_MT.pdf. Acesso em: 13 maio 2026.

Figura 1: Estrutura de um item.



Fonte: Pernambuco (2025, p. 14)

- Enunciado: é a porção inicial que apresenta as informações necessárias para a resolução do item, englobando outros dois recursos basilares: o suporte e o comando.
- Suporte: apresenta-se como texto, imagem, gráfico ou outro recurso essencial para a resolução do item. Ele contextualiza o cenário para o estudante.
- Comando: é, em sua essência, a pergunta ou instrução que aponta de forma clara, coesa e objetiva a tarefa que o discente deve realizar. Ele está intrinsecamente ligado à habilidade que o item deseja analisar.

Em continuidade ao enunciado, apresentam-se as alternativas, que são as opções de respostas oferecidas. Em um item, deve haver exclusivamente uma resposta correta, que é chamada de gabarito (Silva, 2023). As outras opções são alternativas incorretas chamadas de distratores, as quais requerem um rigor técnico elevado em sua elaboração, pois precisam ser plausíveis, não sendo admitida a adoção de valores aleatórios ou absurdos. Segundo o SAEPE (Pernambuco, 2025, p. 15), os distratores devem corresponder a “raciocínios possíveis” por parte dos estudantes ou a erros comuns decorrentes da não consolidação de aprendizagens. Essa configuração torna possível que os próprios erros dos estudantes se tornem dados pedagógicos valiosos sobre o raciocínio equivocado adotado. Para formular esses itens com tamanho rigor técnico contendo um enunciado claro, comando coeso e objetivo e distratores plausíveis faz-se necessário muito empenho, esforço e tempo por parte dos professores. É diante desse desafio que as ferramentas tecnológicas contemporâneas ascendem como possíveis aliadas. E para compreender essa tecnologia devemos destrinchar o campo da inteligência artificial.

2.3 INTELIGÊNCIA ARTIFICIAL

De modo a adentrar no uso da Inteligência Artificial (IA) como ferramenta tecnológica para gerar itens de matemática, faz-se necessário clarificar concepções essenciais para o funcionamento da tecnologia e seus desdobramentos.

De acordo com Silva (2023), a IA vem de uma das ramificações que se encontra no campo da ciência computacional, focalizando em simular capacidades humanas, a exemplo do aprendizado e do raciocínio lógico. Avançando nessa assimilação para McCarthy (2007, p. 2, tradução nossa), a inteligência artificial “É a ciência e a engenharia de criar máquinas inteligentes, especialmente programas de computador inteligentes”. E ainda, para consolidar o paradigma contemporâneo em uma visão mais transversal, Russell e Norvig (2022, p. 4) definem a Inteligência Artificial puramente como “o estudo de agentes que recebem percepções do ambiente e executam ações”.

Diante dessas concepções teóricas, assume-se neste trabalho o entendimento de que a inteligência artificial supera as imitações passivas do ser humano. De fato, adota-se a concepção de que ela age como um agente que interpreta e responde de acordo com o estímulo feito em seu meio (os dados e os *prompts* usados). É com base nessa resposta analítica e na execução estruturada da sua resposta, que essa ferramenta tecnológica se consolida e se torna viável a fim de ser um apoio ao professor, principalmente na atividade de criação e revisão de itens matemáticos. Contudo, para que essa ferramenta inteligente compreenda os estímulos passados pelo professor nas suas criações pedagógicas torna-se necessária a interpretação da nossa forma de comunicação.

Um subdomínio da inteligência artificial que ficou muito conhecido foi o do Processamento de Linguagem Natural (PLN); ele trabalha identificando, lendo e processando textos em linguagem natural, como é o caso do português (Silva, 2023).

Recentemente, o subdomínio do PLN teve um crescimento que culminou no surgimento dos Grandes Modelos de Linguagem (GML ou LLMs, do inglês *Large Language Models*). Conforme Silva (2023), os LLMs são redes neurais artificiais que foram amplamente alimentadas com uma vasta quantidade de dados em forma de texto e que possuem a capacidade de, com um simples estímulo, produzir materiais coesos e consistentes sobre os diversos assuntos para os quais foram treinados. Quando os grandes modelos vão além desse passo inicial de resposta a partir apenas do que foram pré-treinados e passam a gerar material inovador, esse modelo é classificado como uma Inteligência Artificial Generativa (Lima, 2025).

Esse estímulo inicial usado para se comunicar com a ferramenta de IA Generativa é executado através da linguagem natural. Essa instrução ou questionamento dado à IA é chamado de *prompt* (Lima, 2025). Cada *prompt* gera um resultado diferente, que tem mais ou menos qualidade dependendo de como ele foi refinado; ou seja, a qualidade da resposta está diretamente ligada a quão bem foi produzida a pergunta (Lima, 2025).

De acordo com Lima (2025), algumas IAs, como o Gemini, deram um passo além e conseguem simultaneamente processar e gerar materiais através de diferentes fontes, sejam textos, áudios ou imagens. No ensino da matemática, tais características se fazem essenciais, pois possibilitam a compreensão de gráficos, figuras geométricas e outros suportes visuais, integrando o suporte visual às outras partes do item e podendo seguir o comando da questão de forma mais fidedigna.

Como foi descrito, a IA Generativa se faz uma ferramenta muito útil, que pode ser usada nos mais diversos temas. Porém, ainda que tenha essa excelente capacidade, ela não é infalível. O Gemini, assim como outros modelos multimodais, trabalha através da probabilidade estatística das palavras principais da sequência para gerar suas respostas, não atuando como um repositório de conhecimento factual, mas como um processador estatístico de padrões linguísticos (Lima, 2025). Por isso, ela pode apresentar falhas e inconsistências nos materiais gerados, como distorções da

realidade, erros de conteúdo e até referências inexistentes. Tal comportamento incompatível é comumente chamado de alucinação (Lima, 2025). Ademais, os modelos, através de suas bases de dados, podem trazer respostas com vieses preconceituosos (Silva, 2023). Assim, é de suma importância a validação de seus resultados pelos seres humanos.

Nas seções anteriores, detalhamos os elevados critérios e o rigor exigido nas avaliações externas e, conseqüentemente, na construção de seus itens que são sua unidade mais elementar. A partir desta necessidade de conectar o rigor avaliativo a supervisão do professor (devido às limitações da ferramenta), que a IA generativa emerge não como um substituto do professor, mas como sua maior aliada para otimização de tempo e criação de material pedagógico.

