



**UNIVERSIDADE FEDERAL DO OESTE DO PARÁ  
INSTITUTO DE CIÊNCIAS DA EDUCAÇÃO  
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL**

**ANDREY CAMURÇA DA SILVA**

**DESAFIOS E POSSIBILIDADES DO USO DE INTELIGÊNCIA ARTIFICIAL  
GENERATIVA NA ELABORAÇÃO E REVISÃO DE ITENS DE MATEMÁTICA**

**SANTARÉM**

**2023**

ANDREY CAMURÇA DA SILVA

**DESAFIOS E POSSIBILIDADES DO USO DE INTELIGÊNCIA ARTIFICIAL  
GENERATIVA NA ELABORAÇÃO E REVISÃO DE ITENS DE MATEMÁTICA**

Dissertação apresentada ao Mestrado Profissional em Matemática em Rede Nacional para obtenção do título de Mestre em Matemática; Universidade Federal do Oeste do Pará. Orientador: Dr. Mário Tanaka Filho

SANTARÉM

2023

Ficha catalográfica elaborada pelo Sistema Integrado de Bibliotecas (SIBI) da UFOPA  
Catalogação de Publicação na  
Fonte. UFOPA - Biblioteca Unidade Rondon

Silva, Andrey Camurça.

Desafios e possibilidades do uso de inteligência artificial generativa na elaboração e revisão de itens de matemática / Andrey Camurça da Silva. - Santarém, 2023.

112f.: il.

Dissertação (Mestrado). Universidade Federal do Oeste do Pará-UFOPA. Instituto de Ciências da Educação -ICED. Mestrado Profissional em Matemática em Rede Nacional.

Orientador: Mario Tanaka Filho.

1. Inteligência artificial. 2. ChatGPT. 3. Elaboração de itens. I. Tanaka Filho, Mario. II. Título.

UFOPACampus Rondon

CDD 006.35



Universidade Federal do Oeste do Pará  
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL

ATA Nº 19

Aos vinte e dois dias do mês de dezembro do ano de dois mil e vinte e três, às quinze horas, de forma remota, reuniram-se os membros da Banca Examinadora composta pelos(as) professores(as) Drs(as). Prof. Dr. Mário Tanaka Filho (orientador e presidente), Prof. Dr. José Ricardo e Souza Mafra (membro externo), Prof. Dr. Edilan de Sant'Ana Quaresma (membro externo) e Prof. Dr. Claudir Oliveira (membro interno) a fim de argüirem o mestrando Andrey Carmuça da Silva, com a dissertação intitulada DESAFIOS E POSSIBILIDADES DO USO DA INTELIGÊNCIA ARTIFICIAL GENERATIVA NA ELABORAÇÃO E REVISÃO DE ITENS DE MATEMÁTICA. Aberta a sessão pelo presidente, coube ao candidato, na forma regimental, expor o tema de sua dissertação, dentro do tempo regulamentar, em seguida a banca fez as arguições, o candidato respondeu e, após as deliberações na sessão secreta foi:

Aprovado, fazendo jus ao título de Mestre em Matemática.

Reprovado.

Documento assinado digitalmente  
**gov.br** JOSE RICARDO E SOUZA MAFRA  
Data: 26/01/2024 09:53:23-0300  
Verifique em <https://validar.iti.gov.br>

**Dr. JOSE RICARDO E SOUZA MAFRA, UFOPA**

Examinador Externo ao Pro-----

Documento assinado digitalmente

**gov.br**

EDILAN DE SANT ANA QUARESMA

Data: 25/01/2024 22:03:13-0300

Verifique em <https://validar.iti.gov.br>

**Dr. EDILAN DE SANT ANA QUAR**

Examinador Externo ao Programa

Documento assinado digitalmente

**gov.br**

CLAUDIR OLIVEIRA

Data: 24/01/2024 11:23:56-0300

Verifique em <https://validar.iti.gov.br>

Examinador Interno

*Mário Tanaka Filho*  
**Dr. MARIO TANAKA FILHO, UFOPA**

Presidente

**ANDREY CAMURÇA DA SILVA**

Mestrando

Documento assinado digitalmente

**gov.br**

ANDREY CAMURCA DA SILVA

Data: 30/01/2024 11:11:52-0300

Verifique em <https://validar.iti.gov.br>



Dedico este trabalho aos meus pais, Eurides Camurça e Amarildo Lemos, cujo amor e incentivo foram cruciais nessa trajetória. Eles me ensinaram o valor do trabalho duro, da perseverança e, acima de tudo, da integridade.

## **AGRADECIMENTO**

Gostaria de expressar minha profunda gratidão aos meus estimados professores do PROFMAT. Em especial, gostaria de agradecer ao Dr. Sebastian Mancuso, Msc. Miguel Ângelo e Dr. Mario Tanaka, cujo apoio e ensinamentos foram fundamentais para minha jornada acadêmica. Sou imensamente grato à minha companheira, Geovana Tavares, e ao meu irmão, Angel Camurça, que me proporcionaram um lar, em Santarém (PA), sempre que necessário e ofereceram um apoio emocional inestimável. Agradeço à minha coordenadora, Professora Joana, e à minha diretora, Professora Andreza, pela organização dos meus horários e pela flexibilidade que me permitiu cursar o mestrado. Minha gratidão se estende a todos os meus familiares, com destaque para minha querida irmã Eurice Camurça, meus irmãos Amauri e Amarilso, e minha estimada prima Erlene Sousa. Seu constante incentivo e crença em minha capacidade de alcançar meus objetivos acadêmicos foram uma importante fonte de motivação.

## RESUMO

Recentemente, sistemas de Inteligência Artificial (IA) voltados para o processamento de linguagem natural geraram um clima de entusiasmo e preocupações, principalmente no contexto educacional. Chats baseados em Grandes Modelos de Linguagem (GML), como o ChatGPT, demonstraram capacidade de produzir textos em linguagem natural sobre uma ampla gama de assuntos e realizar tarefas diversas, como a elaboração de um item (questão) sobre um determinado assunto de matemática. Na prática, o usuário pode fornecer condições, parâmetros e instruções ao chat, obtendo respostas ou até novas perguntas sobre o tema discutido. Considerando o potencial dessas ferramentas, este estudo investigou a aplicação de ferramentas gratuitas que utilizam IA generativa na construção e revisão de itens de múltipla escolha. Metodologicamente, a pesquisa se caracteriza como aplicada, de natureza qualiquantitativa, e se constitui como bibliográfica, documental e experimental. Os chats utilizados foram o ChatGPT, baseado no Modelo GPT 3.5, e o Bing Chat. O estudo apresentou orientações e protocolos em linguagem natural que podem ser fornecidos aos chats para a elaboração ou revisão de itens de matemática relacionados a habilidades específicas. Alguns dos itens criados a partir desses protocolos e ferramentas compuseram um teste composto por 20 itens que contemplam habilidades preconizadas pela Matriz de Referência do Saeb. Este teste foi respondido por 61 estudantes do 9º ano do ensino fundamental de uma escola pública localizada no município de Terra Santa, Pará. Os resultados indicaram que os chats apresentaram problemas em todas as partes de um item: enunciado, suporte, comando e alternativas. No entanto, os chats também demonstraram capacidade de compreender as habilidades solicitadas e formular situações que as envolvam. Alguns itens produzidos com o auxílio dos chats e revisados pelo autor mostraram bom poder discriminativo, sugerindo que as ferramentas foram úteis para o processo de elaboração de itens de matemática, desde que aprimoradas pelo docente/elaborador.

**Palavras-chave:** Inteligência artificial. ChatGPT. Avaliação. Elaboração de itens. Ensino de matemática.

## ABSTRACT

Recently, Artificial Intelligence (AI) systems focused on natural language processing have generated a climate of motivation and concerns, especially in the educational context. Chats based on Large Language Models (GML), such as ChatGPT, have the capacity to produce texts in natural language on a wide range of subjects and diverse tasks, such as preparing an item (question) on a specific mathematics subject. In practice, the user can provide conditions, parameters and instructions to the chat, obtaining answers or even new questions on the proposed topic. Considering the potential of these tools, this study investigated the application of free tools that use generative AI in the construction and review of multiple choice items. Methodologically, the research is characterized as applied, of a qualitative and quantitative nature, and is constituted as bibliographic, documentary and experimental. The chats used were ChatGPT, based on the GPT 3.5 Model, and Bing Chat. The study presented natural language guidelines and protocols that can be provided to chats for creating or reviewing math items related to specific skills. Some of the items created from these protocols and tools with a test composed of 20 items that include skills recommended by the Saeb Reference Matrix. This test was answered by 61 students in the 9th year of elementary school at a public school located in the municipality of Terra Santa, Pará. The results indicated that the chats presented problems in all parts of an item: statement, support, command and alternatives. However, chats also demonstrated the ability to understand the skills requested and formulate situations involving them. Some items produced with the help of chats and reviewed by the author showed good discriminatory power, indicating that the tools were useful for the process of preparing mathematics items, as long as they were improved by the teacher/developer.

**Keywords:** Artificial intelligence. ChatGPT. Assessment. Item creation. Mathematics teaching.



## LISTA DE ILUSTRAÇÃO

Figura 1 - Linha do tempo com alguns marcos do desenvolvimento da IA.....	19
Figura 2 - Algumas definições de inteligência artificial, organizadas em quatro categorias ...	20
Quadro 1 – Taxonomia de IA.....	22
Figura 3 – Comparação do Aprendizado de Máquina com programação clássica.....	23
Figura 4 - Neurônio biológico e funcionamento de um neurônio artificial.....	24
Figura 5 - Arquitetura de uma Rede Neural Artificial (RNA) genérica.....	25
Figura 6 - Esquema de uma Rede Neural Convolutacional .....	26
Figura 7 - Linha do tempo com as principais mudanças no Saeb: Brasil 1990-2021 .....	42
Figura 8 – Ilustração de uma escala de proficiência.....	44
Figura 9 - Exemplo de item de matemática com suporte gráfico.....	48
Figura 10 - Exemplo de item de matemática sem suporte gráfico .....	48
Figura 11 - Exemplo de item de matemática com texto base: modelo adotado no ENEM.....	49
Figura 12 - Exemplo de item com problemas de assimetria das alternativas.....	50
Figura 13 - Análise Gráfica de Item: elementos.....	53
Figura 14 – Processo e agentes envolvidos na elaboração de itens de múltipla escolha buscando atender alguma habilidade prevista na Matriz de Referência do Saeb 2018 .....	57
Figura 15 – Fluxograma: reconhecimento da Matriz de Referência do Saeb para Elaboração de Itens .....	58
Figura 16 – Instruções resumidas que a IA deve levar em conta na elaboração dos itens.....	59
Quadro 2 – Instruções dadas ao ChatGPT e ao Bing Chat para produção de itens de matemática que contemple uma habilidade específica da Matriz de Referência do Saeb .....	60
Figura 17 – Processo e agentes envolvidos na elaboração de itens de múltipla escolha que atendam alguma habilidade da BNCC .....	61
Quadro 3 – Instruções ao Bing Chat para produção de itens de matemática que contemple uma habilidade da BNCC com base no código alfanumérico .....	62
Figura 18 – Processo e agentes envolvidos na elaboração de itens de múltipla escolha a partir de contexto e habilidade fornecida pelo professor .....	63
Quadro 4 – Instruções dadas ao Bing Chat para produção de item com base em um texto base .....	63
Figura 19 – Processo e agentes envolvidos na elaboração e revisão itens de múltipla escolha: IA como revisora .....	64

Quadro 5 – Instruções dadas aos chats para o aprimoramento de itens com alguns critérios de revisão.....	65
Figura 20 – Item produzido para a habilidade 9A1.1: falhas nas justificativas dos distratores	68
Figura 21 – Item produzido para avaliar a habilidade 9A1.8 .....	69
Figura 22 – Item produzido para avaliar a habilidade <i>EF09MA05</i> .....	71
Figura 23 – Item produzido para avaliar a habilidade <i>EF09MA09</i> .....	73
Figura 24 – Item produzido para avaliar a habilidade <i>EF09MA07</i> .....	74
Figura 25 – Item revisado e aprimorado pelo ChatGPT: habilidade <i>EF09MA06</i> .....	75
Figura 26 – Item revisado e aprimorado pelo ChatGPT: habilidade <i>EF09MA04</i> .....	76
Figura 27 – Conversação no Bing Chat para a criação de um item que atende à habilidade 9A1.7 da MR2018 de matemática.....	78
Figura 28 – Conversação no Bing Chat para a criação de um item que atende à habilidade 9A1.6 da MR2018 de matemática.....	79
Figura 29 – Conversação no Bing Chat para a criação de um item que atende à habilidade 9A2.4 da MR2018 de matemática.....	80
Figura 30 – Conversação no Bing Chat para a criação de um item que atende à habilidade <i>EF09MA04</i> da BNCC .....	81
Figura 31 – Conversação no Bing Chat para a criação de um item que atende à habilidade <i>EF09MA14</i> da BNCC .....	82
Figura 32 – Conversação no Bing Chat para a criação de um item que atende à habilidade <i>EF09MA06</i> da BNCC .....	83
Figura 33 – Conversação no Bing Chat para a criação de um novo item que atende à habilidade <i>EF09MA06</i> da BNCC .....	84
Figura 34 – Conversação no Bing Chat para revisão de um item .....	85
Figura 34 – Conversação no Bing Chat para revisão de um item do ENEM.....	86
Figura 36 – Item 4 e sua análise gráfica .....	89
Figura 37 – Item 13 e sua análise gráfica .....	90
Figura 38 – Item 3 e sua análise gráfica .....	91
Figura 39 – Item 11 e sua análise gráfica .....	92

## LISTA DE TABELAS

Tabela 1 - classificação dos itens de acordo com a discriminação na TCT .....	55
Tabela 2 – Parâmetros clássicos dos itens que compõe o teste: proporção de acerto (dificuldade), discriminação (clássica e correlação ponto bisserial) e consistência interna (alfa de Cronbach) .....	88

## LISTA SIGLAS

AGI	Análise Gráfica de Itens
AM	Aprendizado de Máquina
CAED	Centro de Políticas Públicas e Avaliação da Educação
CNC	Redes Neurais Convolucionais
Enem	Exame Nacional do Ensino Médio
GML	Grandes Modelos de Linguagem
Inep	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MEC	Ministério da Educação
PNL	Processamento de Linguagem Natural
RNR	Redes Neurais Recorrentes
Saeb	Sistema de Avaliação da Educação Básica
TCT	Teoria Clássica de Testes
TRI	Teoria de Resposta ao Item
UFJF	Universidade Federal de Juiz de Fora
UFOPA	Universidade Federal do Oeste do Pará

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>14</b>
<b>2 INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO .....</b>	<b>16</b>
2.1 Um breve histórico da IA .....	16
2.2 Conceitos e definição de IA .....	20
2.3 IA na Educação: desafios e possibilidades .....	28
2.4 Ferramentas de IA para a avaliação da aprendizagem .....	32
<b>3 AVALIAÇÃO EDUCACIONAL NA EDUCAÇÃO BÁSICA .....</b>	<b>35</b>
3.1 Dimensões da avaliação educacional.....	35
3.2 Definições e conceitos utilizados na avaliação em larga escala .....	41
3.2.1 Matriz de referência.....	43
3.2.2 Competências e Habilidades.....	43
3.2.3 Escala de Proficiência.....	44
3.2.4 Item.....	45
3.3 Elaboração de itens de matemática para avaliação da aprendizagem .....	45
<b>4 PROCEDIMENTOS METODOLÓGICOS.....</b>	<b>56</b>
4.1 Classificação da pesquisa .....	56
4.2 Ferramentas utilizadas.....	56
4.3 Sugestões de elaboração de itens de matemática por meio de IA generativa .....	57
4.3.1 Item múltipla escolha com base na Matriz de Referência do Saeb .....	57
4.3.2 Item múltipla escolha com base na BNCC.....	61
4.3.3 Item múltipla escolha a partir de um contexto fornecido .....	63
4.3.4 Aperfeiçoamento de item de múltipla escolha .....	64
4.4 Aplicação de um teste para avaliação da aprendizagem em matemática .....	66
<b>5 CONSTRUÇÃO DE ITENS POR MEIO DE IA GENERATIVA.....</b>	<b>67</b>
5.1 Elaboração de itens no ChatGPT.....	67
5.1.1 Itens construído com base em uma habilidade fornecida: ChatGPT .....	67
5.1.2 Itens construído com base em uma habilidade e um contexto fornecido: ChatGPT .....	72
5.1.3 Aperfeiçoamento e revisão de item com o ChatGPT .....	74
5.2 Elaboração de itens no Bing Chat.....	77
5.2.1 Itens construído com base em uma habilidade fornecida: Bing Chat.....	77
5.2.2 Itens construído com base em uma habilidade e um contexto fornecido: Bing Chat .....	83
5.2.3 Aperfeiçoamento e revisão de item com o Bing Chat .....	84
5.3 Qualidade psicométrica de alguns dos itens produzidos.....	87
<b>6 CONSIDERAÇÕES FINAIS.....</b>	<b>94</b>

<b>REFERÊNCIAS .....</b>	<b>97</b>
<b>APÊNDICES .....</b>	<b>100</b>

## 1 INTRODUÇÃO

A avaliação da aprendizagem é fundamental para o processo educativo, uma vez que permite verificar o nível de domínio dos estudantes sobre os conteúdos e as habilidades trabalhadas em sala de aula, bem como identificar as dificuldades e os avanços de cada educando. Os processos avaliativos também subsidiam o planejamento pedagógico e a tomada de decisões sobre as intervenções necessárias para a melhoria da qualidade do ensino, tanto na sala de aula quanto em um sistema de ensino como um todo (LUCKESI, 2013, 2022; SOUSA, 2000).

Nesse sentido, a elaboração de itens de avaliação é uma tarefa que exige competência e critério por parte do docente, pois envolve a seleção dos conteúdos e das habilidades mais relevantes, a formulação de situações-problema que estimulem o raciocínio e a criatividade dos educandos, a definição dos parâmetros de correção e de feedback, entre outros aspectos (VIANNA, 1978).

No caso da matemática, a elaboração de itens de avaliação é ainda mais desafiadora à medida que requer o domínio de conceitos abstratos e complexos, a articulação entre diferentes áreas do conhecimento, a contextualização de problemas reais e a adequação ao nível de desenvolvimento cognitivo dos estudantes.

Uma forma de orientar e padronizar a elaboração de itens de matemática para avaliações de larga escala é recorrer às matrizes de referência de avaliação externa, como a do Sistema de Avaliação da Educação Básica (Saeb), que definem as habilidades esperadas para cada ano/série e as categorias de conhecimento que as compõem. Essas matrizes são baseadas em documentos curriculares nacionais, como a Base Nacional Comum Curricular (BNCC) e o Parâmetros Curriculares Nacionais (PCNs) (BRASIL, 2010, 2023b).

No entanto, nem sempre os docentes dispõem de tempo e de conhecimento suficientes para elaborar itens de matemática que contemplem as habilidades preconizadas pelas matrizes de referência, o que compromete a qualidade da avaliação e, conseqüentemente, da aprendizagem. Além disso, muitas vezes os docentes se baseiam em modelos de itens já existentes, que podem estar desatualizados, descontextualizados ou inadequados para o público-alvo.

Diante desse cenário, surge a possibilidade avaliar o uso de ferramentas tecnológicas que empregam a inteligência artificial (IA) para auxiliar o docente na elaboração e na revisão de itens de matemática. A IA é um ramo da ciência da computação que busca criar sistemas

capazes de simular o comportamento inteligente humano, como o aprendizado, o raciocínio, a percepção, a comunicação, entre outros (DELIPETREV; TSINARAKII; KOSTIĆ, 2020).

Uma das aplicações da IA é o processamento de linguagem natural (PLN), que consiste em analisar, compreender e gerar textos em linguagem natural, como o português. Um dos avanços recentes nessa área foi o desenvolvimento de grandes modelos de linguagem (GML), que são redes neurais artificiais treinadas com expressiva quantidade de dados textuais, capazes de produzir textos coerentes e consistentes sobre diversos assuntos, a partir de um estímulo inicial (SEJNOWSKI, 2023; YU, 2023a).

Um exemplo de GML é o GPT 3.5 desenvolvido pela OpenAI. Uma das formas de acessar a esse modelo de linguagem é por meio de chats inteligentes, que são interfaces conversacionais que permitem ao usuário interagir com o modelo de linguagem por meio de texto. Um desses chats é o ChatGPT, que é uma plataforma gratuita, que permite ao usuário informar fazer perguntas sobre diversos temas. Outra ferramenta interessante com essa funcionalidade é o Bing Chat, mas que também funciona como uma ferramenta de busca.

Considerando o potencial dessas ferramentas, este estudo teve como objetivo investigar a aplicação de ferramentas gratuitas que utilizam IA generativa na construção e revisão de itens de matemática de múltipla escolha. Para isso, foram utilizados dois chats inteligentes: o ChatGPT, baseado no GPT-3.5, e o Bing Chat, que faz uso do modelo GPT 4.

O estudo apresenta orientações e protocolos em linguagem natural que podem ser fornecidos aos chats para a elaboração ou revisão de itens de matemática relacionados a habilidades específicas. Alguns dos itens criados a partir desses protocolos e ferramentas compuseram um teste composto por 20 itens que contemplam habilidades preconizadas pela Matriz de Referência do Saeb. Este teste foi respondido por 61 estudantes do 9º ano do ensino fundamental de uma escola pública localizada no município de Terra Santa, Pará.

Os resultados indicaram que os chats apresentaram problemas em todas as partes de um item: enunciado, suporte, comando e alternativas. No entanto, os chats também demonstraram capacidade de compreender as habilidades solicitadas e formular situações que as envolvam. Alguns itens produzidos com o auxílio dos chats e revisados pelo autor mostraram bom poder discriminativo, sugerindo que as ferramentas foram úteis para o processo de elaboração de itens de matemática, desde que aprimoradas pelo docente/elaborador.

Em síntese, o estudo apresenta um panorama sobre inteligência artificial, tais como histórico, conceitos, modelos mais basilares e suas interfaces com a educação (Seção 2). Discute duas dimensões da avaliação educacional, seus conceitos e sugestões para elaboração de itens (Seção 3). Propõe potenciais protocolos e recomendações que podem ser fornecidos aos



sistemas de chat inteligentes para a construção de itens que cumpram os principais requisitos estabelecidos nos manuais de elaboração de itens (Seção 4). Discute a elaboração, a revisão e a análise de itens criados com auxílio de duas ferramentas que empregam inteligência artificial (Seção 5).

Portanto, o presente estudo coaduna com o avanço do conhecimento sobre o uso da IA generativa na educação matemática, bem como para a reflexão sobre os desafios e as possibilidades dessa tecnologia para a avaliação da aprendizagem. Ao mesmo tempo, oferece subsídios para o desenvolvimento de estratégias que possam facilitar e aperfeiçoar o trabalho do docente na elaboração e na revisão de itens de matemática.

## **2 INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO**

O desenvolvimento de Inteligência Artificial (IA) é parte da aspiração humana de criar sistemas que possam realizar tarefas cognitivas habitualmente feitas por humanos. Embora seu desenvolvimento tenha passado por alguns vales, no início da terceira década do terceiro milênio se dispõe de IA capazes de responder a demandas humanas complexas em linguagem natural, constituindo um tema de discussão em todos os setores da sociedade. Nesta seção, apresenta-se um histórico do desenvolvimento de IA, seus conceitos e modelos mais recentes, bem como possíveis impactos no campo da educação.

### **2.1 Um breve histórico da IA**

Embora não se saiba com exatidão quando surgiram os protótipos de Inteligência Artificial (IA), Haenlein e Kaplan (2019) atribuem sua origem à década de 1940. No campo da ficção científica, o conto *Runaround* de Isaac Asimov apresentava robôs dotados de inteligências inspiradas na inteligência humana e que operavam dentro de certas regras.

Na mesma época, mas no mundo real, o matemático inglês Alan Turing trabalha no desenvolvimento de uma máquina que fosse capaz de decifrar o código Enigma, usado pelo exército alemão durante a 2ª Guerra Mundial. A *The Bombe*, como foi chamada, foi capaz de executar uma tarefa que na época era impossível até para os melhores matemáticos. A máquina criada é considerada o primeiro protótipo do computador eletromecânico em funcionamento. O êxito da *The Bombe* fez Turing se questionar sobre a inteligência de tais máquinas, o que o levou em 1950 à publicação do artigo intitulado “*Computing Machinery and Intelligence*” (Máquinas de Computação e Inteligência, tradução livre), no qual descreve máquinas

inteligentes e propõe um teste para avaliar se uma máquina é ou não inteligente – o Teste de Turing. Outro marco importante ainda na década de 1940 foi o desenvolvimento de um modelo de neurônio artificial, basilar para as redes neurais artificiais proposto por Warren McCulloch e Walter Pitts (HAENLEIN; KAPLAN, 2019).

Segundo Sichman (2021), a IA adentra no espaço acadêmico na década de 1950 - momento em que o desenvolvimento do computador eletrônico e da computação ganhavam notoriedade. Nesse período, mais precisamente em 1956, Marvin Minsky e John McCarthy, cientistas da computação, organizaram o primeiro evento acadêmico sobre IA. O evento “Dartmouth Summer Research Project on Artificial Intelligence (DSRPAI)” ocorreu em Dartmouth College em Nova Hampshire e foi onde o termo “Inteligência Artificial” foi definitivamente adotado pela comunidade científica (BUCHANAN, 2005).

A conferência de Dartmouth despertou ânimo para as pesquisas em IA. Foram duas décadas de avanços, marcados pela criação da linguagem de programação LISP de John McCarthy e do computador ELIZA, desenvolvido por Joseph Weizenbaum entre 1964 e 1966. O ELIZA foi uma ferramenta de processamento de linguagem natural com capacidade de simular conversas com um humano (HAENLEIN; KAPLAN, 2019). Durante esse período, as abordagens simbólicas dominaram, com foco no processamento de linguagem natural e na resolução de problemas (BUCHANAN, 2005).

Entre as décadas de 1960 e 1970, os pesquisadores de IA se viram diante do primeiro “Inverno da IA” – termo usado para referir à má fase no desenvolvimento de IA, tanto por falta de investimentos quanto pelos entraves técnicos como o baixo poder computacional, processamento e memória disponível na época (DELIPETREV; TSINARAKII; KOSTIĆ, 2020).

Segundo Haenlein e Kaplan (2019), uma das razões para primeiro Inverno da IA foi a abordagem específica adotada pelos primeiros sistemas de IA, como o ELIZA e o *General Problem Solver*. Esses sistemas eram constituídos por coleções de regras que assumiam que a inteligência humana poderia ser formalizada e reconstruída como uma série proposições lógicas.

Entre as décadas de 1970 e 1990 surgiram Sistemas Especialistas, que eram programas de computador baseados em regras para resolver problemas específicos. Nesse ínterim, houve alguns avanços em redes neurais artificiais e técnicas de aprendizado de máquina, como o aprendizado por reforço e o aprendizado supervisionado. Entre os marcos, podem ser citados: (1972) a criação da linguagem de programação Prolog, que permitia o desenvolvimento de diversas ferramentas de IA; (1982) o processamento de informações via “rede de Hopfield” e o

ressurgimento do campo da IA do conexionismo; (1983) criação do algoritmo ID3 – precursor dos algoritmos usados no aprendizado de máquina e processamento de linguagem natural (DELIPETREV; TSINARAKII; KOSTIĆ, 2020).

Nos anos 1990, os Sistemas Especialistas criados visavam “obter conhecimento especializado humano em forma de computador e divulgá-lo como um programa para computadores pessoais (PCs)” (DELIPETREV; TSINARAKII; KOSTIĆ, 2020, p. 9 tradução nossa), isto é, eram produtos criados para inferências e que permitiam resolver problemas como diagnósticos de doenças, prescrição de medicamentos, entre outros.

Assim, tais sistemas davam conta de desafios cada vez maiores, mas reduzidas a tarefas específicas para qual era programada, como jogar uma partida de xadrez, onde as posições das peças podem ser mapeadas por um sistema cartesiano. Os Sistemas Especialistas alcançavam bons resultados nas áreas que se prestavam, mas não eram capazes de lidar a complexidade do mundo real (HAENLEIN; KAPLAN, 2019).

O paradigma dos Sistemas Especialistas enfrentou limitações, que começavam com o processo de aquisição de conhecimento. Segundo Delipetrev, Tsinarakli e Kostić (2020, p.10) “(...) a pesquisa de sistemas especialistas concentrou-se em ferramentas para aquisição de conhecimento, para ajudar a automatizar o processo de design, depuração e manutenção de regras definidas pelos especialistas”. Portanto, os extensos conjuntos de regras para produção de inferências exigiam trabalho humano especializado e custosos – pouco disponíveis. Ao mesmo tempo, a linguagem compilada (como o C) vinha ganhando espaço na comunidade acadêmica em função do desenvolvimento de computadores pessoais cada vez menores e acessíveis, que foram substituindo gradativamente as caras máquinas Lisp.

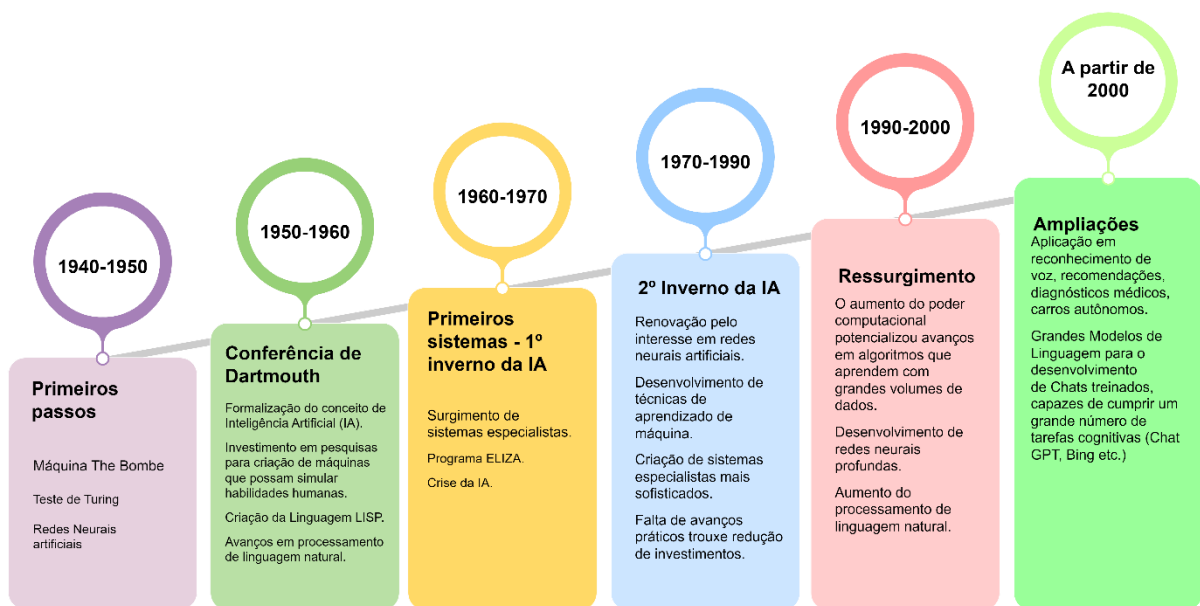
As dificuldades e limitações da IA nesse período levaram ao segundo Inverno da IA (entre 1970 e 1990), marcado pela falta de investimento e pelo ofuscamento do termo “IA” nos projetos subsequentes. Outros termos como “sistemas inteligentes”, “inteligência computacional”, “sistemas cognitivos” e “raciocínio” passaram a fazer parte do vocabulário dos pesquisadores para evitar a fuga de investimentos (DELIPETREV; TSINARAKII; KOSTIĆ, 2020).

Entre 1990 e 2010, a IA alcançou diversas áreas com soluções úteis, tais como mineração de dados, robótica industrial, mecanismos de pesquisa, softwares bancários e sistemas de recomendações. Seu ressurgimento foi impulsionado pelo aumento da capacidade computacional, disponibilidade de grandes conjuntos de dados e avanços em algoritmos de Aprendizado de Máquina. No âmbito acadêmico, esse desenvolvimento pode ser explicado pelo fortalecimento e rigor com que a IA passou a ser abordada enquanto disciplina científica. Sua

formalização por uma linguagem matemática compartilhada levou diversos grupos em várias partes do mundo a colaborarem com o avanço da IA, especialmente para a criação de modelos que aprendem com grandes bases de dados (DELIPETREV; TSINARAKII; KOSTIĆ, 2020).

Na década seguinte a IA não parou de crescer, especialmente no seu subcampo de Aprendizagem de Máquina (Machine Learning) e Aprendizado Profundo (Deep Learning). Segundo Haenlein e Kaplan (2019), um feito que marca o êxito do Aprendizado de Máquina ocorreu em 2016, quando o programa AlphaGo, desenvolvido pelo Google, foi capaz de derrotar um campeão mundial do jogo de tabuleiro Go – por muito tempo considerado uma tarefa impossível para uma IA. Os avanços nesses campos da IA produziu sistemas cada vez mais robustos de recomendações, reconhecimento de voz, reconhecimento de imagens, diagnóstico médico assistido por computador, processamento de linguagem natural, tradução automática, entre outros. Uma linha do tempo com alguns marcos do desenvolvimento da IA é apresentada na Figura 1.

Figura 1 - Linha do tempo com alguns marcos do desenvolvimento da IA



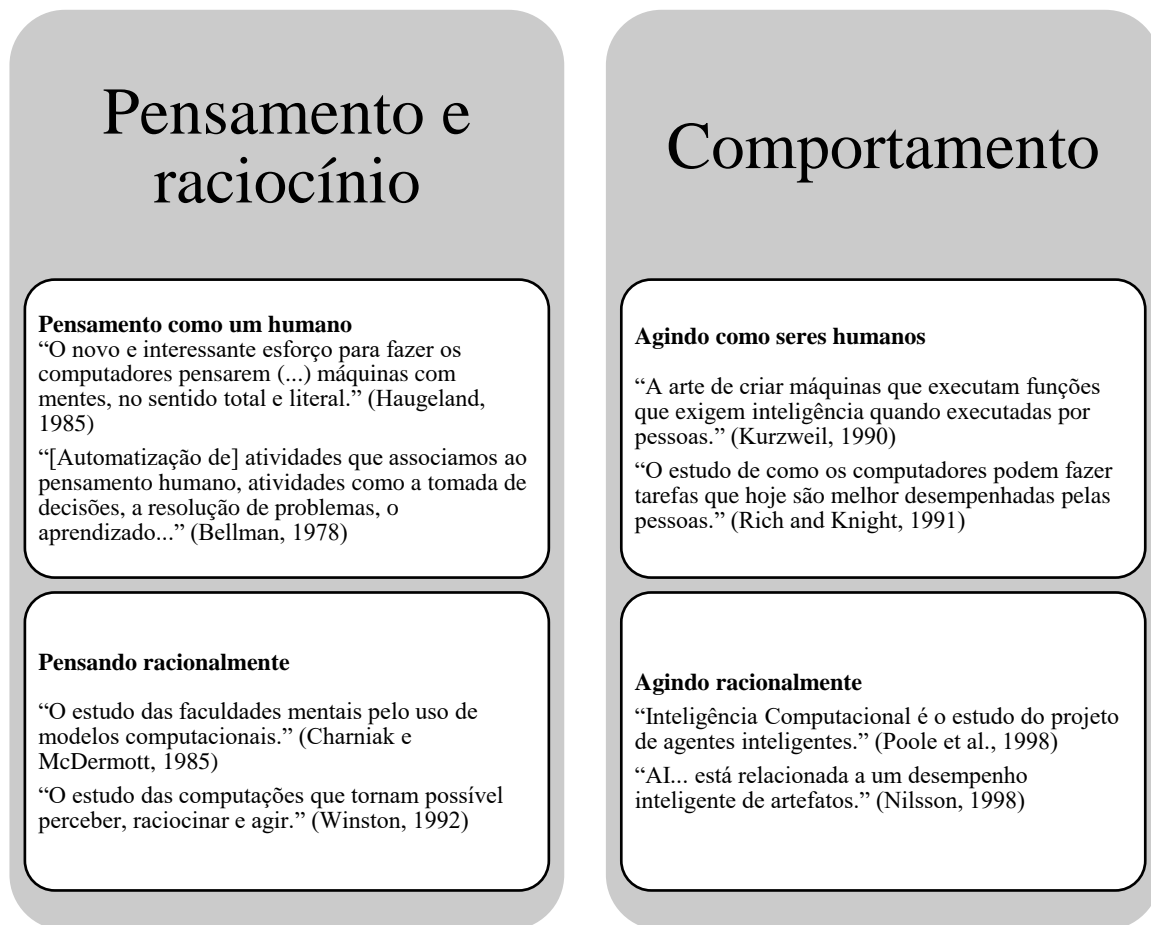
Fonte: Elaborado pelo próprio Autor.

Pode-se dizer que no período mais recente os sistemas baseados em IA ganharam protagonismo pela vasta aplicação que apresentam em praticamente todos os setores da sociedade. Por essa razão, a partir dos anos 2000 identifica-se um movimento de ampliação da atuação da IA. Ferramentas de reconhecimento facial e de voz, chats inteligentes e algoritmos de recomendações com base no interesse do indivíduo estão presentes no dia a dia das pessoas, movimentando um mercado em expansão.

## 2.2 Conceitos e definição de IA

Segundo Russell e Norvig (2013), IA é definida em duas dimensões, a saber, a do “pensamento e raciocínio” e do “comportamento”. Em cada uma dessas dimensões, há duas abordagens: 1) centrada nos seres humanos, como parte de uma ciência empírica; 2) racionalista, envolvendo uma combinação de matemática e engenharia. A Figura 2 exibe algumas definições dentro dessas quatro categorias.

Figura 2 - Algumas definições de inteligência artificial, organizadas em quatro categorias



Fonte: Adaptado de Russell e Norvig (2013).

Na abordagem empirista e, portanto, baseada no pensamento e comportamento humano, percebe-se que a IA é definida como sistemas ou máquinas capazes de pensar como seres humanos ou agir como seres humanos. O parâmetro de inteligência a ser alcançada nessa perspectiva é a inteligência humana. Em particular, Harasim (2015, p. 31) define IA como “uma área da Ciência da Computação que busca fazer os computadores pensarem e se comportarem como seres humanos”. Para Coppin (2010, p. 4) IA “envolve utilizar métodos baseados no

comportamento inteligente de humanos e outros animais para solucionar problemas complexos”.

Em contrapartida, a abordagem racionalista entende IA como modelos computacionais baseados em faculdades mentais que podem perceber, analisar e agir de maneira racional, frente uma tarefa dada (RUSSELL; NORVIG, 2013).