2.4 A INTELIGÊNCIA ARTIFICIAL GENERATIVA NA ELABORAÇÃO DE ITENS

Para efeitos deste trabalho, as buscas realizadas no repositório do programa de Mestrado Profissional em Matemática da Rede Nacional (PROFMAT²) evidenciaram que o tópico de Inteligência Artificial Generativa aplicada à criação de itens vem se destacando recentemente. Trabalhos como os de Carvalho (2025), Lima (2025) e Freitas (2026), exploraram o uso de IAs para gerar itens contextualizados, itens relacionados a Matemática Financeira e itens inéditos baseados no Enem respectivamente. Contudo no que tange às avaliações sistêmicas com alto rigor técnico, destaca-se a dissertação de Silva (2023), que trata especificamente da geração de itens através da IA generativa alinhada com a matriz de referência.

Em sua pesquisa empírica, o autor aborda a aplicação dos grandes modelos de linguagem na criação e revisão de itens de múltipla escolha, utilizando-se de seu *prompt* (comando em linguagem natural) autoral, para servir de guia para a IA gerar itens alinhados com as habilidades e os rigores técnicos do Saeb. O estudo culminou na aplicação dos 20 itens gerados em um teste para uma turma de 9º ano contendo 61 estudantes da rede pública do estado do Pará.

Os resultados demonstraram que a inteligência artificial, quando utilizada de forma autônoma apenas pelo intermédio do *prompt*, apresenta problemas iniciais em cada parte da estrutura do item: enunciado, comando, distratores e gabarito. Entretanto quando o item gerado passa pelo *prompt* e pela revisão do professor atinge níveis de excelente poder discriminativo e alta qualidade psicométrica similar aos itens originais do Saeb.

Buscando avançar nesse processo explorado por Silva (2023), destaca-se atualmente o uso do Gemini, uma inteligência artificial multimodal que foi criada e recebe atualizações constantes pelo Google (Lima, 2025). O Gemini, como modelo, se sobressai em comparação com os outros modelos basilares pelo seu acesso sem interrupções à internet, o que lhe traz a possibilidade de responder a perguntas mais atuais sobre os diversos temas, baseando-se em seus desdobramentos de busca e revelando os links das fontes utilizadas (Lima, 2025).

A fim de que a IA possa cumprir com o papel rigoroso de gerar itens, o professor deve fornecer um *prompt* já refinado. Para que a inteligência artificial generativa minimize falhas ou alucinações e possa fornecer itens com validade técnica elevada em conformidade com as avaliações externas, esse *prompt* deve ser refinado e conter

²Programa de pós-graduação stricto sensu em matemática, ofertado em rede nacional e reconhecido pela CAPES, que concede o título de Mestre em matemática, e visa atender prioritariamente professores de matemática da educação básica, especialmente os da rede pública de ensino. O repositório com as dissertações do programa está disponível em: <https://profmatsbm.org.br/dissertacoes/>. Acesso em: 27 abr. 2026

todas as características esperadas de um item de avaliação de larga escala (Silva, 2023). O Gemini, enquanto IA multimodal, cumpre bem esse papel criativo, pois há a possibilidade, mediante solicitação do professor, de gerar uma resposta personalizada seja ela criativa, informativa, concisa ou de uma forma que se adeque ao público-alvo (Lima, 2025).

Um ponto relevante do Gemini na área educacional é a sua forma de responder didaticamente. Segundo Lima (2025), em seus testes práticos, pôde-se observar que as justificativas das respostas apresentavam ampla riqueza de detalhes e explicações, tornando o processo muito mais completo. Apesar de sua alta aptidão, a utilização do Gemini a fim de gerar itens fica condicionada a uma revisão minuciosa, pois o modelo apresenta limitações quanto à formatação de símbolos matemáticos e ordenação matemática (Lima, 2025).

Embora a IA seja uma excelente ferramenta para otimizar o tempo, o papel do docente permanece insubstituível. Conforme adverte Silva (2023), os resultados obtidos na geração dos itens não são garantidos pelo uso indiscriminado da ferramenta, mas sim pelo cuidado, expertise, supervisão e pelo conhecimento pedagógico de quem a utiliza.

3 METODOLOGIA

Esta pesquisa se caracteriza como um estudo de natureza qualitativa e exploratória. O procedimento metodológico consistiu na elaboração e refinamento de *prompts* de comando submetidos à IA Generativa Gemini, do Google. A análise seguiu o critério de comparação técnica entre os itens gerados e as competências exigidas pela Matriz de Referência de Matemática do SAEPE.

Conforme Gil (2023), pesquisas desse tipo têm como objetivo a aplicação prática e, sobretudo, a solução para uma demanda específica. No contexto deste trabalho, busca-se solucionar um problema vivenciado por muitos docentes: o de gerar itens para avaliações, propondo o uso da Inteligência Artificial como um meio de otimizar o tempo gasto e manter o alto rigor técnico. Em relação aos seus objetivos, a pesquisa possui caráter exploratório, visto que investiga uma tecnologia recente cujas aplicações educacionais ainda estão sendo desvendadas.

No que diz respeito à abordagem, optou-se pela perspectiva qualitativa. Segundo Gil (2023), as pesquisas qualitativas priorizam o aprofundamento das descrições verbais em detrimento das métricas e dos resultados puramente mensuráveis, embasando-se na compreensão de nuances que não poderiam ser alcançadas unicamente por meios estatísticos.

Quanto aos procedimentos técnicos que compõem esta pesquisa, eles delimitam-se em bibliográfico e experimental. A pesquisa bibliográfica é fundamentada na revisão de literatura já publicada, que traz uma ampla gama de possibilidades, além de amparo e solidez em virtude do material usado como base (Gil, 2023). Adicionalmente, a etapa experimental, conforme Gil (2023), dá-se ao restringir um objeto de estudo e controlar seus diferentes aspectos, em caráter de observação exclusiva de seus resultados e de suas implicações.

3.1 PROCEDIMENTOS PARA O LEVANTAMENTO BIBLIOGRÁFICO

O levantamento de dados desta revisão ocorreu na base digital de dissertações do Programa de Mestrado Profissional em Matemática em Rede Nacional (PROFMAT). Priorizaram-se publicações mais recentes, compreendidas entre os anos de 2023 e 2026, devido à atualidade do tema e às suas constantes atualizações no campo da educação matemática por meio da Inteligência Artificial Generativa.

A fim de filtrar os trabalhos analisados, foram adotados critérios de inclusão e exclusão de caráter orientador. Quanto aos critérios de inclusão, destacam-se: (i) trabalhos que tratassem do tema de geração, revisão ou resolução de itens de matemática por meio direto da inteligência artificial; e (ii) trabalhos com foco e alinhamento às matrizes de referência e avaliações padronizadas. Já quanto aos critérios de exclusão, descartaram-se com maior relevância: (i) pesquisas sem a elaboração estruturada de questões; e (ii) produções que não apresentassem a validação prática do uso da inteligência artificial.

Na etapa de busca, utilizou-se o termo “Inteligência Artificial” na ferramenta de filtro do repositório e, subsequentemente, após análise individual, os termos “Matemática” e “Elaboração de itens”, que trouxeram o rigor adequado na delimitação do campo de atuação. Durante a exploração dos materiais filtrados, foram identificados trabalhos relevantes, como os de Carvalho (2025) e Freitas (2026), que discorreram acerca de itens relacionados à Matemática Financeira e itens inéditos baseados no Enem, respectivamente.

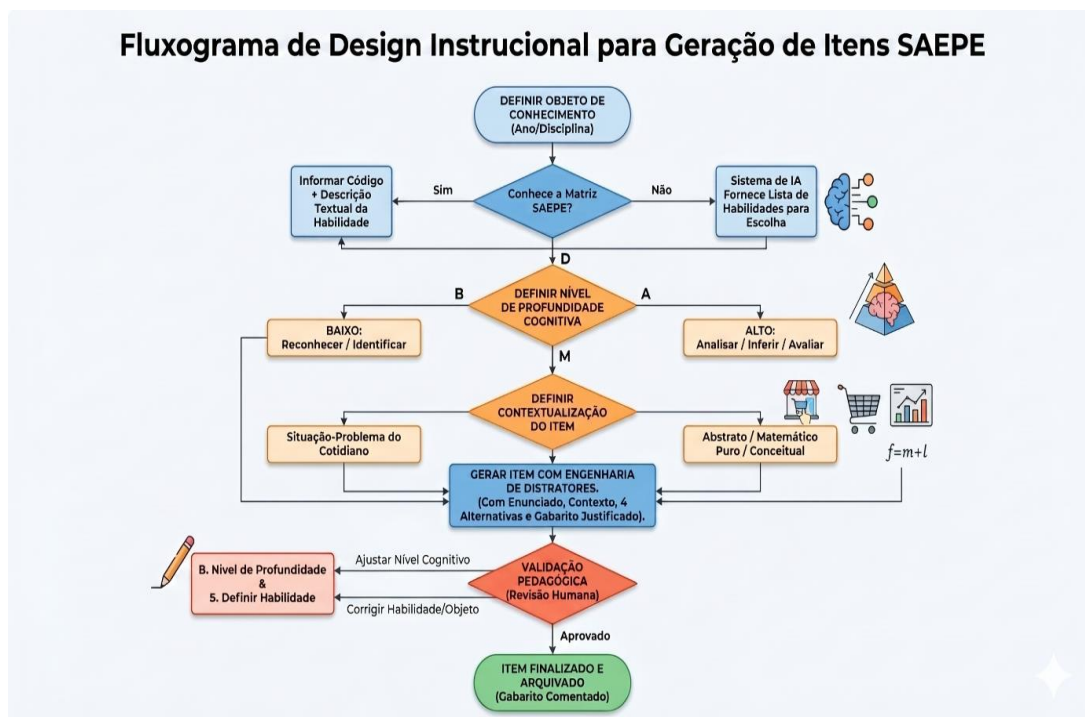
No entanto, ao aplicar com rigor os critérios de inclusão e exclusão, um trabalho destacou-se dos demais: a dissertação de Silva (2023). O estudo tratava não apenas da criação de itens utilizando a inteligência artificial com linguagem natural, como também aplicava o rigor técnico necessário aos itens de avaliações de larga escala e suas matrizes de referência através de um *prompt*. Diante disso, o trabalho de Silva (2023) foi selecionado como pilar metodológico desta pesquisa.

3.2 APLICAÇÃO PRÁTICA E ESTRUTURAÇÃO DO *PROMPT*

Buscando trazer mais solidez, o presente estudo apropriou-se da estrutura do *prompt* já consolidada por Silva (2023), na qual se seguia um roteiro passo a passo junto à máquina. No entanto, com o intuito de simplificar e otimizar o trabalho do docente, optou-se por uma adaptação metodológica para uma nova ferramenta, o modelo Gemini, e para um novo contexto: a geração de itens alinhados às habilidades e competências exigidas pela Matriz de Referência do SAEPE.

No modelo proposto por Silva (2023), o docente interage com a inteligência artificial usando um *prompt* autoral em linguagem natural, estabelecendo um diálogo em forma de “perguntas e respostas” que, por fim, traduz-se no item com todos os requisitos cumpridos. Em vez de manter essa interação sistemática, na presente adaptação o docente atua de forma mais direta, com uma instrução focada e de comando único. Conforme ilustrado no Fluxograma de Design Instrucional para Geração de Itens (Figura 2), o docente deve definir previamente o objeto de conhecimento, o descritor e a habilidade de forma integral, além do nível de profundidade cognitiva esperado para a questão e se ela será contextualizada ou puramente matemática, antes de qualquer interação com a máquina.

Figura 2 – Fluxograma de geração de itens.



Fonte: Elaborado pelos autores com auxílio da IA Gemini (2026)

Após o docente atender a esses requisitos iniciais previstos no fluxograma, inicia-se a etapa de definição das diretrizes finais junto à inteligência artificial do Gemini, por meio da emissão de um comando único em linguagem natural. A fim de garantir a reprodutibilidade do experimento, o *prompt* padrão adotado para a geração dos itens desta pesquisa foi estruturado da seguinte maneira:

Aja como um professor especialista em elaboração de itens de matemática para avaliações em larga escala. Sua tarefa é criar um item de múltipla escolha inédito e contextualizado para estudantes do 9º ano do ensino fundamental, alinhado rigorosamente à Matriz de Referência do SAEPE.