Definições mais recentes como a apresentada em Haenlein e Kaplan (2019, p. 2), segundo a qual IA é “a capacidade de um sistema de interpretar corretamente dados externos, aprender com esses dados e usar essas aprendizagens para alcançar objetivos e tarefas específicas por meio de adaptação flexível”, retomam a abordagem racionalista. Nessa perspectiva, “IA é um termo genérico que se refere a qualquer máquina ou algoritmo capaz de observar seu ambiente, aprender e, com base no conhecimento e na experiência adquiridos, tomar ações inteligentes ou propor decisões.” (SAMOILI et al., 2021, p. 20, tradução nossa). As duas últimas definições evocam paradigmas recentes da IA, como o Aprendizado de Máquina.

Enquanto área do conhecimento, a IA alcança diversos domínios do conhecimento. Samoili et al. (2021) propuseram uma taxonomia da IA, segundo a qual é dividida em domínio e subdomínio. A disposição dos domínios e seus respectivos subdomínios não pressupõe que eles sejam isolados, pelo contrário, possuem relações importantes. Outro ponto é que tal classificação não restringe os subdomínios apenas ao campo da IA. Por exemplo, no subdomínio da Robótica e Automação a IA exerce um papel particular e que dependem de outros campos, como engenharia mecatrônica, ciência da computação e outros.

Quadro 1 – Taxonomia de IA

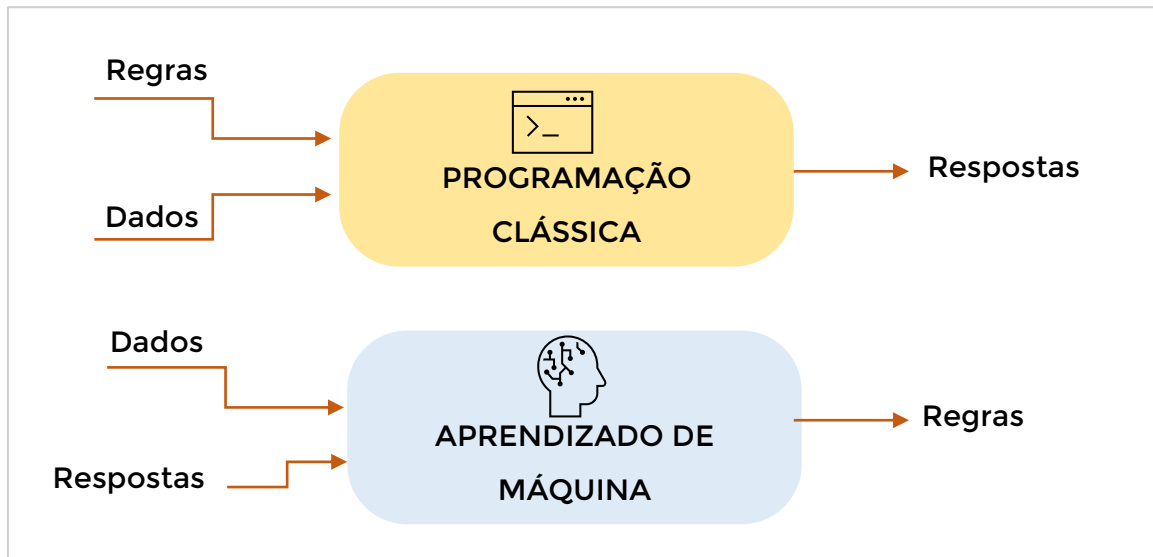
	<b>Domínio de IA</b>	<b>Subdomínio de IA</b>
<b>Essencial</b>	Raciocínio	Representação do conhecimento
		Raciocínio automatizado
		Raciocínio de senso comum
	Planejamento	Planejamento e Programação
		Busca
		Otimização
	Aprendizado	Aprendizado de máquina
	Comunicação	Processamento de linguagem natural
	Percepção	Visão computacional
		Processamento de áudio
<b>Transversal</b>	Integração e Interação	Sistemas multiagentes
		Robótica e Automação
		Veículos conectados e automatizados
	Serviços	Serviços de IA
	Ética e Filosofia	Ética da IA
		Filosofia da IA

Fonte: Samoili et al (2021).

Este trabalho apresenta de forma breve conceitos relativos ao domínio do Aprendizado e da Comunicação, ambos atrelados ao desenvolvimento de *chats* inteligentes capazes de realizar tarefas acadêmicas, como tradução, leitura e síntese de textos, em linguagem natural.

O AM pressupõe que as máquinas “aprendem” a tomar uma decisão baseada em instruções (dados de treinamentos), contrapondo-se aos sistemas que recebem uma programação específica para realizar uma determinada tarefa (SAMOILI et al., 2021). A Figura 3 exibe as principais diferenças do paradigma do AM em relação à programação clássica.

Figura 3 – Comparação do Aprendizado de Máquina com programação clássica



Fonte: Adaptado de Delipetrev, Tsinarakii e Kostić (2020).

Nesse sentido, a tarefa de um sistema segundo esse paradigma é aprender a categorizar dados de entrada de acordo com um conjunto de classificações supervisionadas ou não supervisionadas por humanos. “O sistema, então, tenta aprender, a partir desses dados de treinamento, a como classificar estes mesmos dados (geralmente, uma tarefa relativamente fácil) e como classificar novos dados ainda não observados” (COPPIN, 2010, p. 233).

Uma definição formal bastante abrangente do AM é:

Por aprendizado, nos referimos à capacidade dos sistemas de aprender, decidir, prever, adaptar e reagir automaticamente às mudanças, melhorando a partir da experiência, sem serem explicitamente programados. O AM está amplamente incluído na grande maioria dos esforços para identificar as categorias de IA, como a abordagem algorítmica básica para alcançar a IA, independentemente do tipo de aprendizado, ou seja, reforço, supervisionado, semi-supervisionado, não supervisionado (SAMOILI et al., 2021, p. 24 tradução nossa).

A definição acima elucidada que sistemas baseados em IA são capazes de aperfeiçoarem, seja por reforço, supervisão ou sem supervisão. Enquanto domínio da IA, o AM adquiriu subcampos e técnicas específicas. Subcampos clássicos que lidavam com Árvore de Decisão, Aprendizado por Hábito, Aprendizado por Conceito, K-vizinhos Mais Próximos, entre outros, deram lugar às modernas técnicas baseadas em Redes Neurais Artificiais (RNA), dentre as quais cabe mencionar o campo do Aprendizado Profundo (Deep Learning) e do Aprendizado por Reforço (Reinforcement Learning). Algumas das RNAs utilizadas são: Redes Neurais Convolucionais (CNC) e Redes Neurais Recorrentes (RNR). Esse tipo de IA é baseado num modelo bastante simplificado do cérebro humano, em que sua arquitetura são redes neurais

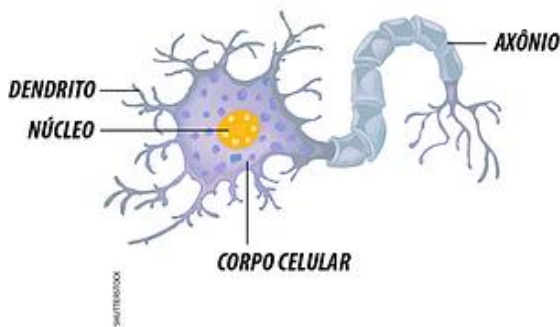


capazes de aprender com dados externos. Por exemplo, Redes Neurais Convolucionais levam o aprendizado com base em dados estruturados em grades, como imagens. Portanto, sistemas treinados nesse campo são usados para reconhecimento de imagem e vídeo, detecção de objetos e segmentação de imagem.

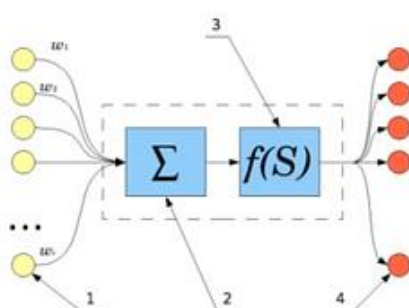
A unidade básica desse modelo de IA é o neurônio, que faz referência ao funcionamento da unidade básica do cérebro humano. Assim como o neurônio biológico recebe sinais de entrada (os dendritos), processa esses sinais em seu núcleo e produz um sinal de saída diferente (axônio), os neurônios artificiais fazem a leitura de dados (*inputs*), expressos por vetores de números reais, processam esses dados por meio de operações matemáticas e apresentam uma saída. A Figura 4 mostra os elementos de um neurônio biológico e de um neurônio. No neurônio artificial, as informações de entrada ( $x_i$ ) são multiplicadas por seus pesos sinápticos  $w_i$ , estes aprendidos durante o processo de treinamentos. Em seguida, esses sinais passam por uma combinação linear ( $\Sigma$ ), que então são agregados e geralmente corrigidos por um valor chamado “viés” ( $b_{ias}$ ). Os resultados  $S$  dessas operações passam por uma função ( $f$ ) conhecida como função de ativação, que limitam as saídas a um intervalo de valores (DATA SCIENCE ACADEMY, 2022b). Algumas funções de ativação comuns são a função sigmoide, a tangente hiperbólica e a função ReLU (*Rectified Linear Unit*).

Figura 4 - Neurônio biológico e funcionamento de um neurônio artificial

a) Esquema de um neurônio biológico



b) Esquema de um neurônio artificial



$$S = \sum_{i=1}^n x_i \cdot w_i + bias$$

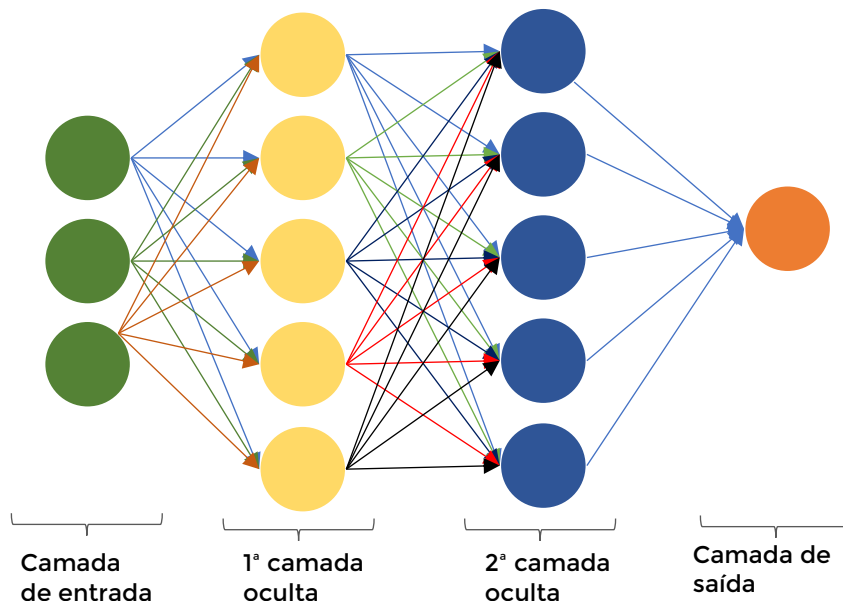
$$out = f(S)$$

Os outputs (out) do neurônio artificial são os valores que ele produz como resultado do processamento das entradas que recebe. Em síntese, esses valores são obtidos pela aplicação de uma função de ativação, que limita a saída em um intervalo determinado. Essa função de ativação é inspirada no limiar de excitação do neurônio biológico, que define se o impulso elétrico deve ou não ser transmitido para outros neurônios. Assim, os outputs do neurônio artificial simulam o comportamento do neurônio biológico na comunicação e propagação de informações (DATA SCIENCE ACADEMY, 2022b).

Assim como um neurônio humano isolado não é capaz de revelar como ocorre o processo de aprendizado humano, o neurônio artificial isolado não é capaz de aprender, mas quando dispostos em rede eles são capazes de cumprir uma gama enorme de tarefas, até a complexa tarefa de reconhecimento facial – amplamente usada em aplicativos bancários e sites governamentais.

Na Figura 5, abaixo, é possível observar uma arquitetura simplificada de uma Rede Neural Artificial (RNA) genérica. Dispostos em camadas, cada neurônio, representados por um círculo, envia informações a todos os neurônios da camada vizinha. Essa informação é basicamente um número resultante das operações matemáticas descritas anteriormente (Figura 4).

Figura 5 - Arquitetura de uma Rede Neural Artificial (RNA) genérica

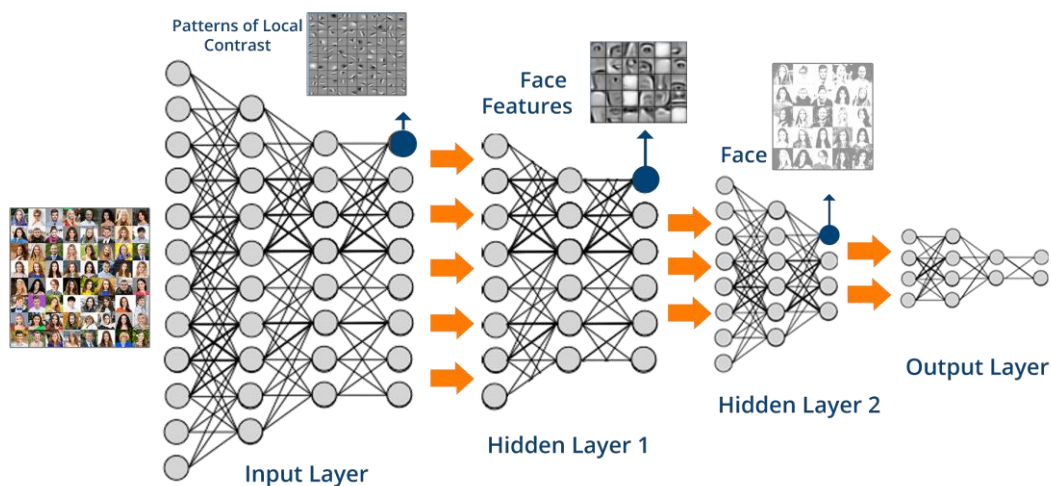


Fonte: Elaborado pelo próprio Autor.

Sem entrar nos detalhes técnicos da programação de RNA, seu funcionamento segue um conjunto de etapas que variam de acordo com a técnica de AM empregada e da complexidade do problema que se busca resolver. Em geral, um conjunto de dados de treinamento é informado na primeira camada, os neurônios dessa camada modificam os dados recebidos e passam para os neurônios que estão imediatamente à sua direita. No primeiro momento, modificar a informação significa multiplicar o dado recebido por um peso (número real). Então, os neurônios das camadas ocultas executam a soma dos dados que recebem e passam esse somatório pela função de ativação adotada no modelo. Os resultados de todas essas operações produzem os dados de entradas dos neurônios da próxima camada. Esse processo é feito para todos os neurônios até alcançar a camada de saída. A conexão entre dois neurônios, indicada por uma seta, tem sempre um peso o qual é modificado com o treinamento da RNA. Treinar uma RNA é aprender o peso de cada conexão entre neurônios da rede, isto é, à medida que se avalia o erro associada à saída, os pesos devem então serem ajustados para minimizar tal erro. Assim, o ajuste dos pesos na etapa da retroprogramação é o que torna possível o aprendizado em uma RNA (DATA SCIENCE ACADEMY, 2022a).

Quanto mais camadas tiver a rede maior é sua capacidade de lidar com tarefas mais complexas de aprendizado. Um exemplo de arquitetura baseada em RNA é o esquema da Rede Neural Convolutiva apresentada na Figura 6 (SANTANA, 2018). A tarefa dessa rede é realizar o reconhecimento facial. Cada uma das camadas apresentadas são RNA que aprenderam a realizar tarefas específicas, para então fornecer dados de entradas para as camadas posteriores. Essas múltiplas camadas interligadas constituem uma arquitetura típica de Aprendizado Profundo.

Figura 6 - Esquema de uma Rede Neural Convolutiva



Fonte: Santana (2018).

No domínio da comunicação, o Processamento de Linguagem Natural (PLN) foi mais um dos campos da IA que apresentou avanços acelerados na segunda década do terceiro milênio, graças aos avanços em AM, especialmente no âmbito do Aprendizado Profundo. Uma definição abrangente para esse subdomínio da IA é:

PLN, como principal tarefa da comunicação, refere-se à capacidade da máquina de identificar, processar, entender e/ou gerar informações em comunicações humanas escritas e faladas. É considerado um subdomínio de IA de várias estratégias nacionais e especialistas em IA abrangendo aplicações como geração de texto, mineração de texto, classificação e tradução automática (SAMOILI et al., 2021, p. 24 tradução nossa).

Conforme apresentado na definição acima, PNL é um subdomínio da IA que tem aplicação em muitas dimensões da produção do conhecimento, afetando todos os setores produtivos, os meios de comunicação e a comunidade científica em geral. Nesse campo, a IA é capaz de realizar uma grande variedade de tarefas de linguagem natural, incluindo tradução automática, resumo de texto, previsão de palavras, classificação de texto, análise de sentimento e geração de texto (SEJNOWSKI, 2023).

Alguns exemplos de produtos disponíveis para PNL são ferramentas de tradução de textos bastante fidedignas como “*Machine Translated by Google*”, chatbots inteligentes como o ChatGPT, versão GPT 4, da *Open IA* e o Bing Chat, da *Microsoft*. Os dois últimos são IAs generativas produzidos por Grandes Modelos de Linguagens (GML), um subcampo do Aprendizado Profundo. Diferente de *chatbots* programados para realizar uma tarefa específica, como os de autoatendimentos, esses *chatbots* são capazes de extrair informações valiosas de grandes quantidades de dados textuais e gerar trabalhos complexos semelhantes aos humanos, como escrever um conto ou uma resenha de uma obra conhecida. São capazes de produzir respostas e *feedbacks* em linguagem natural sobre um número quase irrestrito de assuntos (YU, 2023). O modelo por trás dessas ferramentas utiliza bilhões de parâmetros de treinamentos – o que é possível graças à disponibilidade cada vez maior de dados na rede de internet e do aumento do poder computacional e de memória.

Em suma, a IA ganhou protagonismo pela vasta capacidade de resolver problemas e proporcionar avanços em diversos setores. Em especial, a IA apresenta aplicações importantes na educação, resolvendo problemas relacionados à gestão educacional e alterando de alguma forma a relação dos estudantes e professores com o conhecimento. Nesse sentido, a próxima subseção apresenta algumas interfaces da IA com o campo educacional, bem como seus desafios e possibilidades.

### 2.3 IA na Educação: desafios e possibilidades

O uso da IA na educação tem uma história que remonta a várias décadas. Du Boulay (2022) cita como uma das experiências pioneiras da IA na educação, voltada para a promoção da aprendizagem, um sistema de tutoria e autoria chamado *Scholar*, descrito em 1970. A ferramenta baseada em redes semânticas se prestava ao ensino de geografia, sendo capaz de “raciocinar” sobre seu conhecimento geográfico da América do Sul para concluir informações que não foram programadas *a priori* tanto sobre o assunto da região que conhecia quanto de outras regiões.

“O sistema rastreia quais bits do domínio geográfico foram compreendidos pelo aluno marcando as partes relevantes da rede semântica, criando assim um modelo evolutivo do conhecimento do aluno” (DU BOULAY, 2022, p. 1). Embora bastante limitada, a ferramenta mencionada se distinguia de outros sistemas instrucionais assistidos por computador, pois era capaz de se adaptar ao indivíduo durante o processo de interação aluno-sistema.

Os avanços com sistemas especialistas no final da década de 1970 produziram ferramentas que eram capazes de tutorar programação elementar, ensinar algoritmos próprios dos sistemas especialistas e conteúdos específicos de matemática, como equações quadráticas. Em geral, esses sistemas baseados em IA constituíam o que se conhece como “ferramentas voltadas para o aluno”, isto é, ferramentas que apresentam modelos explícitos do que deve ser ensinado, como deve ser ensinado e um modelo evolutivo de compreensão e habilidades alcançadas, proporcionados por uma interface que permite a interação aluno-sistema (DU BOULAY, 2022).

Os avanços em IA relatados na Subseção 2.1, sobretudo no domínio do Aprendizado de Máquina, elevou a gama de aplicações da IA no campo educacional, produzindo ferramentas assistivas para professores e outras para ajudar os administradores educacionais (YU, 2023).

Em um estudo com objetivo de verificar a incidência e os assuntos tratados pelos pesquisadores brasileiros na interface entre educação e inteligência artificial, Gatti (2019) identificou diferentes áreas de aplicação da IA na educação, como EAD, Foco em Ensino-Aprendizagem, Gestão, *Learning Analytics* (Mineração de dados), Personalização, Avaliação e Inclusão. A maioria dos estudos se concentram no ensino superior, sendo incipientes as aplicações da IA na educação básica.

Contudo, o rápido desenvolvimento de soluções educacionais baseadas em inteligência artificial abre a possibilidade de sua inclusão na educação básica. Isso vai desde a personalização do ensino até a avaliação equitativa e eficaz da aprendizagem dos alunos.

Assim, sistemas recentes baseados em IA, voltados para alunos, apresentam interfaces intuitivas e possuem aspirações no processo de humanização da relação aluno-sistema. Incorporam constructos relativos à aprendizagem humana, como por exemplo a compreensão da motivação do aluno, mentalidade, sentimentos e emoções.

Embora essa evolução ajude a humanizar a interação entre sistemas e alunos, ela abre mais espaço para questões éticas relacionadas à privacidade e aos tipos de dados coletados e armazenados. Esse foco ampliado envolveu o desenvolvimento de técnicas para tentar avaliar os estados emocionais e motivacionais transitórios dos alunos, a fim de estimular estados mentais positivos, como concentração engajada, e combater estados mentais negativos, como frustração ou tédio. (DU BOULAY, 2022, p. 4, tradução nossa).

Os sistemas baseados em IA estão presentes nos mais variados dispositivos utilizados para tarefas rotineiras do dia a dia, como no smartphone, mas ainda adentram lentamente em escolas, conforme visto em Gatti (2019) e Du Doulay (2022). Em particular, ferramentas educacionais com foco nos professores, cujo objetivo é ajudá-los a compreender as particularidades de seus alunos visando a tomada de decisões adequadas para o estímulo à aprendizagem, costumam se restringir a modalidades como a educação à distância, sendo incipientes em redes educacionais públicas.

Algumas soluções voltadas para os professores têm potencial de ajudá-los na alocação de tempo, concentrando seus esforços na aprendizagem dos estudantes que mais precisam e na compreensão dos problemas comuns em uma classe. Um exemplo de ferramenta com essa finalidade é a *Teachy*, uma *edtech* brasileira que faz uso de IA em boa parte de suas soluções.

No âmbito dos sistemas voltados para gestores, são empregados especialmente na análise dos dados gerados em contextos educacionais desde a sala de aula até redes de ensino como um todo, como distribuição de carga horária e alocação de estudantes para escolas que atendem melhor suas necessidades. Oferecem soluções, como por exemplo, para o monitoramento do engajamento dos estudantes em cursos abertos massivos (MOOCs<sup>1</sup>), elucidando dificuldades individuais ou do grupo em relação ao conteúdo ofertado, a partir de diferentes padrões de envolvimento (DU BOULAY, 2022, p. 7).

Em todo caso, os sistemas modernos baseados em IA para a educação capturam quantidades cada vez maiores de dados dos sujeitos, tais como informações de identificação,

---

<sup>1</sup> Os Massive Open Online Courses (MOOCs) são cursos online gratuitos e abertos, sem requisitos mínimos. Criados em 2006, eles permitem o desenvolvimento pessoal e profissional, com atividades e testes online. Representam uma inovação na educação a distância, ampliando o acesso ao conhecimento.

desempenho acadêmico e comportamento, trazendo para discussão os desafios éticos relativos à privacidade. Soluções usadas para tomada de decisão no âmbito do ensino, gestão ou em qualquer outra dimensão educacional, deve operar dentro de uma estrutura ética e jurídica.

Segundo Yu (2023), a IA pode ser usada para tomar decisões importantes sobre os alunos, como a seleção de cursos, a atribuição de notas e a recomendação de carreiras, o que pode levar a preocupações com a justiça e a equidade dessas decisões. A nível de exemplo, se a IA for treinada com dados que refletem preconceitos ou desigualdades, ela pode perpetuar esses preconceitos em suas decisões, contribuindo para a manutenção dos problemas educacionais vigentes. Yu (2023) ressalta a importância de estabelecer leis e regulamentos correspondentes para proteger a privacidade e os direitos dos alunos, com objetivo na garantia de que a aplicação da IA na educação esteja em conformidade com a ética e os padrões morais da educação. Nesse sentido:

(...) o envolvimento da IA na tecnologia educacional também deve ser exigido para fazer o melhor e tratar os alunos com igualdade. Para ferramentas voltadas para o aluno, deve-se esperar que os projetistas da tecnologia educacional garantam que a tecnologia fará o melhor possível nas circunstâncias, seja ensinando, dando aulas particulares, orientando ou aconselhando os alunos. Deve-se também esperar que a tecnologia trate os alunos de maneira equitativa e não favoreça um aluno em detrimento de outro, inadvertida ou deliberadamente (DU BOULAY, 2022, p. 9, tradução nossa).

Para elucidar esse ponto de vista, suponha uma situação na qual um sistema baseado em IA é elaborado para ser um curador dos conteúdos que o estudante deve estudar e das tarefas que ele deve realizar durante um processo formativo. Para isso, seleciona cada tarefa com base no histórico educacional do estudante. Essa ferramenta hipotética baseada em IA tem “em mãos” uma decisão eticamente carregada. Se seu critério de seleção decorre de análise de informações como idade, sexo, sucesso ou insucesso em percursos educativos anteriores, motivação ou capacidade de aprendizagem autorregulada, e dependendo dos dados usados para o treinamento da IA, há riscos de que tais critérios de seleção dos conteúdos/tarefas possam apresentar vieses e contribuir para manutenção ou ampliação das desigualdades educacionais presentes no grupo de sujeitos que participam do processo formativo em questão.

Assim, os aspectos éticos e riscos associados ao uso de IA tornou-se focal entre pesquisadores e instituições ao redor do mundo. Um exemplo de estratégia foi apresentado em 2020 pela Comissão Europeia (BIDARRA; SIMONSEN; HOLMES, 2020) para ampliar e aperfeiçoar as relações da IA com a educação, visando sobretudo estimular o uso crítico e responsável de IA nesse contexto. A ação proposta define três eixos: i) aprendizagem com IA,

que visa estimular o estudo das aplicações de IA voltadas a educação, sobretudo para a atenuação das desigualdades educacionais; ii) aprendizagem sobre a IA para compreender o funcionamento da IA; iii) aprendizagem para a IA, que focaliza a avaliação de impactos da IA na sociedade, aspectos éticos e privacidade.

Em 2019, a Conferência Internacional sobre Inteligência Artificial e Educação, intitulada “Planejando a educação na era da IA: liderar o avanço”, foi realizada em Pequim, na China. O evento contou com a presença de 50 ministros e vice-ministros de governo, além de aproximadamente 500 representantes de agências das Nações Unidas, instituições acadêmicas, sociedade civil e setor privado para discutir o assunto (UNESCO, 2019). A conferência apresentou 10 recomendações de ações a serem implementadas por governos e partes interessadas no âmbito da inteligência artificial na educação.

- 1) *Planejando a IA nas políticas educacionais*: alinhar a IA na educação com políticas públicas, definir prioridades estratégicas e desenvolver estratégias coerentes em todo o sistema para a IA na educação.
- 2) *IA para gerenciamento e entrega da educação*: integrar ou desenvolver tecnologias e ferramentas de IA relevantes para aprimorar a coleta e o processamento de dados, tornando o gerenciamento e o fornecimento de educação mais equitativos, inclusivos, abertos e personalizados.
- 3) *IA para capacitar professores e o ensino*: fornecer treinamento e desenvolvimento profissional para professores e educadores para que possam usar a IA de forma eficaz em suas práticas de ensino.
- 4) *IA para aprendizagem e avaliação da aprendizagem*: usar a IA para personalizar a aprendizagem e avaliar o progresso dos alunos de forma mais eficiente e precisa.
- 5) *Desenvolvimento de valores e habilidades para a vida e para o trabalho na era da IA*: desenvolver habilidades e valores necessários para a vida e o trabalho na era da IA, como pensamento crítico, criatividade e colaboração.
- 6) *IA como forma de oferecer oportunidades de aprendizagem ao longo da vida para todos*: usar a IA para fornecer oportunidades de aprendizagem ao longo da vida para todos, independentemente de idade, gênero, localização ou status socioeconômico.
- 7) *Promover o uso equitativo e inclusivo da IA na educação*: garantir que o uso da IA na educação seja equitativo e inclusivo, levando em consideração as necessidades de grupos marginalizados e vulneráveis.



- 8) *IA com igualdade de gênero e IA para igualdade de gênero*: garantir que a IA seja usada para promover a igualdade de gênero na educação e evitar a perpetuação de estereótipos de gênero.
- 9) *Garantir o uso ético, transparente e auditável dos dados e algoritmos educacionais*: garantir que o uso da IA na educação seja ético, transparente e auditável, protegendo os dados dos alunos e garantindo que os algoritmos usados sejam justos e imparciais.
- 10) *Monitoramento, avaliação e pesquisa*: monitorar e avaliar continuamente o uso da IA na educação para garantir que esteja alinhado com os objetivos educacionais e os valores éticos, e realizar pesquisas para aprimorar a compreensão do impacto da IA na educação.

Dentre as ações recomendadas, a de número 4 chama atenção para o potencial que a IA possui na avaliação da aprendizagem, visto que, se implantada de forma responsável elas podem propiciar avaliações formativas personalizadas, cujo objetivo é o aperfeiçoamento dos processos de ensino e aprendizagem escolar.

Por fim, cabe mencionar que as interfaces da IA com a educação impõe desafios contemporâneos que vão além da otimização das ferramentas que possam direcionar e aperfeiçoar os processos de ensino, como também: a necessidade de conscientização e regulamentação referentes à privacidade de dados; transparência dos algoritmos para que as decisões tomadas possam ser compreendidas e supervisionadas por profissionais da educação; adoção de ações que proponham além da aprendizagem com IA, a aprendizagem sobre IA e a aprendizagem para a IA.

## **2.4 Ferramentas de IA para a avaliação da aprendizagem**

Sistemas baseados em IA estão cada vez mais eficientes na leitura de dados que podem identificar estados emocionais, preferências e até o nível de compreensão do sujeito sobre um determinado assunto. Com isso, soluções voltadas para a avaliação da aprendizagem já estão disponíveis para uso em diferentes contextos, inclusive escolar. Parte dessas soluções são produzidas pelas "EdTechs", empresas de tecnologia que atuam na área educacional.

A nível de exemplo, a ferramenta *Quizizz*, que pode ser usada gratuitamente, é capaz de gerenciar um banco de perguntas e ofertar ao respondente aquelas que podem ser compatíveis com seus conhecimentos prévios. Os bancos adaptativos oportunizam uma avaliação progressiva das habilidades a serem verificadas por meio de exames personalizados

para cada aluno. Essas soluções atacam umas das principais dificuldades do docente que lida com salas de aula lotadas, o da personalização do ensino.

O Quizizz permite ainda a elaboração de itens para avaliação da aprendizagem com o auxílio de ferramentas de IA generativa. O docente fornece um link da web que contenha o conteúdo que pretende abordar na avaliação. Posteriormente, a plataforma gera automaticamente testes e itens em diversos formatos. Estes podem ser impressos ou atribuídos como atividade gamificada, permitindo que os alunos respondam por meio de dispositivos móveis ou computadores. Esta abordagem facilita a criação de avaliações personalizadas, promovendo uma experiência de aprendizado mais envolvente.

Outro aspecto relevante do uso de ferramentas como o *Quizizz* é a capacidade de fornecer relatórios detalhados sobre o nível de aprendizado de cada estudante. Essas informações podem ser úteis para os professores, pois permitem identificar áreas de dificuldade e sucesso, auxiliando na intervenção e reformulação dos planos de ensino.

Há ferramentas que podem ser usadas para revisar, classificar e avaliar diversos exames e outros testes não numéricos, como redações, de maneira rápida e precisa. Há aplicações que permitem avaliar o nível de fluência de leitura do estudante em uma dada etapa da sua escolarização. Com tantos dados educacionais disponíveis até mesmo do contexto da sala de aula, ferramentas baseadas em IA podem ser usadas para obter dados sobre o desempenho e comportamentos dos alunos, permitindo que os professores tomem decisões que o ajudem a planejar suas aulas.

Na visão do pesquisador, o uso de sistemas inteligentes apresenta quatro vantagens principais nos processos de avaliação da aprendizagem. Primeiramente, a IA permite a elaboração de itens personalizados. As ferramentas de IA podem gerar e atribuir tarefas aos estudantes de acordo com seu nível de conhecimento e habilidades, proporcionando uma experiência de aprendizado mais individualizada.

Em segundo lugar, a IA pode ser empregada para corrigir tarefas de forma automática. Isso libera tempo para que os professores se concentrem em atividades de ensino e orientação, otimizando o uso de seus recursos.

A terceira vantagem é a capacidade de fornecer feedback personalizado. As ferramentas de IA podem destacar as áreas fortes dos alunos e os pontos que ainda precisam ser aperfeiçoados, oferecendo orientações específicas para cada estudante.

Por fim, a IA pode ser utilizada para analisar dados de desempenho de forma rápida e eficiente. A identificação de padrões e tendências nos dados pode ser útil para os professores, fornecendo *insights* sobre o progresso dos alunos.

Contudo, é crucial destacar o desafio enfrentado por uma parcela significativa de alunos e professores da educação básica, particularmente aqueles situados em comunidades rurais, no que diz respeito ao acesso a essas tecnologias. A disponibilidade de internet e tecnologia emergem como obstáculos fundamentais para a promoção de uma educação pública acessível e equitativa no Brasil.

Em síntese, a Seção 2 apresentou um breve histórico da inteligência artificial (IA), desde seus primórdios até os avanços recentes, destacando os principais modelos e paradigmas que a sustentam. Em seguida, discutiu as possibilidades e os desafios da aplicação da IA na educação, tanto para apoiar os processos de ensino e aprendizagem quanto para avaliar a aprendizagem dos alunos. Por fim, abordou as questões éticas e regulatórias relacionadas ao uso da IA na educação, enfatizando a necessidade de garantir a privacidade, a equidade e a transparência dos dados e algoritmos educacionais.

Em consonância com os objetivos propostos neste estudo, abordaremos na seção seguinte alguns aspectos da avaliação educacional na educação básica, que são essenciais para fundamentar teoricamente as decisões e as escolhas no processo de elaboração de itens com o auxílio da IA generativa.

### **3 AVALIAÇÃO EDUCACIONAL NA EDUCAÇÃO BÁSICA**

O avaliar no contexto educacional é um empreendimento que traz consigo desafios teóricos e metodológicos em praticamente todas as suas dimensões, quer sejam na sala de aula, no âmbito institucional, em programas e projetos educativos, no currículo ou no âmbito de sistemas. Essas dimensões e alguns conceitos da avaliação educacional são brevemente discutidos na Subseção 3.1. A Subseção 3.2 discute termos comuns nas avaliações externas, mas que estão se tornando comuns no âmbito escolar. Por fim, a Subseção 3.3 apresenta algumas diretrizes relevantes para elaboração e revisão de itens de múltipla escolha, dado que esse modelo de item é amplamente utilizado em testes padronizados.

#### **3.1 Dimensões da avaliação educacional**

A avaliação envolve coletar, analisar e sintetizar dados para atribuir valor ou qualidade ao objeto avaliado, comparando-o com um padrão de qualidade pré-estabelecido. Isso leva a uma tomada de posição a favor ou contra o objeto, resultando em uma decisão de mantê-lo como está ou agir sobre ele (LUCKESI, 2013).

No âmbito da avaliação educacional, o ato de avaliar se distingue em seus diferentes contextos. Sousa (2000) descreve cinco dimensões da avaliação, a saber:

- i) avaliação de sala de aula, cujo foco é o ensino e a aprendizagem e visa avaliar o desempenho dos educandos e o progresso em relação aos objetivos educacionais estabelecidos;
- ii) avaliação institucional, que se ocupa da avaliação da qualidade e eficácia da instituição educacional como um todo, incluindo aspectos como gestão, infraestrutura, recursos humanos e financeiros;
- iii) avaliação de programa e projetos educativos, a qual avalia a eficácia e eficiência de programas e projetos educativos específicos, incluindo aspectos como planejamento, implementação e resultados alcançados;
- iv) avaliação de currículo, que visa avaliar a adequação e relevância do currículo em relação aos objetivos educacionais estabelecidos e às necessidades dos alunos;
- v) avaliação de sistema, que lida com avaliações de desempenho do sistema educacional como um todo, incluindo aspectos como políticas públicas, financiamento e gestão.

Nesse estudo, é apresentado conceitos relativos à avaliação de sala de aula e à avaliação de sistema, mais conhecida como avaliação de larga escala. A avaliação de sala de aula tem como foco o acompanhamento escolar do aluno, possibilitando a tomada de decisões durante o processo instrucional. É por meio dessas avaliações que o professor adquire informações para promover reajustes em suas práticas, como propor novas intervenções, aperfeiçoar sua metodologia, e assim busca oferecer um ensino que leve em conta as particularidades dos indivíduos e do grupo.

Diferentemente da avaliação de sistema, que utiliza testes objetivos e é realizada em grande escala, a avaliação de sala de aula exige a formação de professores que apresentem uma atitude avaliativa, isto é, são predispostos a reflexões constantes sobre suas estratégias metodológicas, as aprendizagens dos alunos e as condições educacionais oferecidas pela escola. A principal característica da avaliação de sala de aula é a sua natureza formativa à medida que busca fornecer *feedbacks* aos alunos e ao professor sobre as habilidades e competências desenvolvidas e as que precisam ser desenvolvidas pelo indivíduo em cada etapa da sua escolarização (SOUSA, 2000).

A avaliação tem com objeto a qualidade e pode contribuir para a formação integral dos estudantes ao permitir que o educador avalie não apenas o desempenho do aluno, mas também suas habilidades, competências e atitudes, uma vez que a avaliação qualitativa vai além de resultados obtidos, mas também o processo de aprendizagem, as dificuldades e as estratégias utilizadas para superá-las (LUCKESI, 2012). A avaliação escolar, portanto, tem como foco a aprendizagem, o aperfeiçoamento do ensino e, conseqüentemente, o êxito no desenvolvimento das competências necessárias para a vida.

No Brasil, a avaliação da aprendizagem enquanto campo teórico começou a ser discutida nos anos 1960 e 1970, substituindo o paradigma do avaliar como sinônimo de proceder com exames escolares – perspectiva a qual o docente assume o papel de “examinador”. A Lei de Diretrizes e Bases da Educação, de 1996, introduziu oficialmente a expressão “avaliação da aprendizagem”. Apesar das mudanças na legislação, a prática escolar ainda prioriza os exames em detrimento da avaliação no seu sentido mais amplo. Dentre outros aspectos, examinar se caracteriza pela classificação e seletividade do educando, ao passo que avaliar é um ato voltado pelo seu diagnóstico e pela sua inclusão (LUCKESI, 2013).