A habilidade a ser avaliada é o **[INSERIR CÓDIGO E DESCRIÇÃO DO DESCRITOR]**.

Nível de profundidade da questão: **[INSERIR NÍVEL: BAIXO, MÉDIO OU ALTO E UMA BREVE DESCRIÇÃO DA APLICAÇÃO]**.

Contextualização: **[INSERIR SUGESTÃO DE CENÁRIO OU CONTEXTO COTIDIANO PARA O PROBLEMA]**.

Em relação aos distratores: Os distratores devem ser baseados em erros matemáticos comuns nesta etapa de ensino. **[INSERIR ORIENTAÇÕES ESPECÍFICAS SOBRE OS ERROS QUE A MÁQUINA DEVE SIMULAR NAS ALTERNATIVAS INCORRETAS, EX: ERRO DE SINAL, ESQUECIMENTO DE FÓRMULA, ORDEM DE PRECEDÊNCIA]**.

O item gerado deve conter obrigatoriamente a seguinte estrutura: Enunciado e Suporte: Um texto claro, criando a situação-problema. Comando: A pergunta ou instrução objetiva. Alternativas: Quatro opções de resposta (A, B, C e D), sendo 1 gabarito e 3 distratores plausíveis baseados nos erros citados acima. Solução e Justificativa: O cálculo detalhado do gabarito e a explicação do erro hipotético exato de cada distrator.

Fonte: Elaborado pelos autores com auxílio da IA Gemini (2026)

Essa adaptação confere maior precisão e eficiência ao processo, tendendo a evitar alucinações da máquina, em virtude do maior controle das informações fornecidas previamente pelo docente. Com isso, a Inteligência Artificial é direcionada ao seu papel principal: a geração estrutural do item e a criação de distratores plausíveis. O ciclo encerra-se com a intervenção pedagógica para validação humana e os ajustes finais, de modo a atender a todos os requisitos técnicos esperados de um bom item do SAEPE.

Com a metodologia, a Inteligência Artificial Gemini e a estrutura de comando bem definidas, a pesquisa avança para a sua etapa experimental. No capítulo seguinte, serão apresentados, analisados e discutidos os resultados obtidos a partir da interação com essa ferramenta tecnológica emergente: a Inteligência Artificial Gemini.

Com o intuito de garantir a transparência, o rigor científico e a reprodutibilidade desta pesquisa, a transcrição integral de todos os comandos (*prompts*) e respostas geradas durante a fase experimental deste trabalho foi arquivada em nuvem. O registro autêntico da interação com a Inteligência Artificial Gemini encontra-se disponível para consulta pública através do link <https://gemini.google.com/share/1611f53a1e9c>. Acesso em: 04 de maio 2026.

4. RESULTADOS E DISCUSSÕES

Este capítulo se destina a exibir os resultados das interações com a inteligência artificial Gemini. A qual foi submetida ao *prompt* em linguagem natural a fim de gerar itens válidos tecnicamente e alinhados a matriz de referência do SAEPE para o 9º ano de ensino fundamental na disciplina de matemática.

Os dados obtidos demonstram que, ao fornecer comandos estruturados — contendo suporte, enunciado e comando — a IA foi capaz de reproduzir a estrutura técnica de um item de larga escala. No entanto, observou-se o fenômeno da alucinação (Lima, 2025) em distratores que envolviam cálculos de geometria complexa, o que reforça a necessidade da supervisão pedagógica mencionada por Silva (2023).

Quanto a avaliação de qualidade dos itens gerados, ficaram a cargo dos mesmos critérios técnicos de elaboração de questões consolidado por Silva (2023), os quais foram baseados em guias oficiais do Inep (Brasil, 2010) e do CAEd (2008). O foco das análises recai sobre o alinhamento das questões, as habilidades, clareza na linguagem, formatação, e mais incisivamente sobre os distratores.

4.1 AVALIAÇÃO DO DESCRITOR D08: PROPRIEDADES DOS POLÍGONOS

O primeiro teste busca avaliar o Descritor 08 do SAEPE que lida com problemas relacionados a propriedades dos polígonos. A interação junto a inteligência artificial do Gemini foi realizada através de bloco único conforme descrito na metodologia, e nela foi exigida além dos requisitos básicos, que fosse uma questão contextualizada e definiu restrições nos distratores a fim de minimizar possíveis alucinações, o resultado na íntegra está demonstrado a seguir.

Descritor 08: Resolver problema utilizando propriedades dos polígonos (soma de seus ângulos internos, número de diagonais, cálculo da medida de cada ângulo interno nos polígonos regulares).

Enunciado: Uma arquiteta está projetando o design de uma nova praça de alimentação em um shopping. O centro dessa praça contará com um grande mosaico no chão em formato de um hexágono regular. Para garantir o corte preciso das pedras

de mármore que compõem as bordas desse mosaico, a arquiteta precisa calcular a soma de todos os ângulos internos dessa figura.

Comando: Qual é a soma dos ângulos internos do mosaico hexagonal projetado pela arquiteta?

Alternativas:

- (A) 1080°
- (B) 720°
- (C) 180°
- (D) 9°

Solução e Justificativa:

Para resolver este item, o estudante deve aplicar a fórmula da soma dos ângulos internos de um polígono convexo:

$$S_i = (n-2) \times 180^\circ$$

Onde n é o número de lados do polígono. No caso de um hexágono, $n = 6$.

$$S_i = (6-2) \times 180^\circ$$

$$S_i = 4 \times 180^\circ$$

$$S_i = 720^\circ$$

Gabarito: Alternativa B

Análise dos Distratores (Justificativa do Erro):

- (A) **1080°:**(Erro de fórmula – esquecimento da subtração): O estudante utiliza a fórmula sem subtrair 2 do número de lados ($6 \times 180^\circ = 1080^\circ$). Este é um erro clássico de memorização da fórmula;
- (B) **720°:** GABARITO: Aplicação correta do conceito e da operação aritmética;
- (C) **180°:**(Confusão conceitual): O estudante confunde a soma dos ângulos internos de um polígono qualquer com a soma dos ângulos de um triângulo, ou interpreta erroneamente que “cada unidade” da fórmula equivale a apenas um triângulo isolado;
- (D) **9°:**(Confusão entre fórmulas distintas): O estudante confunde a fórmula da soma dos ângulos com a fórmula do número de diagonais: $d = \frac{n(n-3)}{2}$. Ao aplicar $d = \frac{6(6-3)}{2}$, ele encontra 9° e marca o valor numericamente igual, ignorando a unidade de medida em graus.

Fonte: Elaborado pelos autores com auxílio da IA Gemini (2026). Conversa original disponível em: <https://gemini.google.com/share/1611f53a1e9c>. Acesso em: 04 de maio 2026.

A fim de analisar a qualidade técnica do item gerado pelo Gemini, o mesmo será julgado com base nos parâmetros do Quadro 1, tomando como base os critérios apontados por Silva (2023, p. 51).

Quadro 1 – Validação Técnica do Item Gerado (Descritor D08).

Requisito Técnico (Inep/CAEd)	Status	Parecer Pedagógico
1. Alinhamento à Habilidade: O item avalia apenas a habilidade solicitada e não mede mera memorização?	Atende	O item exige a aplicação do conceito de soma dos ângulos internos de um polígono para o contexto de um polígono de 6 lados.
2. Clareza e Linguagem: Enunciado diversifica fontes (cotidiano), usa norma culta, sem jargões ou pegadinhas?	Atende	O item é claro e direto, sem pegadinhas e evidenciando situação cotidiana.
3. Formatação e Simetria: O comando é objetivo e as alternativas possuem a mesma extensão e padrão?	Não Atende	O comando é apresentado de forma interrogativa, as alternativas, no entanto não possuem mesma simetria, onde variam de 4 a 1 caractere, onde fica evidente a discrepância.
4. Plausibilidade dos Distratores: As alternativas incorretas refletem erros operacionais ou de raciocínios reais?	Não Atende	As alternativas incorretas atendem apenas parcialmente ficando a alternativa "A" como plausível e a "C" e "D" beirando uma alucinação da máquina.

Fonte: Elaborado pelos autores (2026).

O item gerado através do Gemini e o comando único em bloco, trouxe grandes expectativas quando consegue conciliar as abstrações conceituais da soma de ângulos de um polígono regular com um exemplo de criação de um projeto arquitetônico de um piso.

No entanto, o mesmo item que nos traz expectativa também nos faz observar alguns pontos o primeiro deles, que é de caráter opcional, seria a inserção de uma imagem para ilustrar a situação, tornando o item muito mais completo se tratando de um item da área de geometria. Essa mudança pontual pode ser gerada no próprio Gemini com o auxílio do docente, mas evidencia a limitação do comando.

O ponto mais grave se encontra nos distratores "C" e "D", a alternativa "C" apesar de plausível junto a seu distrator encontra-se longe da realidade onde trata da confusão do conceito de soma dos ângulos de um polígono regular com a soma dos ângulos internos de um triângulo qualquer 180° , plausível, mas distante do contexto da questão em que apresenta um hexágono regular, um polígono que pode ser decomposto em 6 triângulos iguais. Quanto a alternativa "D" a máquina alucina por completo onde ainda que seja plausível pois há essa possibilidade como na alternativa "C" na alternativa "D" se distancia ainda mais pois na alternativa "A" a alternativa tem 4 caracteres na "B" e "C" são 3 porém ao gerar a alternativa D, a inteligência artificial leva a uma resolução com apenas 1 caractere, ficando muito distante e afetando a simetria das alternativas como um todo, o ideal seria se manter nos 3 caracteres ou em uma tolerância chegando a 2 caracteres.

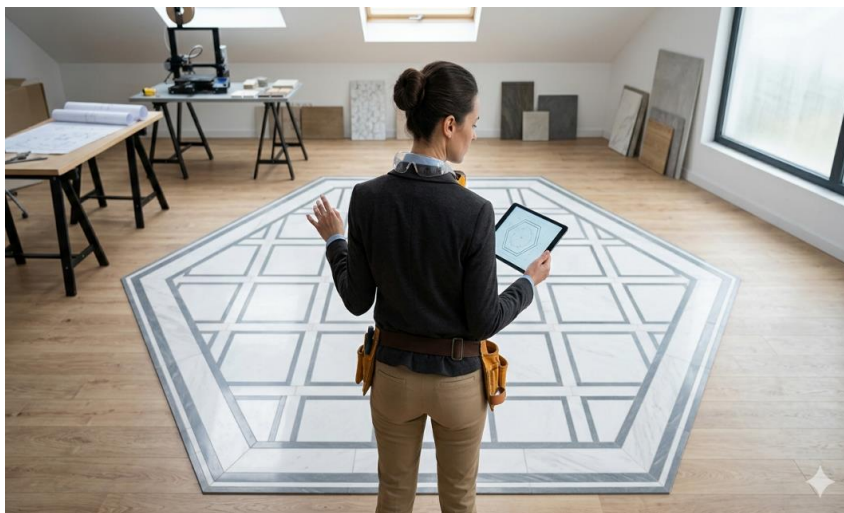
A fim de simular a intervenção do docente nessa pesquisa os autores fizeram as modificações necessárias para sanar os pontos alertados e assim tornar o item tecnicamente melhor e viável cumprindo as exigências do Quadro 1.

A seguir, apresenta-se o item já com as modificações aplicadas.

Descritor 08: Resolver problema utilizando propriedades dos polígonos (soma de seus ângulos internos, número de diagonais, cálculo da medida de cada ângulo interno nos polígonos regulares).

Enunciado: Uma arquiteta está projetando o design de uma nova praça de alimentação em um shopping. O centro dessa praça contará com um grande mosaico no chão em formato de um hexágono regular. Para garantir o corte preciso das pedras de mármore que compõem as bordas desse mosaico, a arquiteta precisa calcular a soma de todos os ângulos internos dessa figura.

Suporte:



Fonte: Elaborado pelos autores com auxílio da IA Gemini (2026)

Comando: Qual é a soma dos ângulos internos do mosaico hexagonal projetado pela arquiteta?