Para efeito de distinção da avaliação da aprendizagem escolar do procedimento de verificação da aprendizagem, guiada pelo paradigma do exame, Luckesi (2013) descreve avaliação como um ato que ultrapassa a o conceito de “verificar” a configuração do objeto

avaliado à medida que exige uma decisão sobre e o que fazer com ele, isto é, ao passo que verificar “congela” o objeto, a avaliação o direciona em um caminho dinâmico de ação.

O modo de trabalhar com os resultados da aprendizagem escolar — sob a modalidade da verificação — reifica a aprendizagem, fazendo dela uma “coisa” e não um processo. O momento de aferição do aproveitamento escolar não é ponto definitivo de chegada, mas um momento de parar para observar se a caminhada está ocorrendo com a qualidade que deveria ter. Neste sentido, a verificação transforma o processo dinâmico da aprendizagem em passos estáticos e definitivos. A avaliação, ao contrário, manifesta-se como um ato dinâmico que qualifica e subsidia o reencaminhamento da ação, possibilitando consequências no sentido da construção dos resultados que se deseja (LUCKESI, 2013, p. 54).

Para evitar que a avaliação de sala de aula promova exclusão, o educando precisa ser visto como um todo, levando em conta não apenas o aspecto cognitivo, tipicamente mensurados por testes avaliativos, mas também dimensões socioafetivas e contextuais que impõe condições e restrições ao desenvolvimento de sua aprendizagem. Deseja-se que a avaliação em sala de aula, promovida continuamente para avaliar o domínio ou não das habilidades e competências previstas no currículo, propicie oportunidade para o educando demonstrar o que aprendeu e, para isso, é necessário que o professor crie situações que permitam a ele expressar suas aprendizagens (SOUSA, 2000).

No decorrer do processo de avaliação da aprendizagem escolar, cabe ao professor coletar, analisar e sintetizar objetivamente as manifestações das condutas dos alunos, atribuir uma qualidade a essa configuração da aprendizagem com base em um padrão preestabelecido e tomar uma decisão sobre as condutas a serem seguidas. Isso pode incluir a reorientação imediata da aprendizagem ou o encaminhamento dos alunos para passos subsequentes, dependendo da qualidade da aprendizagem (LUCKESI, 2013).

A conduta avaliativa descrita por Luckesi (2013) esbarra em desafios práticos, sobretudo na educação pública. Em primeiro lugar, o professor lida com salas de aulas lotadas que dificultam a recomposição da aprendizagem dos educandos que não desenvolveram as habilidades esperadas e, em segundo, há pressão para o cumprimento de um currículo extenso, com dezenas de habilidades e competências, as quais requerem um tempo muito maior para o seu domínio. Outro ponto, diz respeito às jornadas extensas de trabalho que dificultam a implementação de uma postura avaliativa e reduzem o avaliar à verificação da aprendizagem. Embora esforços individuais pautados numa conduta avaliativa voltada para a promoção da aprendizagem possam surtir efeitos em realidades particulares, a mudança será efetiva quando a avaliação de sala de aula for ressignificada nas secretarias de educação, nos centros de formação de professores e na comunidade escolar como um todo.

Se avaliar a aprendizagem escolar é um conjunto de ações voltadas para o diagnóstico e o replanejamento do ensino, as avaliações de sistema (avaliação externa e de larga escala) atuam em um campo muito maior. Esse tipo de avaliação tem como objetivo fornecer subsídios para o desenvolvimento de políticas educacionais em escala federal, estadual e municipal. Atua na análise do sistema educacional como um todo, incluindo a estrutura, a organização, a gestão e os resultados obtidos. Essas avaliações são empregadas para identificar as desigualdades de educação oferecida nas escolas dos diferentes sistemas de ensino e tomar decisões no sentido de corrigir injustiças (SOUSA, 2000).

A avaliação do sistema de ensino é constituída pela investigação da qualidade do desempenho de todos os componentes educativos formais do país, que, em uma visão ascendente, investiga a qualidade do resultado do desempenho dos professores em sala de aula, a seguir, aborda a qualidade do desempenho das escolas e, subsequentemente, aborda o desempenho da organização municipal, estadual e federal do ensino; evidentemente, incluindo as escolas particulares dos diversos níveis de ensino; numa visão descendente, essa estrutura segue do mais para o menos amplo. A avaliação de larga escala inclui em seu foco investigativo o sistema de ensino nacional, incluindo todas as suas partes (LUCKESI, 2022, p. 196).

Assim, compete ao Estado controlar a qualidade da educação oferecida em suas escolas e implantar medidas que permitam garantir a equidade no processo educacional. A avaliação de sistema é fundamental para a formulação de políticas públicas na área educacional, pois permite que o Estado tenha uma visão ampla e detalhada do sistema educacional e possa tomar decisões embasadas em dados concretos (SOUSA, 2000).

A avaliação sistema no Brasil tem uma história de mais de três décadas, iniciada com o Sistema de Avaliação da Educação Básica (Saeb) em 1990. Sob responsabilidade do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), o Saeb é um conjunto de avaliações externas que permite realizar um diagnóstico da educação básica brasileira e de fatores externos que podem interferir na aprendizagem do estudante. Os testes cognitivos e os questionários são aplicados a cada dois anos em carácter amostral ou censitário (BRASIL, 2018b). Ao longo dos anos, outras avaliações externas como o Exame Nacional do Ensino Médio (Enem), Programa Internacional de Avaliação de Estudantes (Pisa) e o Programa de Avaliação da Alfabetização (Proalfa), foram sendo incorporadas na educação brasileira.

A disseminação da cultura da avaliação externa na educação brasileira passou a ser tema de diversos estudos, além de fornecer subsídios para implementação de políticas públicas educacionais, como mencionado anteriormente. Não demorou para que sistemas de ensino a níveis municipal e estadual passassem a investir em sistemas próprios de avaliações externas. Pelo menos 23 estados brasileiros registraram algum sistema de avaliação externa, de acordo

com um relatório do Instituto Reúna<sup>2</sup> (INSTITUTO REÚNA, 2020). Parte desses investimentos são motivados pelo atraso com que chegam os resultados dos testes cognitivos realizados pelo Sistema Avaliação da Educação Básica (Saeb), o que impossibilita seu uso para tomadas de decisão no âmbito do ensino (GIMENES et al., 2013).

No entanto, a consolidação de sistemas de avaliações próprios requer expressivos investimentos, o que nem sempre é viável. Em municípios pouco populosos e sem recursos significativos, a criação de um sistema de avaliação próprio é praticamente inimaginável, tanto pela dificuldade de encontrar profissionais do campo da avaliação qualificados quanto pelas dificuldades no próprio processo de operacionalização dos testes/questionários e da análise de seus resultados. Outro ponto, são as especificidades que os itens (questões) que compõe os testes devem ter para avaliarem de forma justa a aprendizagem dos sujeitos.

Em meio a dificuldades no monitoramento da aprendizagem dos estudantes, sobretudo no período pós-pandemia da COVID 19, as redes de educação municipais e estaduais tiveram acesso a uma plataforma de avaliações diagnósticas e formativa do Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF). A ferramenta contou com testes cognitivos de matemática, língua portuguesa e avaliação de fluência de leitura, disponíveis para aplicação em sala de aula. Lançada em outubro de 2021, a plataforma teve suas atividades encerradas em maio de 2023.

Nesse contexto, é possível notar que as avaliações externas, primeiramente conduzida em âmbito nacional, passaram a adentrar nas redes de ensino municipais e estaduais por meio de sistemas próprios e, mais recentemente, por meio de plataformas de avaliações que podem ser usadas diretamente por professores e coordenadores escolares como forma de diagnosticar os níveis de aprendizagem de seus estudantes, cujos testes cognitivos têm formatos semelhantes aos testes do Saeb. Essa aproximação das avaliações externas com a sala de aula tem efeitos discutidos na literatura com diferentes enfoques.

Um dos enfoques atenta para o risco de as escolas serem levadas a uma exaustiva rotina de avaliações externas, o que pode subtrair tempo útil para o desenvolvimento do trabalho pedagógico e favorecer o desgaste desse tipo de ação avaliativa (GIMENES et al., 2013). Outro ponto questionado a respeito das implementações de avaliações externas são os riscos do estreitamento do currículo para focar nas matérias avaliadas nos testes cognitivos. O

---

<sup>2</sup> Um total de 23 estados já registraram experiências de avaliação em larga escala, são eles: Acre, Alagoas, Amazonas, Amapá, Bahia, Ceará, Espírito Santo, Goiás, Mato Grosso do Sul, Maranhão, Minas Gerais, Pará, Paraíba, Paraná, Pernambuco, Piauí, Rio de Janeiro, Rio Grande do Sul, Rio Grande do Norte, Rondônia, Santa Catarina, São Paulo e Tocantins.



estreitamento também pode ocorrer nas práticas e na escolha dos conteúdos, no caso em que os professores se vêm diante de pressões externas para focar tão somente nas habilidades descritoras previstas nas matrizes de referência dos testes cognitivos, se atendo aos modelos de itens (questões) cobrados nesses testes. Assim, a avaliação de sistema pode ser vista como uma interferência externa no trabalho pedagógico da escola, a qual impõe padrões e critérios externos à realidade da escola (MACHADO; ALAVARSE, 2014).

Por outro lado, as práticas pedagógicas escolares podem ter ganhos com as avaliações externas quando elas forem usadas para promover a cultura da avaliação, da autoavaliação e para incentivar a reflexão crítica sobre as práticas de ensino na busca pela excelência (MACHADO; ALAVARSE, 2014). Ao passo que as avaliações externas são basilares na implementação de políticas educacionais, os resultados de desempenho revelados nessas avaliações têm potencial de contribuir para a adoção de práticas pedagógicas mais efetivas, uma vez que os resultados podem ser utilizados como base para a tomada de decisões e o planejamento de intervenções educacionais.

Com essa aproximação e apropriação do formato das avaliações externas no âmbito escolar, termos técnicos das avaliações externas, tais como, “avaliação de larga escala”, “matriz de referência”, “eixos cognitivos”, “escala de proficiência”, “Teoria de Resposta ao Item”, “Item”, dentre outros, ganharam popularidade, mas permanecem desconhecidos por boa parte dos profissionais da educação. A consequência do desconhecimento de tais conceitos é de que pouco se compreende os significados pedagógicos dos resultados dessas avaliações, dificultando sua interpretação e seu uso no aperfeiçoamento do ensino.

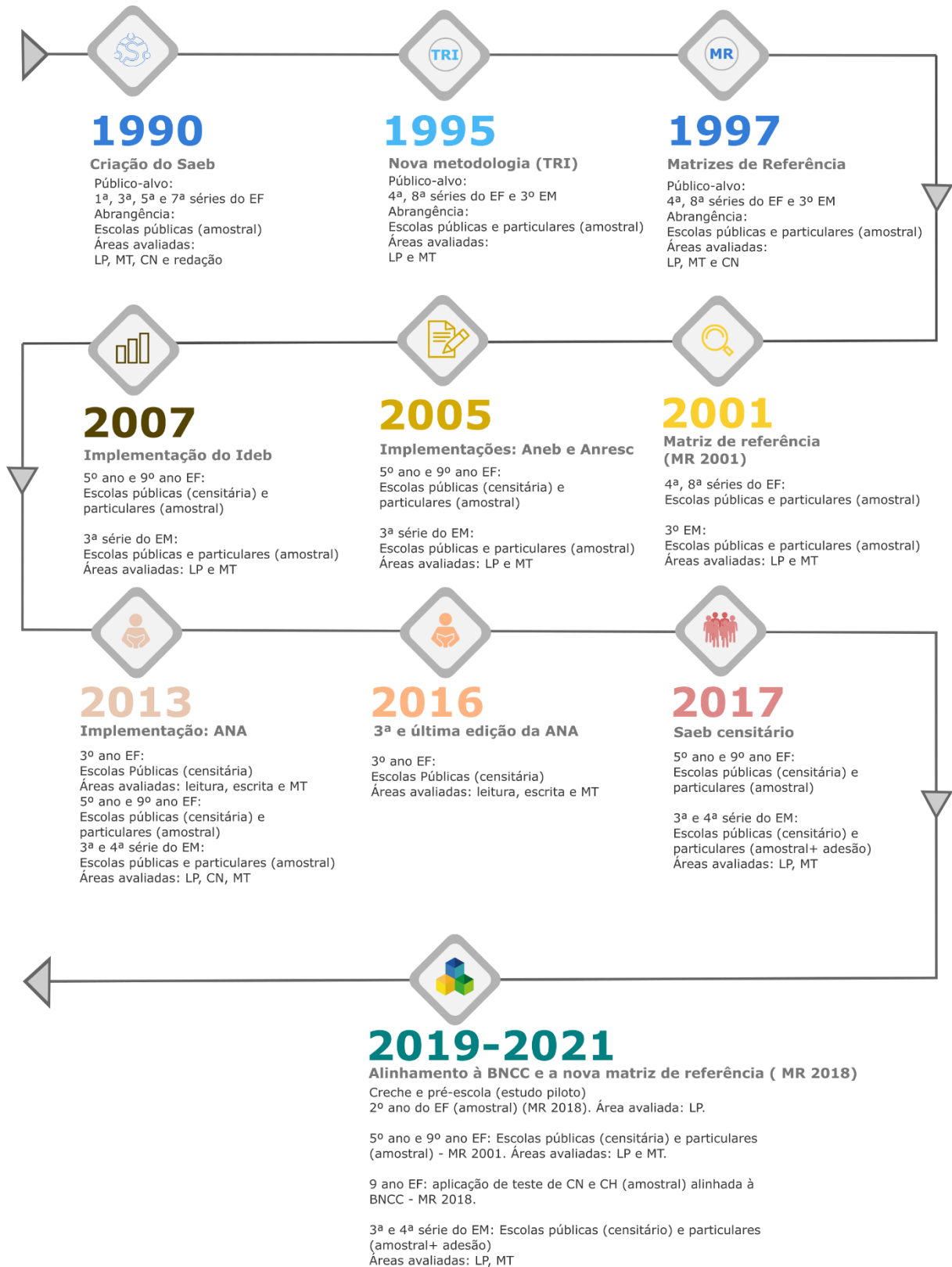
Diante do exposto, a Subseção 3.2 apresenta algumas definições, de forma sucinta, e conceitos amplamente usadas no contexto das avaliações externas, mas que precisam ser compreendidas pelos profissionais da educação básica, pois, como mencionado anteriormente, as avaliações externas estão cada vez mais presentes nas escolas e exercem sobre elas influências importantes.

### **3.2 Definições e conceitos utilizados na avaliação em larga escala**

Esta seção apresenta conceitos e definições utilizados na área de avaliação educacional, especialmente nas avaliações nacionais de sistemas que, no Brasil, são conduzidas pelo Inep. O desenho das avaliações externas da educação básica coordenadas a níveis municipal e estadual se respaldam nos procedimentos empregados pelo Inep, tanto para fundamentar a elaboração de itens, visando à composição de testes de avaliação em larga escala, quando nas metodologias de análise e apresentação dos seus resultados. O Inep é o órgão responsável pelo Saeb, que realiza um conjunto de avaliações externas de larga escala. O Saeb busca diagnosticar a situação da educação básica no Brasil e identificar os fatores que podem influenciar o desempenho dos estudantes (BRASIL, 2023b).

A linha do tempo exibida na Figura 7 mostram as principais mudanças ocorridas no Saeb desde sua implementação. O aperfeiçoamento do Saeb inclui adoção da Teoria de Resposta ao Item (TRI) em 1995, a fim de tornar os resultados de rendimento comparáveis ao longo do tempo; consolidação de matrizes de referências (1997, 2001, 2018); criação do Índice de Desenvolvimento da Educação Básica (Ideb) (2008); ampliação do público-alvo, tornando o Saeb censitário para toda a educação básica (níveis fundamental e médio) e, mais recentemente avaliações da educação infantil (2019).

Figura 7 - Linha do tempo com as principais mudanças no Saeb: Brasil 1990-2021



Fonte: Inep, 2023. Elaborado pelo próprio Autor.

### 3.2.1 Matriz de referência

Em avaliação educacional, Matriz de Referência é um documento base para construção de itens. Nesse tipo de avaliação, itens são as conhecidas “questões”. Para tanto, uma Matriz de Referência define objetivamente competências e habilidades que se espera que o educando tenha alcançado em uma etapa específica de sua escolarização.

No contexto do Saeb, duas matrizes foram propostas para elaboração dos testes cognitivo, a saber, a Matriz de Referência introduzida no ano de 2001 e a Matriz de Referência de 2018. Essa última faz parte do processo de alinhamento dos testes cognitivos do Saeb com a Base Nacional Comum Curricular (BNCC). Esse alinhamento é importante porque a BNCC orienta os currículos e propostas pedagógicas de todas as escolas do país, com o objetivo de garantir que todos os estudantes tenham acesso aos aprendizados essenciais para suas formações, incluindo conhecimentos, competências e habilidades que devem ser desenvolvidas ao longo do ensino básico (BRASIL, 2023a).

Apesar de ser alinhada à BNCC, a Matriz de 2018 não se confunde com o currículo, isto é, não define quais habilidade e competências devem ser desenvolvidas no processo de escolarização. O documento se apropria de habilidades e competências pontuais, que podem ser avaliadas por meio de um instrumento de medida, neste caso, por um conjunto de itens objetivos (BRASIL, 2023a).

Estados e municípios que dispõem de sistemas próprios de avaliação da qualidade do ensino ofertado por suas redes possuem matrizes específicas, mas que têm o currículo básico (BNCC) como alicerce para definição das habilidades a serem avaliadas por meio dos testes cognitivos.

### 3.2.2 Competências e Habilidades

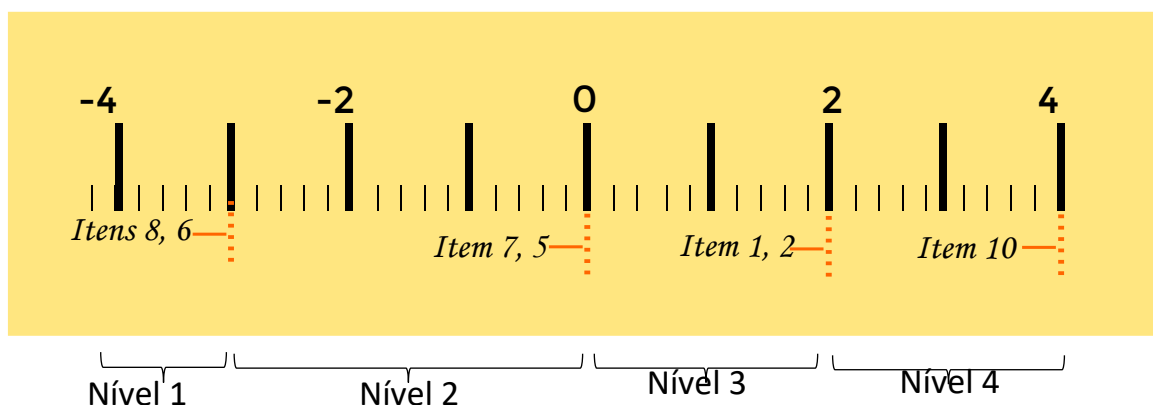
Embora tenha múltiplos significados, em avaliação habilidades são esquemas que orientam as operações mentais e as operações concretas. Habilidades ou *skills* são saberes processuais elementares que podem ser avaliados a partir um curso de ações, sejam elas operacionais ou mentais. Por outro lado, competências evocam domínio global de uma situação, requerendo diversas habilidades para tal domínio. Uma competência é, portanto, mais ampla e apresenta um número maior de parâmetros a serem geridos, isto é, “as competências estão relacionadas a um conjunto de situações, e as habilidades a operações ou a esquemas que podem funcionar como recursos a serviço de múltiplas competências” (PERRENOUD, 2013, p. 49).