Alternativas:

- (A) 1080°
- (B) 720°
- (C) 540°
- (D) 360°

Solução e Justificativa:

Para resolver este item, o estudante deve aplicar a fórmula da soma dos ângulos internos de um polígono convexo:

$$S_i = (n-2) \times 180^\circ$$

Onde n é o número de lados do polígono. No caso de um hexágono, $n = 6$.

$$S_i = (6-2) \times 180^\circ$$

$$S_i = 4 \times 180^\circ$$

$$S_i = 720^\circ$$

Gabarito: Alternativa B

Análise dos Distratores (Justificativa do Erro):

- (A) **1080°:** (Erro de fórmula – esquecimento da subtração): O estudante utiliza a fórmula sem subtrair 2 do número de lados ($6 \times 180^\circ = 1080^\circ$). Este é um erro clássico de memorização da fórmula;

- (B) **720°**: GABARITO: Aplicação correta do conceito e da operação aritmética;
- (C) **540°**:(Erro de fórmula – identificação incorreta de polígono): O estudante utiliza a fórmula da forma correta porém confunde o hexágono que tem 6 lados com um pentágono que tem 5 desconsiderando o suporte $(5 - 2) \times 180^\circ = 540^\circ$. Este é um erro clássico de identificação de polígono;
- (D) **360°**:(Confusão entre fórmulas distintas): O estudante confunde a propriedade da “Soma dos Ângulos Internos” com a “Soma dos Ângulos Externos”, que para qualquer polígono convexo é 360° , olhando a questão e através de memorização marca diretamente sem reflexão confundindo o comando da questão.

4.2 AVALIAÇÃO DO DESCRITOR D18: OPERAÇÕES COM NÚMEROS INTEIROS

No segundo teste a preferência foi dada a um descritor de um eixo da matemática diferente do Descritor 08 que trabalha com propriedade de polígonos, o descritor escolhido foi o Descritor 18 tem sua delimitação no campo das 4 operações básicas e potenciação, sendo um descritor crucial no 9º ano pois trabalha questões de grande entrave, são elas as regras de sinais e precedência nas operações matemáticas.

Assim como no primeiro teste o *prompt* seguiu as interações necessárias junto a inteligência artificial do Gemini e realizou através de bloco único conforme descrito na metodologia a geração do item, e nela foi exigida além dos requisitos básicos, que fosse uma questão contextualizada e definiu restrições nos distratores a fim de minimizar possíveis alucinações, o resultado pode ser analisado a seguir:

Descritor 18: Efetuar cálculos com números inteiros, envolvendo as operações (adição, subtração, multiplicação, divisão, potenciação).

Enunciado: Lucas começou a semana com um saldo de –20 reais (negativo) em sua conta bancária. Para tentar regularizar sua situação, ele fez um depósito de 50 reais. No entanto, logo após o depósito, o banco cobrou 3 parcelas fixas de um serviço de manutenção, cada uma no valor de 15 reais.

Comando: Qual o saldo final da conta bancária de Lucas após essas movimentações?

Alternativas:

- (A) 75 reais
- (B) 15 reais
- (C) –15 reais
- (D) –45 reais

Solução e Justificativa:

Para chegar ao resultado correto, o estudante deve montar a expressão numérica correspondente à situação:

$$-20 + 50 + 3 \times (-15)$$

Prioridade das Operações: Primeiro, resolvemos a multiplicação (as 3 parcelas de dívida):

$$3 \times (-15) = -45$$

Soma e Subtração: Agora, seguimos com a ordem da esquerda para a direita:

$$-20 + 50 = +30$$

$$+30 - 45 = -15$$

O saldo final de Lucas é de -15 reais (ele continua devendo 15 reais).

Gabarito: Alternativa C

Análise dos Distratores (Justificativa do Erro):

- (A) **75 reais** (Erro de precedência das operações): O aluno resolve a adição/subtração antes da multiplicação. Ele faz: $-20 + 50 = 30$, depois multiplica $30 \times 15 = 450$ (ou soma os valores de forma desordenada). Outra possibilidade é ele somar os valores absolutos ignorando a natureza dos débitos: $30 + 45 = 75$.
- (B) **15 reais** (Erro na regra de sinais): O aluno realiza os cálculos corretamente, mas erra o sinal na operação final. Ele calcula $30 - 45$ e, por ver que o resultado numérico é 15, assume que o saldo é positivo, ignorando que o valor absoluto do débito (45) é maior que o do crédito (30).
- (C) **GABARITO:** O aluno respeita a prioridade da multiplicação e opera corretamente com os sinais de números inteiros.
- (D) **-45 reais** (Erro de interpretação/Dados parciais): O aluno identifica apenas o valor total das parcelas cobradas $3 \times 15 = 45$ e esquece de considerar o saldo anterior e o depósito realizado, ou se confunde ao tentar somar o saldo inicial negativo com o novo débito.

Fonte: Elaborado pelos autores com auxílio da IA Gemini (2026). Conversa original disponível em: <https://gemini.google.com/share/1611f53a1e9c>. Acesso em: 04 de maio 2026.

A fim de julgar a qualidade técnica do item gerado pela máquina é necessário analisar junto aos parâmetros do Quadro 2 se o mesmo atende aos critérios, tomando como base as especificações apontadas por Silva (2023, p. 51).

Quadro 2 – Validação Técnica do Item Gerado (Descritor D18).

Requisito Técnico (Inep/CAEd)	Status	Parecer Pedagógico
1. Alinhamento à Habilidade: O item avalia apenas a habilidade solicitada e não mede mera memorização?	Atende	O item exige modelagem adequada para sua resolução.
2. Clareza e Linguagem: Enunciado diversifica fontes (cotidiano), usa norma culta, sem jargões ou pegadinhas?	Atende	O uso da linguagem de valores monetários é adequado para problemas envolvendo valores negativos.
3. Formatação e Simetria: O comando é objetivo e as alternativas possuem a mesma extensão e padrão?	Atende	Comando e alternativas apresentam simetria e estão alinhados a unidade monetária (reais)
4. Plausibilidade dos Distratores: As alternativas incorretas refletem erros operacionais ou de raciocínios reais?	Atende	Os distratores cumprem seu papel na íntegra, simulando erros de procedência real.

Fonte: Elaborado pelos autores (2026).

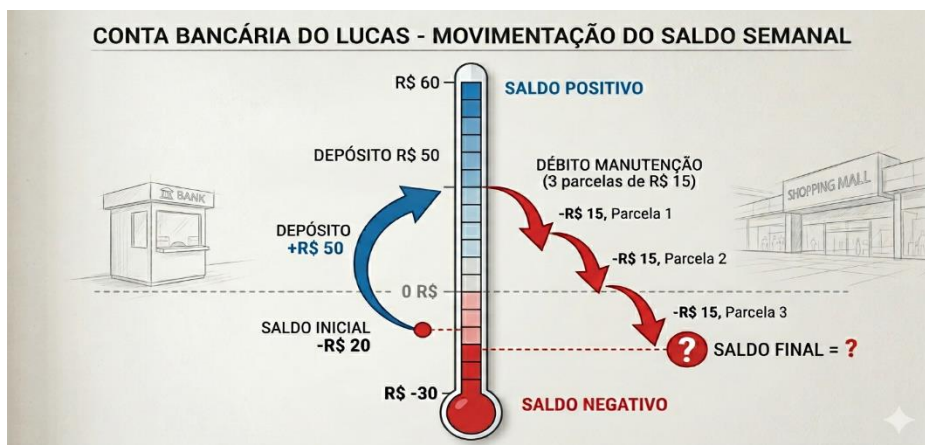
O item gerado para o descritor 18 pela inteligência artificial diferente do gerado para o descritor 08, atendeu todos os requisitos do quadro de avaliação, trazendo luz a sua viabilidade, porém ainda que viável o mesmo não deve ser limitado as capacidades do comando, da inteligência artificial ou mesmo do docente que está operando a máquina através de interações via *prompt*, todo item gerado ainda que viável pode ser melhorado, através de refinamentos via *prompt*, ou com base no conhecimento dos docentes alinhadas a especificidades exigidas em itens avaliação externa de larga escala, cada item após passar pelo refinamento do docente terá resultado diferente tendo em vista os diferentes níveis de conhecimento, as percepções e vivências de cada um, tornando assim itens diversos e viáveis mas com diferentes níveis de profundidade.

A seguir, apresenta-se o item já com os devidos refinamentos.

Descritor 18: Efetuar cálculos com números inteiros, envolvendo as operações (adição, subtração, multiplicação, divisão, potenciação).

Enunciado: Lucas começou a semana com um saldo de –20 reais (negativo) em sua conta bancária. Para tentar regularizar sua situação, ele fez um depósito de 50 reais. No entanto, logo após o depósito, o banco cobrou 3 parcelas fixas de um serviço de manutenção, cada uma no valor de 15 reais.

Suporte:



Fonte: Elaborado pelos autores com auxílio da IA Gemini (2026)

Comando: Qual o saldo final da conta bancária de Lucas após essas movimentações?

Alternativas:

- (A) –55 reais
- (B) 15 reais
- (C) – 15 reais
- (D) – 45 reais

Solução e Justificativa:

Para chegar ao resultado correto, o estudante deve montar a expressão numérica correspondente à situação:

$$-20 + 50 + 3 \times (-15)$$

Prioridade das Operações: Primeiro, resolvemos a multiplicação (as 3 parcelas de dívida):

$$3 \times (-15) = -45$$

Soma e Subtração: Agora, seguimos com a ordem da esquerda para a direita:

$$-20 + 50 = +30$$

$$+30 - 45 = -15$$

O saldo final de Lucas é de -15 reais (ele continua devendo 15 reais).

Gabarito: Alternativa C

Análise dos Distratores (Justificativa do Erro):

- (A) **-55 reais** (Erro de precedência das operações): O aluno resolve tentando assimilar as operações semelhantes antes de realizar a multiplicação $-20 - 15 = -35 \times 3 = -105$ e após isso realiza a soma $-105 + 50$ resultando em -55.
- (B) **15 reais** (Erro na regra de sinais): O aluno realiza os cálculos corretamente, mas erra o sinal na operação final. Ele calcula $+30 - 45$ e, por ver que o resultado numérico é 15, assume que o saldo é positivo, ignorando que o valor absoluto do débito (45) é maior que o do crédito (30).
- (C) **GABARITO:** O aluno respeita a prioridade da multiplicação e opera corretamente com os sinais de números inteiros.
- (D) **-45 reais** (Erro de interpretação/Dados parciais): O aluno identifica apenas o valor total das parcelas cobradas $3 \times 15 = 45$ e esquece de considerar o saldo anterior e o depósito realizado, ou se confunde ao tentar somar o saldo inicial negativo com o novo débito.

A fim de julgar a qualidade técnica do item gerado pela máquina é necessário analisar junto aos parâmetros do Quadro 2 se o mesmo atende aos critérios, tomando como base as especificações apontadas por Silva (2023, p. 51).

Quadro 3 – Validação Técnica do Item Refinado (Descritor D18).

Requisito Técnico (Inep/CAEd)	Status	Parecer Pedagógico
1. Alinhamento à Habilidade: O item avalia apenas a habilidade solicitada e não mede mera memorização?	Atende	O item exige modelagem adequada para sua resolução.
2. Clareza e Linguagem: Enunciado diversifica fontes (cotidiano), usa norma culta, sem jargões ou pegadinhas?	Atende	O uso da linguagem de valores monetários é adequado para problemas envolvendo valores negativos.
3. Formatação e Simetria: O comando é objetivo e as alternativas possuem a mesma extensão e padrão?	Atende	Comando e alternativas apresentam simetria e estão alinhados a unidade monetária (reais)
4. Plausibilidade dos Distratores: As alternativas incorretas refletem erros operacionais ou de raciocínios reais?	Atende	Os distratores cumprem seu papel na íntegra, simulando erros de procedência real.

Fonte: Elaborado pelos autores (2026).

5 CONSIDERAÇÕES FINAIS

A presente pesquisa teve como eixo central investigar a viabilidade do uso da Inteligência Artificial Generativa, especificamente o modelo Gemini, como ferramenta auxiliar na elaboração de itens de matemática com rigor técnico e pedagógico similar aos encontrados em avaliações de larga escala, alinhados à Matriz de Referência do SAEPE. Diante dos desafios enfrentados pelos docentes que dispõem de tempo escasso para uma produção que exige alto nível de detalhamento, buscou-se compreender se a máquina poderia atuar como uma parceira na otimização desse processo.