A nível de exemplo, a competência 5 de matemática para o ensino fundamental previsto na BNCC traz o seguinte texto: “Utilizar processos e ferramentas matemáticas, inclusive tecnologias digitais disponíveis, para modelar e resolver problemas cotidianos, sociais e de outras áreas de conhecimento, validando estratégias e resultados” (BRASIL, 2018a, p. 267). Dentre várias outras habilidades esperadas que um estudante do 9º ano desenvolva para a competência 5, a habilidade “(EF09MA05) Resolver e elaborar problemas que envolvam porcentagens, com a ideia de aplicação de percentuais sucessivos e a determinação das taxas percentuais, preferencialmente com o uso de tecnologias digitais, no contexto da educação financeira” (BRASIL, 2018a, p. 317) especifica um dos domínios importantes para o desenvolvimento de tal competência.

### 3.2.3 Escala de Proficiência

Escala é uma régua, isto é, um conjunto de números ordenados, que permite comparar e interpretar, a partir dos itens de um teste e da proficiência, o quanto um determinado grupo de sujeitos sabe ou é capaz de realizar a respeito das habilidades e competências avaliadas (BRASIL, 2023a). Na Figura 8, observa-se uma régua que exibe uma escala de proficiência que varia de -4 a 4, composta por quatro níveis distintos de proficiência. Cada nível é acompanhado por itens específicos que facilitam a interpretação da respectiva escala, conhecidos como itens âncoras. A partir da proficiência alcançada pelo indivíduo ou por um grupo de indivíduos é possível identificar as habilidades potencialmente desenvolvidas.

Figura 8 – Ilustração de uma escala de proficiência



Fonte: Elaborado pelo próprio Autor.

As escalas de proficiência são produzidas no âmbito da Teoria de Resposta ao Item (TRI) e são usadas nas avaliações externas de larga escala, permitindo que se façam interpretações pedagógicas dos resultados demonstrados pelos sujeitos avaliados, assim como permitem posicionar os itens na mesma escala e interpretá-la quanto ao seu nível de dificuldade, capacidade de distinguir sujeitos mais proficientes dos sujeitos menos proficientes e estimar a probabilidade de um sujeito com baixo nível de proficiência responder o item corretamente.

Assim, a escala de proficiência pode ser conceituada como um instrumento analítico que possibilita a categorização dos indivíduos avaliados em distintos níveis de habilidade. Cada estrato da escala sinaliza as habilidades que o educando provavelmente domina ao ser enquadrado nesse nível. Esta característica confere à escala de proficiência um valor significativo para a análise da qualidade educacional, uma vez que facilita interpretações pedagógicas dos testes aplicados (BRASIL, 2023a; KLEIN, [s.d.]).

#### 3.2.4 Item

O item é a unidade mais elementar dos instrumentos de avaliação da aprendizagem e das demais variáveis investigadas nas avaliações externas. No caso da avaliação da aprendizagem escolar, o item são as populares “questões” produzidas com a finalidade de mobilizar atitudes, procedimentos e análises por parte do educando, a respeito de uma competência ou simplesmente de uma habilidade. Existem duas categorias de itens: os de resposta livre, nos quais o educando escreve sua própria resposta, e as de resposta orientada, também conhecidos como itens objetivos ou de múltipla escolha, onde o aluno escolhe uma resposta dentre várias opções fornecidas (BRASIL, 2010; VIANNA, 1978). Há uma variedade de itens de respostas orientadas, conforme discute Vianna (1978), mas nesse estudo é dado enfoque nos itens tipicamente construídos para avaliações externas – o que será tema da Subseção 3.3.

### **3.3 Elaboração de itens de matemática para avaliação da aprendizagem**

Para que a avaliação seja um instrumento significativo no ensino escolar, é importante que tanto a prática educativa quanto a avaliação sejam conduzidas com rigor científico e técnico (LUCKESI, 2013). Assim, a avaliação da aprendizagem requer elaboração de instrumentos que sejam fidedignos e válidos, para não correr risco de penalizar o sujeito avaliado com testes e provas que não mensuram o que propõem mensurar.

Essa subseção apresenta algumas dicas para elaboração de itens cuja finalidade é avaliar a aprendizagem do estudante em uma certa etapa da educação escolar, na componente curricular matemática. Em particular, trata da construção dos itens objetivos de múltipla escolha, pois são itens que facilitam o processo de leitura das respostas e a análise estatística dos resultados.

Antes de tratar do que propõe essa subseção, é importante elucidar que os itens de múltipla escolha nos modelos produzidos para avaliação de larga escala não podem ser tomados como único instrumento de avaliação da aprendizagem ou como o instrumento mais preciso e apropriado para tal objetivo.

Os itens presentes nas avaliações de larga escala são produzidos por especialistas da área, testados e validados com rigor psicométrico apropriado, mas mensuram aspectos muito elementares do currículo. Os múltiplos comportamentos desempenhados no ambiente da sala de aula, as atitudes, os valores, as percepções individuais que os indivíduos possuem no seu íntimo não são capturadas por itens como esses.

Assim, defende-se que a adoção de itens de múltipla escolha seja parte das estratégias avaliativas do professor, mas não a única. Fluência de leitura, oralidade, escrita, organização do pensamento em etapas lógicas, habitualmente requeridas para resolução de um problema de matemática, devem ser estimuladas por meio outras estratégias avaliativas que enriquecem o processo formativo, tais como debates, dramatizações, estudos de caso, elaboração de projetos, experimentos, dentre outras.

Os itens de múltipla escolha são elaborados com a finalidade de avaliar uma habilidade e eventualmente uma competência. Esta última finalidade é mais comum no Exame Nacional do Ensino Médio (ENEM). Em avaliações externas de larga escala voltadas para o monitoramento da qualidade dos sistemas de ensino, como no caso do Saeb, os testes cognitivos são compostos por itens que buscam avaliar uma única habilidade (BRASIL, 2010; CAED, 2008).

Desta forma, para cada item o elaborador escolhe uma habilidade prevista na Matriz de Referência e desempenha um processo criativo e técnico para garantir que o item avalie exclusivamente a habilidade escolhida (BRASIL, 2010; VIANNA, 1978). No contexto da avaliação da aprendizagem escolar empreendida pelo professor, a BNCC é o documento que fornece as habilidades a serem avaliadas.

Retomando ao tema do processo de elaboração de itens no formato de múltipla escolha, o elaborador precisa ter familiaridade com a habilidade a qual pretende avaliar, além de compreender o estágio de desenvolvimento cognitivo do público-alvo e a etapa de escolarização

a que o item se destina, para que ele saiba avaliar o nível de dificuldade do item proposto e a adequação da linguagem (VIANNA, 1978).

O item de múltipla escolha tem como elementos: enunciado, que inclui comando para respostas e eventualmente um suporte, e as alternativas. Nos itens de matemática, os enunciados costumam ser um texto claro e objetivo que envolve uma situação-problema, produzido pelo próprio elaborador, que visa estimular o estudante a buscar conhecimentos aprendidos e as habilidades desenvolvidas a fim de resolver o problema proposto (CAED, 2008).

O enunciado pode vir acompanhado de um suporte, que podem ser figuras (gráficos, fotografias, ilustração etc.) ou textos retirados de fontes como jornais e revistas científicas, cuja finalidade é fornecer elementos para a elucidação do contexto e dos dados relevantes para a resolução do problema (CAED, 2008).

O comando para resposta é a pergunta do problema, isto é, esclarece para o educando o que ele precisa responder a respeito das informações apresentadas no enunciado, podendo ser na forma de complementação ou de interrogação (VIANNA, 1978). Em alguns itens, o suporte escolhido dispensa a criação de um texto para o enunciado, fornecendo informações suficientes para a redação do comando. Com ou sem suporte, todas as informações necessárias para a resposta do problema precisam ser apresentadas (BRASIL, 2010; CAED, 2008).

As alternativas do item são compostas por uma única resposta correta, conhecida como gabarito, e por alternativas incorretas que são chamadas de distratores (BRASIL, 2010). Apesar de apresentarem resultados incorretos, os distratores devem ser plausíveis e coerentes com o contexto do problema (VIANNA, 1978). É desejável que alguns distratores são elaborados com base em eventuais erros conceituais ou procedimentais cometido na resolução do problema (CAED, 2008).

A Figura 9 mostra um exemplo de item de matemática e seus elementos. O enunciado do item é um texto conciso que requer um suporte gráfico, neste caso, uma reta numerada com alguns pontos marcados. O comando apresentado é uma sentença a ser completada. Enunciado e suporte fornecem, nesse item, todas as informações necessárias para que o educando possa responder ao comando. Dentre as alternativas apresentadas, há apenas um gabarito e três distratores, sendo que todos os distratores são plausíveis.



Figura 9 - Exemplo de item de matemática com suporte gráfico

**Enunciado**

Observe na reta numérica alguns pontos igualmente espaçados.

**Comando**

O ponto que tem abscissa -6 está representado pela letra

**Gabarito**

(A) L.

(B) M.

(C) Q.

(D) R.

**Distratores**

**Alternativas**

Fonte: Elaborado pelo próprio Autor.

Um outro exemplo de item de múltipla escolha pode ser visto na

Figura 10. O item possui um enunciado sem suporte e um comando para resposta na forma interrogativa. O enunciado apresenta um contexto e informações suficientes para que o educando possa responder ao comando. Esse formato de item é comum nos materiais didáticos de matemática, nas avaliações externas e nas avaliações de sala de aula.

Figura 10 - Exemplo de item de matemática sem suporte gráfico

**Enunciado**

Uma comerciante comprou um produto por R\$ 150,00 e quer vendê-lo com margem de lucro de 20%. Ela sabe que para calcular o preço de venda, ela deve somar o valor do lucro ao preço de compra.

**Comando**

Qual será o preço de venda do produto?

(A) R\$ 165,00.

(B) R\$ 170,00.

**Gabarito** (C) R\$ 180,00.

(D) R\$ 195,00.

**Distratores**

**Alternativas**

Fonte: elaborado com apoio do Bing Chat.

Os itens apresentados Figura 9 e

Figura 10 são típicos na avaliação da aprendizagem, sobretudo nas avaliações externas. Seus formatos são baseados nas orientações propostas em CAEd (2008). Um outro guia de elaboração de itens voltados para avaliações externas conduzidas pelo Inep pode ser visto em Brasil (2010). Nesse guia, os itens possuem três partes: Texto-Base, Enunciado e Alternativas. As partes de um item segundo esse documento podem ser identificadas no exemplo da Figura 11.

Figura 11 - Exemplo de item de matemática com texto base: modelo adotado no ENEM

**Texto-base**

Para uma temporada das corridas de Fórmula 1, a capacidade do tanque de combustível de cada carro passou a ser de 100 Kg de gasolina. uma equipe optou por utilizar uma gasolina com densidade de 750 gramas por litro, iniciando a corrida com o tanque cheio. Na primeira parada de reabastecimento, um carro dessa equipe apresentou um registro em seu computador de bordo acusando o consumo de quatro décimos da gasolina originalmente existente no tanque. Para minimizar o peso desse carro e garantir o término da corrida, a equipe de apoio reabasteceu o carro com a terça parte do que restou no tanque na chegada ao reabastecimento.

Disponível em: [www.superdanilor1page.com.br](http://www.superdanilor1page.com.br). Acesso em: 6 jul. 2015 (adaptado).

**Enunciado**

A quantidade de gasolina utilizada, em litro, no reabastecimento foi

**Gabarito**

(A)  $\frac{20}{0,075}$

(B)  $\frac{20}{0,75}$

(C)  $\frac{20}{7,5}$

(D)  $20 \times 0,075$

(E)  $20 \times 0,75$

**Alternativas**

Distratores

Fonte: Inep. ENEM 2017.

Uma vez que as partes de um item foram apresentadas, é importante esclarecer como esses elementos devem ser dispostos visualmente para evitar que o educando seja induzido ao

erro, seja por falta de informações ou clareza, ou pela criação e disposição inadequada do enunciado e das alternativas.

Nos dois itens usados como exemplo (Figura 9 e

Figura 10), as informações do problema são fornecidas antes do comando para resposta. Após ler o comando para resposta, espera-se que o respondente seja levado de volta para as informações do problema para que trace uma estratégia de resolução e, somente após alcançar uma possível resposta, é que ele recorre às alternativas. As alternativas exibidas em ambos os itens apresentam simetria, isto é, aproximadamente a mesma extensão, e estão ordenadas de forma lógica (progressão textual ou ordem alfabética). Os distratores de ambos os itens são plausíveis, pois possui estrutura semelhante à alternativa correta. Alguns dos distratores inclusive apresentam eventuais erros que os respondentes podem cometer.

Por outro lado, no exemplo da Figura 12, o item apresenta disposição indesejada de suas partes, além de que mobiliza tão somente a memorização da definição de semelhança de triângulos. Neste item, após ler o enunciado e o comando para resposta, o respondente é levado diretamente às alternativas, sem precisar executar nenhum procedimento, analisar ou sistematizar os dados do problema. Ao chegar às alternativas, é fácil notar que a alternativa A apresenta uma extensão maior do que as demais, dando a impressão de ser mais completa e podendo induzir o sujeito ao gabarito por exclusão. Esse é um efeito indesejado, pois espera-se que o item dê oportunidade ao sujeito de refletir, analisar e agir de forma consciente sobre a situação proposta.

Figura 12 - Exemplo de item com problemas de assimetria das alternativas

Qual das condições abaixo é necessária e suficiente para que dois triângulos sejam considerados semelhantes?

**Enunciado + comando**

(A) A condição necessária e suficiente para que dois triângulos sejam considerados semelhantes é que seus ângulos internos sejam congruentes e que seus lados correspondentes sejam proporcionais. **Gabarito**

(B) Os triângulos são considerados semelhantes apenas se tiverem a mesma área.

(C) Apenas os lados correspondentes dos triângulos precisam ser proporcionais.

(D) Os triângulos são considerados semelhantes apenas se tiverem o mesmo perímetro.

(E) Apenas os ângulos internos dos triângulos precisam ser congruentes. **Distratores**

Fonte: Plataforma *Teachy*.

Em conformidade com Brasil (2010) e CAED (2008), ao elaborar um item de múltipla escolha para a avaliação da aprendizagem é importante levar em consideração as seguintes instruções.

- 1) Utilizar uma linguagem apropriada ao nível escolar.
- 2) Se certificar de alinhar o comando à habilidade que pretende avaliar.
- 3) Se certificar de que o item avalia uma única habilidade e que ele não apresenta “pegadinhas”, isto é, elementos que induzam o respondente ao erro.
- 4) Elaborar itens com pontuação correta. Por exemplo, comando a ser completado tem alternativas minúsculas com ponto, ao passo que comandos na forma interrogativa têm alternativas iniciais maiúsculas.
- 5) Evitar itens cujas respostas dependam de outros itens, nem elaborar itens que meçam único e exclusivamente a capacidade de memorização do estudante.
- 6) O enunciado e as alternativas devem ser formulados de maneira positiva e redigidos de acordo com a norma culta da língua portuguesa.
- 7) Devem ser evitados termos como “sempre”, “nunca”, “todo(a)”, “totalmente”, “absolutamente”, “completamente” e “somente”.
- 8) Cada item deve apresentar apenas um problema, ser elaborado em linguagem clara e objetiva, apresentar pontuação adequada e apenas um gabarito.
- 9) Revisar os itens várias vezes para melhorar sua qualidade antes de aplicá-lo ou enviá-lo à coordenação.

No enunciado, recomenda-se:

- 1) diversificar fontes, priorizando situações do cotidiano do estudante;
- 2) a apresentação clara de todas as informações essenciais para a resposta ao comando;
- 3) não utilizar expressões negativas, como “não é correto afirmar que”;
- 4) evitar enunciados que levem o estudante à resposta sem que seja necessário realizar qualquer procedimento ou análise;
- 5) evitar expressões como “Assinale a alternativa correta”, “Qual das alternativas...”, “A alternativa que indica...”, e similares;
- 6) se certificar de que o enunciado esteja contemplando a habilidade que o item pretende avaliar;
- 7) o texto deve seguir a norma culta da língua e não ser escrito em 1ª pessoa.

Em relação ao suporte, ele pode incluir gráficos, figuras, ilustrações e tabelas. Se esses materiais forem retirados de outras fontes, como livros, revistas ou jornais, é obrigatório apresentar a bibliografia completa de acordo com as normas da Associação Brasileira de Normas Técnicas (ABNT) (BRASIL, 2010). É proibido utilizar materiais que não estejam relacionados ao item em questão. As imagens de gráficos, figuras e tabelas devem ser nítidas e bem-posicionadas (CAED, 2008).

As alternativas incorretas devem ser plausíveis, ou seja, ter semelhanças ou similaridade em relação à alternativa correta (VIANNA, 1978). Evitar alternativas que induzam ao erro, que contenham detalhes irrelevantes ou conteúdos absurdos, que sejam mutuamente excludentes (a menos que o descritor exija), ou que induzam o estudante a acertar o item por exclusão (CAED, 2008). As alternativas devem ser ordenadas de maneira lógica (por progressão textual ou ordem alfabética), ter aproximadamente a mesma extensão, ser redigidas com vocabulário adequado ao público-alvo e apresentar respostas completas. Evitar construir alternativas demasiadamente longas.

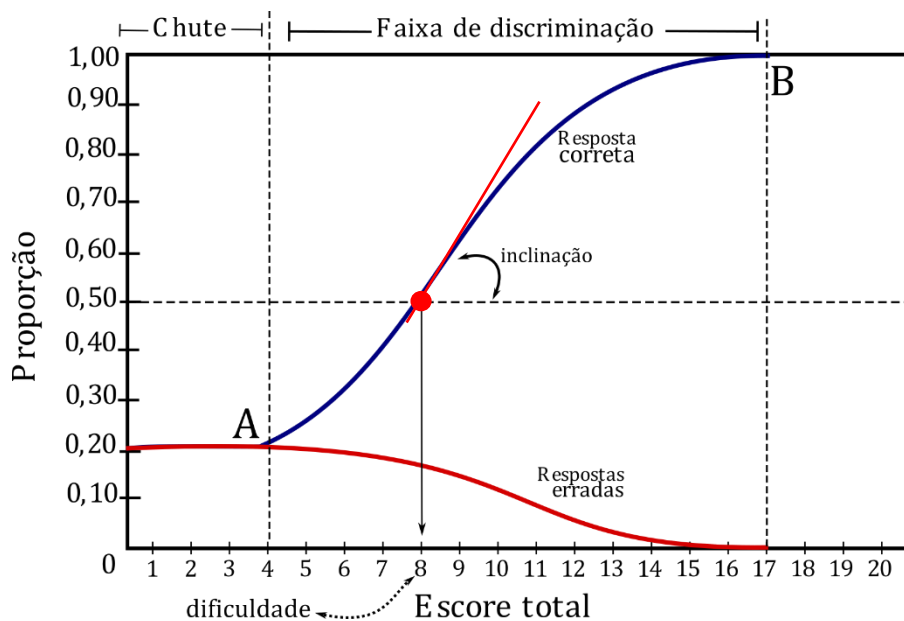
Embora as diretrizes para a elaboração de itens sejam rigorosamente seguidas, é crucial que esses itens também exibam atributos estatísticos, como discriminação, dificuldade e a propensão ao acerto casual. Geralmente, esses parâmetros são fornecidos com o auxílio da Teoria de Resposta ao Item (TRI) durante a etapa de pré-teste, conduzida por especialistas nesse tipo de análise. No entanto, existem abordagens mais simplificadas, associadas à Teoria Clássica dos Testes (TCT), que podem ser empregadas por professores sem formação em psicometria para avaliar a qualidade dos itens.

Caso um professor queira avaliar um conjunto de itens utilizado por ele apresentam índices aceitáveis de discriminação, propensão adequada de acerto casual e os níveis de dificuldade de cada item, ele pode recorrer a recursos gráficos de fácil análise. Uma proposta de análise de itens, conhecida como Análise Gráfica de Itens (AGI), apresenta de forma visual informações sobre a qualidade do item.

A AGI é uma representação na qual o eixo das abscissas exibe os escores dos indivíduos, considerados como indicadores de proficiência na TCT. No eixo das ordenadas, é apresentada a proporção de respostas dadas a cada alternativa para cada grupo de proficiência. Esta representação gráfica pressupõe que a probabilidade de um indivíduo responder corretamente a um item aumenta com a sua proficiência. Portanto, um item de qualidade é caracterizado pelo aumento proporcional da resposta correta à medida que o escore dos indivíduos aumenta (BATENBURG; LAROS, 2002).

A Figura 13 mostra um item teórico de um teste de 20 itens, cada um com 5 alternativas para resposta, sendo apenas uma correta. A curva azul indica a proporção de respostas corretas em função do escore total dos participantes, enquanto a curva vermelha decorre da proporção de respostas erradas em função do escore total. Nota-se que à medida que aumenta o escore total aumenta também a proporção de sujeitos com bons escores que optam pela resposta correta – propriedade que se espera de um bom item.

Figura 13 - Análise Gráfica de Item: elementos



Fonte: Silva (2019).

Para o teste em questão, a faixa de respostas aleatórias, ou "chute", é compreendida no intervalo  $[0, 4]$ . Isso se deve ao fato de que um indivíduo que opta por responder todos os itens de forma aleatória tem uma probabilidade de acerto de  $1/5$  para cada item. Considerando que o teste é composto por 20 itens, a expectativa é que o número de acertos seja 4.

Para escores superiores a 4, observa-se um aumento na proporção de acertos e uma diminuição na proporção de respostas erradas. Isso indica que, a partir desse escore, o item começa a discriminar os indivíduos com base em sua habilidade (LAROS, 2017).

Após a curva (representada em azul) atingir seu ponto máximo, que neste caso ocorre quando o escore é igual a 17, o item deixa de ser capaz de discriminar, pois todas as respostas já foram avaliadas. Portanto, para o item em análise, a faixa de discriminação compreende o intervalo  $(0, 17]$  (LAROS, 2017).

Uma maneira visual de avaliar a capacidade discriminatória de um item envolve a construção de uma linha paralela ao eixo das abscissas que intersecta o eixo das ordenadas em 0,50. A interseção desta linha com a curva das respostas corretas é então avaliada (Figura 13). Um indicativo de alta capacidade discriminatória do item é uma inclinação positiva acentuada neste ponto de interseção. Inclinações negativas podem sugerir que o item tem dificuldade em discriminar entre sujeitos proficientes e não proficientes. Ao examinar a abscissa do ponto em análise, identifica-se o índice de dificuldade do item. No exemplo em questão, a dificuldade do item é 8, o que sugere que sujeitos com escores superiores a 8 podem ter uma probabilidade maior de responder corretamente ao item. Quanto maior esse índice, maior é o nível de dificuldade do item para o público avaliado.

É crucial enfatizar que esta é uma representação teórica. Na prática, pode haver casos em que o intervalo de discriminação seja mais restrito, ou até mesmo onde o item não produza uma discriminação significativa. No entanto, a expectativa geral para essa representação gráfica é que um aumento no escore resulte em um aumento na proporção de respostas corretas.

Outros índices, como a correlação ponto bisserial, o índice de discriminação (D), a dificuldade (determinada pela proporção de acertos) e a consistência interna (alfa de Cronbach), permitem uma análise abrangente dos itens quando suas propriedades psicométricas. Esses índices são amplamente utilizados no contexto da TCT e podem ser calculados com relativa facilidade com o auxílio de planilhas eletrônicas ou pacotes para softwares estatísticos, como o R.

A correlação ponto bisserial é um indicador de discriminação do item baseado na correlação de Pearson para dados dicotomizados, isto é, para quando se atribui 0 para alternativa errada e 1 para alternativa correta (PASQUALI, 2017). A medida em questão é dada por

$$r_{pb} = \frac{\bar{X}_A - \bar{X}_T}{s_T} \sqrt{\frac{p}{q}}$$

em que  $\bar{X}_A$  é a média no teste dos sujeitos que acertam o item,  $\bar{X}_T$  é a média total do teste,  $s_T$  é o desvio padrão do teste,  $p$  é a proporção de sujeitos que acertam o item e  $q = 1 - p$ . O modelo em questão revela que quanto maior for a diferença de média dos escores dos sujeitos que acertam o item e da média total, mais discriminativo é o item. Mas essa medida depende ainda da variabilidade das respostas ao item, indicada pelo desvio padrão e da proporção da proporção de respostas corretas, sendo um número entre -1 e 1. Em geral, medidas de correlação superiores a 0,4 indicam bom poder discriminativo (PASQUALI, 2017).

Uma outra proposta ainda mais simples utilizada para avaliar a discriminação ( $D$ ) é ordenando os sujeitos com base no seu desempenho. Tipicamente, calcula-se a discriminação  $D$  do item fazendo-se a diferença da proporção de respostas corretas dos 27% dos sujeitos com maiores desempenho e dos 27% dos sujeitos com piores desempenho. Uma sugestão de interpretação desse índice de discriminação é proposta por Rabelo (2013), conforme exhibe a Tabela 2.

Tabela 1 - classificação dos itens de acordo com a discriminação na TCT

Valores de discriminação	Classificação
$D < 20$	Item deficiente, que deve ser rejeitado
$20 \leq D < 30$	Item marginal, sujeito a reelaboração
$30 \leq D < 40$	Item bom, mas sujeito a aprimoramento
$D \geq 40$	Item bom

Fonte: Rabelo (2013, p. 136).

No que se refere à consistência interna do instrumento, é comum, em uma abordagem clássica, utilizar o alfa de Cronbach ( $\alpha$ ), dado por

$$\alpha = \frac{n}{n-1} \left( 1 - \frac{\sum s_i^2}{s_T^2} \right)$$

em que  $n$  é o número de itens,  $\sum s_i^2$  é a soma das variâncias dos  $n$  itens e  $s_T^2$  é a variância total dos escores do teste (PASQUALI, 2017). A medida alfa é um número entre 0 e 1. O modelo em questão indica que quanto menor for as variâncias individuais dos itens, maior é a medida alfa e, por conseguinte, maior é a consistência interna do instrumento. Assim, medidas de correlação próximas de 1 indicam que o instrumento possui boa consistência interna.

Nesta Seção, discutiram-se as principais dimensões, conceitos e métodos da avaliação da aprendizagem em matemática no ensino fundamental. Abordaram-se os elementos que compõem um item de múltipla escolha, bem como os critérios de qualidade e as etapas de elaboração e revisão de itens, para que estes avaliem precisamente as competências e habilidades esperadas dos educandos em uma determinada etapa.

Na seção seguinte, serão descritos os procedimentos metodológicos adotados, que incluíram a seleção das ferramentas utilizadas, a definição dos protocolos de interação com os chats inteligentes, a aplicação de um teste com alguns itens gerados e a análise dos resultados obtidos.



## 4 PROCEDIMENTOS METODOLÓGICOS

### 4.1 Classificação da pesquisa

O estudo em questão é de natureza aplicada. Conforme definido por Gil (2008), um estudo aplicado tem como principal objetivo a aplicação prática dos resultados, contribuindo para a resolução de problemas específicos. Em termos de objetivo, a pesquisa adotada é de natureza exploratória. Este tipo de pesquisa tem como objetivo proporcionar uma maior familiaridade com o problema, elucidando elementos que possam auxiliar na sua compreensão e resolução.

A abordagem metodológica adotada é quali-quantitativa, uma junção das metodologias qualitativas e quantitativas. A pesquisa qualitativa busca compreender os fenômenos a partir dos significados que os indivíduos atribuem a eles. Em contrapartida, a pesquisa quantitativa tem como objetivo quantificar os dados e realizar análises estatísticas. Esta abordagem combinada permite uma análise mais completa e robusta dos dados coletados (GIL, 2008).

No que se refere ao método empregado, faz-se uso do método dedutivo que envolve um processo de raciocínio que parte de premissas gerais para chegar a conclusões específicas, pressupondo a razão como o único meio para alcançar o conhecimento verdadeiro (GIL, 2008).

Os procedimentos utilizados na pesquisa são bibliográficos, documental e experimentais. A pesquisa bibliográfica se baseia na análise de materiais já publicados, como livros e artigos científicos. A pesquisa experimental, por sua vez, envolve a manipulação direta das variáveis relacionadas ao objeto de estudo (GIL, 2008).

### 4.2 Ferramentas utilizadas

Nesse estudo recorreu-se a duas ferramentas de IA generativa, a saber, o ChatGPT, que faz uso do modelo GPT-3.5 e o Bing Chat. As duas ferramentas utilizam Grandes Modelos de Linguagens (GLM) para gerar textos, mas não são equivalentes em termo de funcionalidades. O Bing Chat é um serviço gratuito de busca e conversação da *Microsoft*. Ele pode responder a perguntas e criar imagens a partir de descrições. Dependendo do propósito da conversação, o usuário pode optar entre o modo equilibrado, criativo ou preciso. Ela utiliza elementos de *markdown* para formatar as respostas. Durante a conversação, a ferramenta segue regras de segurança e respeito (aspectos éticos).

O Bing Chat tem uma relação de cooperação com o GPT-4, desenvolvido pela OpenAI. Na prática, o modelo por trás do Bing Chat que ajuda a produzir resultados em linguagem natural é o GPT-4, mas cada um tem suas próprias características e funcionalidades. O Bing

Chat se destaca por sua capacidade de realizar buscas na web e apresentar resultados de forma visual e lógica, enquanto GPT-4 se destaca por sua versatilidade na escrita de textos em um número praticamente irrestrito de domínios.

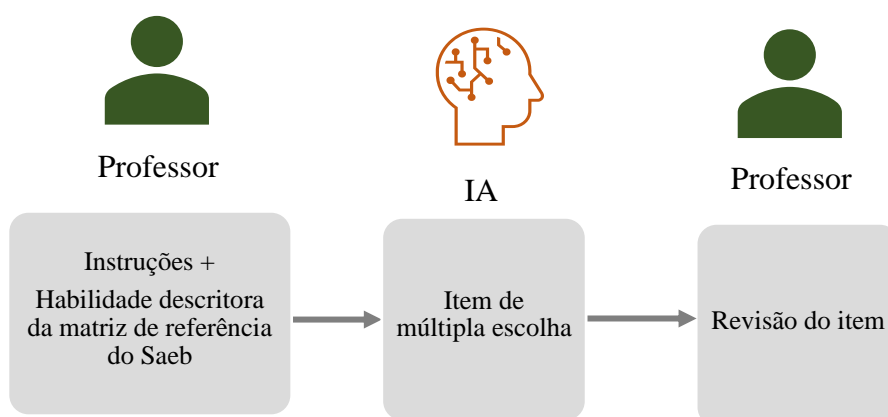
No que se refere ao ChatGPT que faz uso do modelo GPT 3.5, é um chat disponibilizado gratuitamente pela OpenAI. O chat usado, portanto, usa um modelo anterior ao GPT-4, cuja funcionalidade também é gerar textos em linguagem natural sobre diversos domínios. É importante mencionar limitações informadas pela própria ferramenta, tais como a possibilidade de: gerar informações incorretas; produzir instruções prejudiciais ou conteúdos tendenciosos; conhecimento limitado do mundo e eventos após 2021.

### 4.3 Sugestões de elaboração de itens de matemática por meio de IA generativa

#### 4.3.1 Item múltipla escolha com base na Matriz de Referência do Saeb

O método proposto aqui para a criação de itens de múltipla escolha de matemática com o auxílio de uma ferramenta de IA generativa é apresentado na Figura 14. O processo começa com o professor fornecendo à IA instruções claras que definem as partes de um item e o que se espera de cada um deles. Essas instruções também estabelecem um procedimento que a IA deve seguir para realizar a tarefa designada. Depois de receber essas instruções, a IA está preparada para reconhecer as habilidades presentes na Matriz de Referência do Saeb fornecidas pelo professor e gerar uma ou mais questões de múltipla escolha. Em seguida, essas questões são analisadas pelo professor.

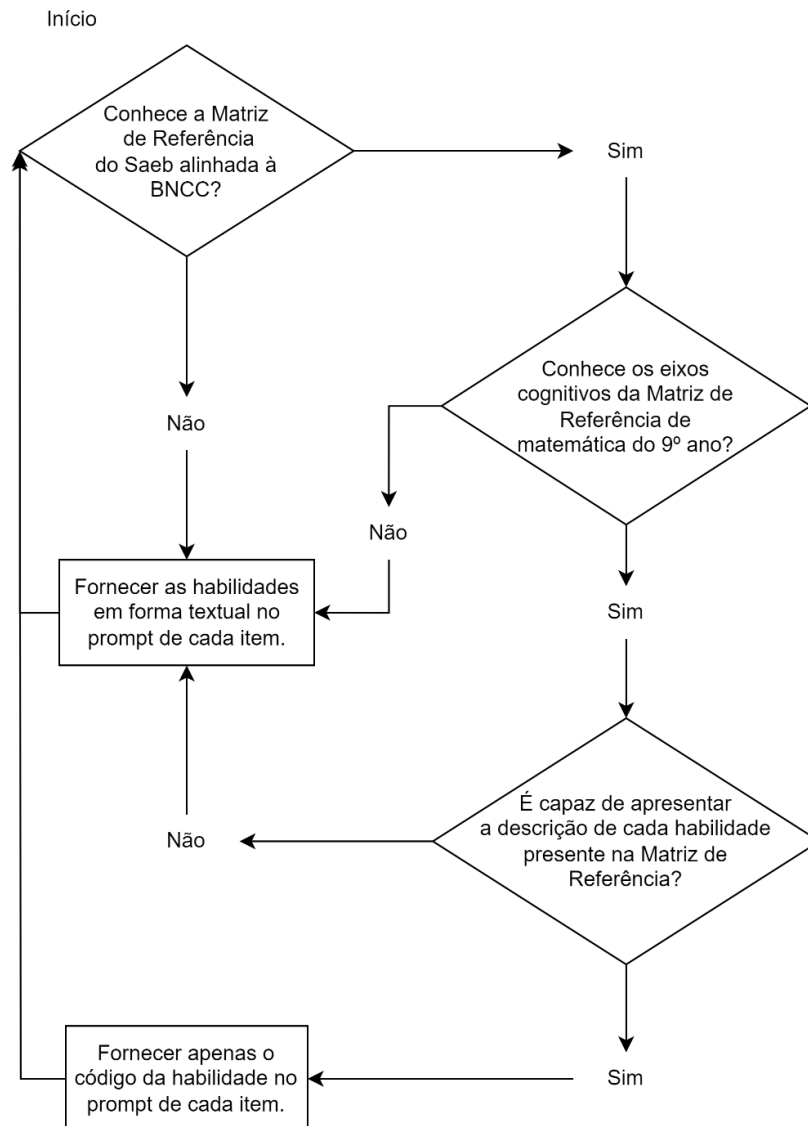
Figura 14 – Processo e agentes envolvidos na elaboração de itens de múltipla escolha buscando atender alguma habilidade prevista na Matriz de Referência do Saeb 2018



Fonte: Elaborado pelo próprio Autor.

Explorando as possibilidades das ferramentas ChatGPT 3.5 e o Bing Chat, foi verificado se elas tinham capacidade de reconhecer a Matriz de Referência do Saeb, bem como seus eixos do conhecimento, eixos cognitivos e suas habilidades. A verificação foi feita segundo o seguinte fluxograma (Figura 15).

Figura 15 – Fluxograma: reconhecimento da Matriz de Referência do Saeb para Elaboração de Itens



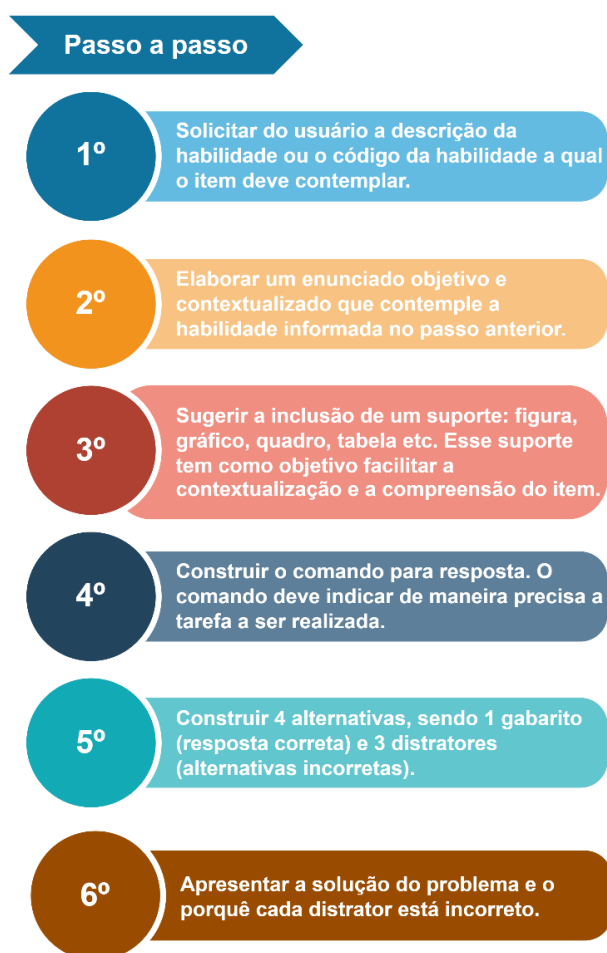
Fonte: Elaborado pelo próprio Autor.

Ao constatar familiaridade com o documento intitulado “Matriz de Referência de Matemática” disponível no site do Inep, foi questionado aos Chats se eles conhecem os eixos cognitivos presentes no documento e a resposta foi positiva. Por fim, ao ser questionado sobre sua capacidade de fornecer a descrição de cada uma das habilidades presentes no documento, as ferramentas apresentaram limitações. Assim, a construção de itens que atenda as habilidades

preconizadas na Matriz de Referência do Saeb por meio da ferramenta ChatGPT, com modelo de linguagem GPT-3.5 e do Bing Chat requer do usuário o fornecimento por escrito da habilidade que o item deve contemplar. É importante mencionar que essa limitação se deu no contexto e data em que esse estudo foi realizado, podendo não existirem mais à medida que as ferramentas forem atualizadas.

Baseado nas instruções apresentadas no Guia de Elaboração de Itens de Matemática do Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF) (CAED, 2008) e no Guia de Elaboração de Itens do Inep (BRASIL, 2010), foi escrito um conjunto de instruções sequenciais que os *chats* devem levar em consideração para elaborar um item de matemática. A sequência de instruções elaborada inicialmente é fornecida na Figura 16.

Figura 16 – Instruções resumidas que a IA deve levar em conta na elaboração dos itens



Fonte: Elaborado pelo próprio Autor.