Os resultados da fase empírica, evidenciados nas análises e intervenções dos Descritores D08 e D18, demonstraram o potencial da máquina em reproduzir e simular a estrutura de um item viável, de acordo com os guias oficiais. Essa performance foi alcançada mediante a utilização de um *prompt* (comando) adequado e refinado pelo pesquisador. Diferentemente da abordagem de Silva (2023) que concedia maior liberdade criativa à IA, abrindo margem para “alucinações” e falhas, esta pesquisa optou por limitar a autonomia do modelo através de restrições pedagógicas. Ainda assim, constatou-se que a máquina não atinge a perfeição autônoma. Como defendido ao longo do estudo, é imprescindível a intervenção do docente para realizar melhorias substanciais, tais como a inserção de suportes visuais, a adequação de enunciados e comandos, e a reescrita de distratores, garantindo o padrão técnico exigido por um autêntico item do SAEPE.

Sendo assim, o papel do professor transmuta-se: ele deixa de ser o elaborador solitário de todo o texto bruto e passa a atuar como um validador técnico e curador do conhecimento. O domínio pedagógico do docente torna-se ainda mais essencial em duas frentes: primeiramente, na criação e no refinamento do *prompt*, pois quanto mais preciso for o comando, menor será o esforço na edição do resultado; e, em segundo lugar, na própria revisão e validação crítica da questão gerada. Esse polimento final humano é indispensável para que o item atinja a qualidade e a precisão esperadas de uma avaliação externa.

Conclui-se, portanto, que a Inteligência Artificial Generativa é uma ferramenta poderosa de auxílio na redação bruta, mostrando-se plenamente viável para o uso na educação matemática ao auxiliar na otimização do tempo de trabalho dos docentes. Contudo, seu sucesso institucional está condicionado à premissa de que toda geração passe, rigorosamente, pelo olhar analítico e pelo refinamento adequado do professor.

Por fim, como proposta para trabalhos futuros, sugere-se a aplicação prática dos itens gerados em uma turma de 9º ano do Ensino Fundamental, a fim de aferir suas qualidades psicométricas e atestar, de forma empírica, a eficácia de discriminação de cada item. Além disso, considera-se a expansão desta pesquisa para a testagem de outras inteligências artificiais generativas e novas modelagens de *prompts*, com o objetivo de otimizar continuamente a produção e reduzir ainda mais o tempo necessário para o refinamento final docente.

REFERÊNCIAS

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Guia de elaboração e revisão de itens**. Brasília: Inep, 2010. Disponível em: https://docs.ufpr.br/~aanjos/CE095/guia_elaboracao_revisao_itens_2012_INEP.pdf. Acesso em: 17 abr. 2026.

CAEd. Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora. **Guia de elaboração de itens:** matemática. Juiz de Fora: CAEd/UFJF, 2008. Disponível em: https://docs.ufpr.br/~aanjos/CE095/3_Guia_De_Elabora%C3%A7%C3%A3o_De_Itens_MT.pdf. Acesso em: 17 abr. 2026.

CARVALHO, Elvis Gonçalves. **A Inteligência Artificial no Ensino da Matemática:** possibilidades de aplicações nas turmas do Ensino Médio. 2025. Dissertação (Mestrado Profissional em Matemática em Rede Nacional) - Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, 2025. Disponível em: https://sca.profmt-sbm.org.br/busca_tcc_det.php?id=171059036. Acesso em: 12 mai. 2026.

FREITAS, Carlos Higor Sousa de. **Dos currículos oficiais à inteligência artificial:** o ensino de funções exponencial e logarítmica na BNCC e a geração de questões inéditas por meio de inteligência artificial generativa. 2026. Dissertação (Mestrado Profissional em Matemática em Rede Nacional) - Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Redenção, 2026. Disponível em: https://sca.profmt-sbm.org.br/busca_tcc_det.php?id=171060808. Acesso em: 13 mai. 2026.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa.** 7. ed. Barueri: Atlas, 2023.

HAYDT, Regina Célia Cazaux. **Avaliação do processo ensino-aprendizagem.** São Paulo: Ática, 1988.

HOFFMANN, Jussara Maria. **Avaliação:** mito e desafio: uma perspectiva construtivista. Porto Alegre: Mediação, 2005.

LIMA, Symon Igor Pinheiro da Silva. **Inteligência artificial no ensino de matemática:** potencialidades e desafios para professores. 2025. Dissertação (Mestrado Profissional em Matemática em Rede Nacional) - Universidade Federal de Alagoas, Maceió, 2025. Disponível em: https://sca.profmt-sbm.org.br/busca_tcc_det.php?id=171059966. Acesso em: 19 fev. 2026.

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

MCCARTHY, John. **What is Artificial Intelligence?** Stanford: Stanford University, 2007. Disponível em: <http://www-formal.stanford.edu/jmc/whatisai.pdf>. Acesso em: 19 abr. 2026.

PERNAMBUCO. Secretaria de Educação e Esportes. **SAEPE 2025:** Revista da Escola - Equipe Pedagógica Matemática. Juiz de Fora: CAEd/UFJF, 2025. Disponível em: https://prototipos.caeddigital.net/arquivos/pe/colecoes/2025/SAEPE_2025_RE_MT.pdf. Acesso em: 13 mai. 2026.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência artificial:** uma abordagem moderna. 4. ed. São Paulo: Pearson, 2022.

SILVA, Andrey Camurça da. **Desafios e possibilidades do uso de inteligência artificial generativa na elaboração e revisão de itens de matemática**. 2023. Dissertação (Mestrado Profissional em Matemática em Rede Nacional) - Universidade Federal do Oeste do Pará, Santarém, 2023. Disponível em: https://sca.proformat-sbm.org.br/busca_tcc_det.php?id=171058058. Acesso em: 13 fev. 2026.

SOUSA, Clarilza Prado de. Dimensões da avaliação educacional. **Estudos em Avaliação Educacional**, São Paulo, n. 22, p. 101-118, dez. 2000. Disponível em: <https://publicacoes.fcc.org.br/eae/article/view/2218>. Acesso em: 01 abr. 2026.