Esse conjunto de instruções para elaboração do item foi apresentado na forma textual, esclarecendo já no preâmbulo a tarefa que a ferramenta deve realizar. Cada instrução foi

enriquecida com breves recomendações abordadas na Subseção 3.3 deste trabalho. As instruções foram apresentadas em 6 passos, que orienta cada um dos *chats* a apresentar um item de matemática de múltipla escolha que tenha: enunciado, suporte, comando, 1 gabarito, 3 distratores, descrição da solução do item e de seus distratores. As instruções são apresentadas no Quadro 2.

Quadro 2 – Instruções dadas ao ChatGPT e ao Bing Chat para produção de itens de matemática que contemple uma habilidade específica da Matriz de Referência do Saeb

Neste chat, a tarefa é elaborar itens inéditos de múltipla escolha de matemática para estudantes do 9º ano.

Essa tarefa deve seguir rigorosamente o seguinte passo a passo:

1º Passo: Pergunte ao usuário sobre qual habilidade presente na matriz de referência do Saeb de 2018 o item deve contemplar.

2º Passo: Leia a habilidade fornecida pelo usuário e pergunte se o enunciado do item deve ser objetivo (1) ou contextualizado (2). Após obter a resposta do usuário, execute o próximo passo.

3º Passo: Elabore um enunciado do item contemplando a habilidade informada. Ainda nessa etapa, apresente o comando para resposta. O comando deve indicar de maneira precisa a tarefa a ser realizada, evitando negações ou qualquer pegadinha que induza o respondente ao erro. O comando pode ser interrogativo ou apresentar uma frase incompleta.

4º Passo: Caso necessário, sugira a inclusão de um suporte: figura, gráfico, quadro, tabela etc. Descreva o suporte de forma que o usuário possa reproduzi-lo em algum software gráfico. Esse suporte tem como objetivo facilitar a contextualização e a compreensão do item.

5º Passo: Construa 4 alternativas, sendo 1 gabarito (resposta correta) e 3 distratores (alternativas incorretas). Os distratores devem ser plausíveis, mas não podem induzir o sujeito ao erro. Eles devem ser resultados de algum erro operacional ou de raciocínio.

6º Passo: I) apresente a solução/gabarito do item de forma detalhada; II) forneça a explicação detalhada do porquê cada distrator está incorreto e quais são os eventuais erros operacionais ou de raciocínio que o respondente possivelmente cometeu ao marcá-lo.

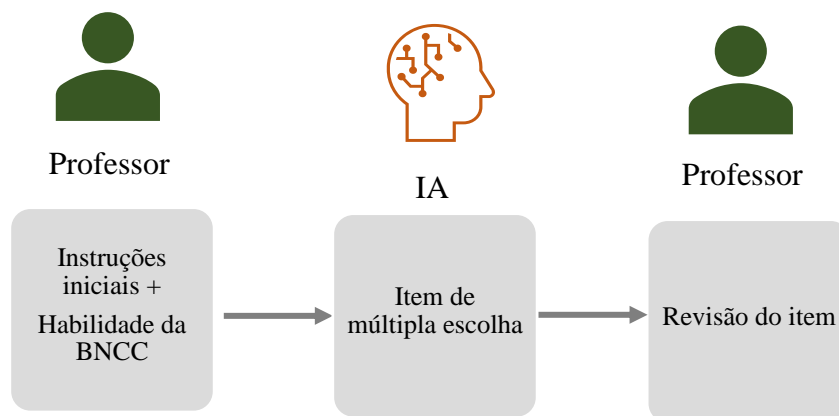
Fonte: Elaborado pelo próprio Autor.

As instruções fornecidas no Quadro 2 foi testada em ambos os chats. Eles foram capazes de compreender o conjunto de tarefas que lhes foram atribuídas. Os itens produzidos adotando esse protocolo são apresentados nos resultados desse trabalho.

#### 4.3.2 Item múltipla escolha com base na BNCC

O processo sugerido para elaboração de itens utilizando habilidades da BNCC é análogo ao processo para elaboração de itens de múltipla escolha com base na Matriz de Referência do Saeb. O que difere é que as instruções precisam esclarecer que o documento de referência que subsidiará a elaboração de itens é a BNCC, cujas competências e habilidades são de amplo conhecimento e estão disponíveis na *internet*.

Figura 17 – Processo e agentes envolvidos na elaboração de itens de múltipla escolha que atendam alguma habilidade da BNCC



Fonte: Elaborado pelo próprio Autor.

Utilizando o mesmo fluxograma da Figura 15 durante a conversação nos *chats*, constatou-se no momento do estudo que o Bing Chat, por ser capaz de fazer busca na *web*, é capaz de reconhecer as habilidades a partir do código alfanumérico. Por exemplo, ao fornecer ao Bing Chat o código EF06MA20, a ferramenta correlacionará este código a uma habilidade específica da BNCC. Esta habilidade, destinada ao 6º ano do Ensino Fundamental, está inserida na componente curricular de Matemática. A habilidade em questão envolve: “Identificar características dos quadriláteros, classificá-los em relação aos lados e ângulos e reconhecer a inclusão e a intersecção de classes entre eles”.

Um conjunto de instruções sugerido nesse trabalho para elaboração de itens com auxílio do Bing Chat é apresentado no Quadro 3. Como mencionado anteriormente, parte das instruções são inteiramente análogas às fornecidas no Quadro 2.

Quadro 3 – Instruções ao Bing Chat para produção de itens de matemática que contemple uma habilidade da BNCC com base no código alfanumérico

Neste chat, o objetivo é desenvolver itens (questões) inéditos de múltipla escolha, focados em matemática e alinhados com as habilidades da Base Nacional Comum Curricular (BNCC). As instruções são as seguintes:

1. Solicite ao usuário que especifique o código ou a descrição da habilidade da BNCC a ser considerada na elaboração do item.
2. Após receber a habilidade do usuário, busque na internet o que a habilidade significa. Siga para o próximo passo.
3. Após entender a que habilidade o código refere, pergunte ao usuário se o item deve ser objetivo (1) ou contextualizado (2). Prossiga com base na resposta do usuário.
4. Desenvolva o item de acordo com a habilidade fornecida. Nesta etapa, inclua um comando claro para a resposta. Este comando deve especificar a tarefa a ser realizada pelo respondente, evitando qualquer tipo de armadilha que possa levar ao erro. O comando pode ser uma pergunta ou uma frase incompleta.
5. Se necessário, sugira a inclusão de um recurso visual, como uma figura, gráfico, quadro ou tabela. Descreva este recurso de forma que o usuário possa reproduzi-lo em um software gráfico. O objetivo deste recurso é facilitar a contextualização e compreensão do item.
6. Elabore quatro alternativas de resposta: uma correta (gabarito) e três incorretas (distratores). Os distratores devem ser plausíveis, mas não devem induzir o respondente ao erro. Eles devem ser resultados de possíveis erros operacionais ou de raciocínio.
7. Finalmente, apresente uma solução detalhada para a questão e explique por que cada distrator está incorreto, identificando os possíveis erros operacionais ou de raciocínio que o respondente pode ter cometido ao escolhê-los.

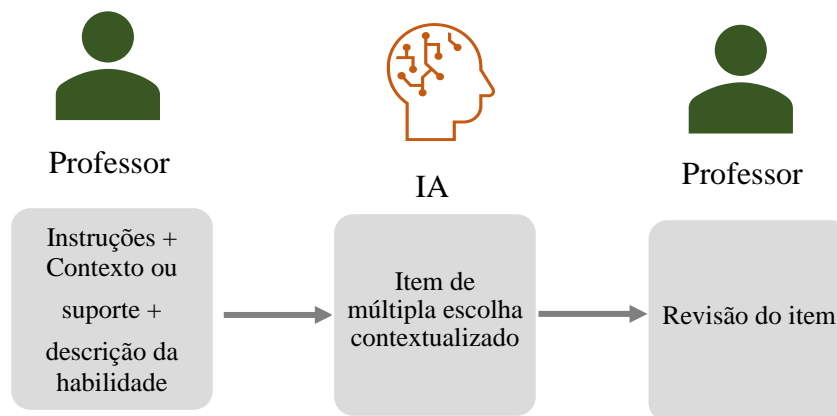
Fonte: Elaborado pelo próprio Autor.

É importante elucidar que as instruções apresentadas no Quadro 3 foi testada somente no Bing Chat. Para que ela seja adequada ao ChatGPT 3.5 a descrição da habilidade precisa ser informada na conversação.

### 4.3.3 Item múltipla escolha a partir de um contexto fornecido

Uma abordagem interessante para a criação de questões de múltipla escolha com o auxílio de IA generativa ocorre quando o professor fornece, além das instruções, um contexto ou suporte para a elaboração do item e a habilidade correspondente (Figura 18).

Figura 18 – Processo e agentes envolvidos na elaboração de itens de múltipla escolha a partir de contexto e habilidade fornecida pelo professor



Fonte: Elaborado pelo próprio Autor.

Com objetivo de testar instruções mais sintéticas, nessa proposta as instruções dadas ao *chat* foram resumidas em 4 passos, mas que esclarecem para a IA o que ela precisa produzir. Essas instruções são exibidas no Quadro 4.

Quadro 4 – Instruções dadas ao Bing Chat para produção de item com base em um texto base

Neste chat, a tarefa é criar itens inéditos de múltipla escolha para matemática. Siga os passos abaixo:

1º Passo: Solicite ao usuário que forneça uma descrição da habilidade que o item deve avaliar.

2º Passo: Após receber a habilidade do usuário, peça a ele que forneça um contexto ou cenário que será a base para o item.

3º Passo: Com base nas informações fornecidas, crie um item de múltipla escolha com quatro alternativas, sendo apenas uma correta. O item deve avaliar de forma precisa a habilidade informada e não ser óbvio. Além disso, o enunciado do item deve ser conciso e esclarecer para o respondente o contexto.

4º Passo: Forneça a solução ou gabarito do item de forma detalhada.

Fonte: Elaborado pelo próprio Autor.

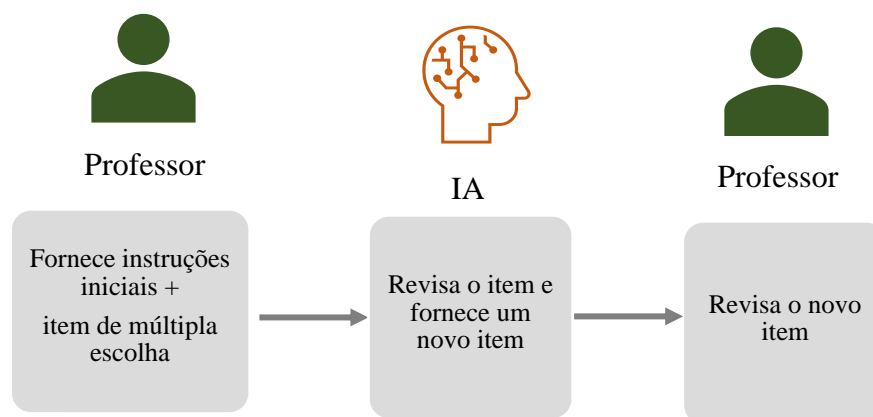


Com a provisão dessas instruções, descrição da habilidade e o contexto ou cenário que servirão como base para a elaboração do item, antecipa-se a produção de itens com características precisas que satisfaçam as expectativas do professor elaborador. Por exemplo, se a competência estiver relacionada ao cálculo da porcentagem de uma quantidade e for do interesse do elaborador contextualizar com o cenário econômico vigente no momento da elaboração, ele pode apresentar à IA um link ou um texto contextual que trate, por exemplo, da Taxa Selic atual.

#### 4.3.4 Aperfeiçoamento de item de múltipla escolha

Nesta proposta, o professor, atuando como elaborador, fornece à IA instruções específicas que devem ser utilizadas para avaliar e aprimorar um item. O elaborador apresenta um item completo, que é então submetido ao processo de aperfeiçoamento pela IA. O item volta para o professor que então revisa o novo item. Essa proposta está ilustrada na Figura 19.

Figura 19 – Processo e agentes envolvidos na elaboração e revisão itens de múltipla escolha: IA como revisora



Fonte: Elaborado pelo próprio Autor.

Em etapas, o professor pode esquematizar esse processo da seguinte forma:

- 1) O professor fornece um item completo. Este item deve incluir o enunciado, alternativas, gabarito e habilidade a qual ele se refere.
- 2) O professor atribui à IA tarefas específicas como: pedir para ela verificar se o texto está de acordo com as normas da língua portuguesa; avaliar se o item atende a habilidade proposta; verificar se o problema tem solução; avaliar se o gabarito é único; verificar se o item é isento de erros conceituais; verificar se alternativas são compatíveis com

problema e se apresentam extensão compatíveis; avaliar a clareza do enunciado, entre outros.

- 3) Com base nas instruções fornecidas, a IA aprimora o item. Este aprimoramento pode envolver a reformulação do enunciado da questão, a alteração das alternativas, a verificação da resposta correta, entre outros.
- 4) A IA retorna o item aperfeiçoado ao professor. O professor, então, revisa este novo item. Se o professor estiver satisfeito com o item, o processo termina aqui. Se não, o professor pode optar por fornecer feedback adicional à IA e o item é submetido novamente à IA para mais aprimoramentos.

As instruções utilizadas nesse trabalho para aprimoramento de itens usando Bing Chat e ChatGPT 3.5 estão disponíveis no Quadro 5.

Quadro 5 – Instruções dadas aos chats para o aprimoramento de itens com alguns critérios de revisão

O objetivo deste chat é aprimorar itens de múltipla escolha de matemática. Para isso, siga o passo a passo:

1º Passo: Solicite ao usuário que forneça um item de múltipla escolha com enunciado, alternativas, gabarito e a habilidade a qual o item se refere.

2º Passo: Verifique se o item:

- Possui todas as informações necessárias para resposta.
- Possui enunciado e comando claro para resposta.
- Está de acordo com as normas padrão da língua portuguesa.
- Atende à habilidade proposta.
- Tem solução única.
- É isento de erros conceituais.
- Apresenta enunciado claro e compreensível.
- Possui 4 alternativas que são compatíveis com o problema e apresentam extensão compatíveis.
- Verifique se o gabarito informado é a resposta correta do problema.

3º Passo: Caso necessário, aprimore o item considerando que ele deve atender à habilidade proposta. Além disso, verifique se o item está alinhado com os conceitos e definições apresentados no Guia de Elaboração e Revisão de Itens do INEP.

4º Passo: Apresente, ao final do item, uma proposta de solução. Certifique-se de que a solução está clara e detalhada, facilitando a compreensão do usuário.

5º Passo: Realize uma revisão final do item para garantir que todas as orientações foram seguidas corretamente e rescreeva o item fazendo todas as correções necessárias, de modo que o item tenha enunciado, comando para resposta e 4 alternativas, com apenas uma correta.

Fonte: Elaborado pelo próprio Autor.

#### 4.4 Aplicação de um teste para avaliação da aprendizagem em matemática

Atendendo ao objetivo de avaliar a qualidade psicométrica de alguns dos itens elaborados ou revisados com auxílio de IA generativa, a presente subseção descreve a aplicação de um teste de matemática que fez parte de um conjunto de intervenções pedagógicas conduzidas pelo pesquisador envolvendo seus educandos do 9º ano do ensino fundamental durante o ano letivo de 2023.

Tendo em vista que no ano de 2023 ocorreram aplicações das avaliações do Saeb em todo Brasil, optou-se por aplicar um instrumento de avaliação da aprendizagem com 20 itens de atendem algumas das habilidades preconizadas na Matriz de Referência dos anos finais do ensino fundamental na componente de matemática. O documento de referência utilizado pelo Saeb para produzir os itens que compõe os testes cognitivos de português e matemática para a edição de 2023 foi a antiga matriz de referência de 2001.

A unidade temática selecionada para a avaliação foi "Números e Operações/Álgebra e Funções", abrangendo conceitos estudados ao longo do ano letivo. O teste em questão foi composto por itens derivados de várias fontes externas, incluindo materiais disponibilizados pelo Inep, Caed e outros extraídos de Assunção, Cavalcanti e Freitas (2019), totalizando 9 itens. Além desses, 11 itens foram criados ou aprimorados com o auxílio do Bing Chat e do ChatGPT, seguindo as orientações propostas na Subseção 4.3.

O teste foi aplicado a um público-alvo de 61 estudantes do 9º ano do ensino fundamental de uma escola pública localizada no município de Terra Santa, Pará. As aplicações ocorreram durante o mês de outubro de 2023, com duração de 1,5 horas. Foi garantido aos estudantes que suas respostas seriam utilizadas anonimamente em um estudo e que isso não acarretaria nenhum prejuízo de qualquer natureza para eles.

Junto com o caderno de questões, os participantes receberam uma folha de resposta que foi corrigida automaticamente com o auxílio de um smartphone e do aplicativo "Prova Fácil". Cada folha de resposta foi digitalizada e as respostas foram armazenadas em um banco de dados, que contém as respostas de todos os participantes avaliados. A ferramenta permite a exportação desse banco de dados no formato CSV, facilitando sua análise por meio de planilhas eletrônicas ou RStudio. Embora a ferramenta não seja gratuita, existe um plano de baixo custo disponível para professores.

## 5 CONSTRUÇÃO DE ITENS POR MEIO DE IA GENERATIVA

Nesse estudo exploratório é proposto um roteiro de conversação que permita a elaboração e revisão de itens nas ferramentas Bing Chat e ChatGPT (linguagem GPT-3.5) que atendam os critérios vistos na Subseção 3.3. Os resultados aqui apresentados seguem rigorosamente as sugestões apresentadas na Subseção 4.3.

### 5.1 Elaboração de itens no ChatGPT

#### 5.1.1 Itens construído com base em uma habilidade fornecida: ChatGPT

Os itens exibidos a seguir foram produzidos com base nas instruções apresentadas no Quadro 2. Após fornecer esse conjunto de instruções ao Chat, sua resposta foi a seguinte: “Vamos começar a elaborar os itens de múltipla escolha de matemática para estudantes do 9º ano, seguindo o passo a passo fornecido. 1º Passo: Qual habilidade presente na matriz de referência do Saeb de 2018 o item deve contemplar?”

O Chat realizou o 1º Passo solicitando ao usuário que forneça uma habilidade da Matriz de Referência do Saeb de 2018. A primeira habilidade fornecida foi: (9A1.1) Resolver uma equação polinomial de 1º grau. Essa habilidade é do eixo do conhecimento “Álgebra” e eixo cognitivo “Compreender e aplicar conceitos e procedimentos”. Neste caso, um item adequado para avaliar essa habilidade deve apresentar enunciado e comando objetivo, que mobilize procedimentos próprios da tarefa “resolver equação polinomial do 1º grau”. Portanto, espera-se que o item apresente um enunciado com uma equação do 1º grau, um comando que pergunte objetivamente sua solução e 4 alternativas, dentre as quais apenas uma delas responde corretamente ao comando.

Após fazer a leitura do da habilidade fornecida, a resposta da ferramenta foi a seguinte: “Ótimo! Vamos trabalhar com a habilidade 9A1.1, que envolve a resolução de equações polinomiais de 1º grau. 2º Passo: O enunciado do item deve ser objetivo ou contextualizado?” (ChatGPT, 2023).

Novamente, a ferramenta cumpriu rigorosamente a sequência de instruções dadas, questionando ao usuário se o comando deve ser objetivo (1) ou se deve ser contextualizado (2). Como a habilidade a ser avaliada por meio do item mobiliza apenas procedimentos, foi sugerido que o enunciado seja objetivo, em oposição a enunciados contextualizados – comuns em itens que mobilizam a aplicação de conceitos, resolução de problemas e inferências. Assim, o item produzido pela ferramenta após as instruções é apresentado na Figura 20.

Figura 20 – Item produzido para a habilidade 9A1.1: falhas nas justificativas dos distratores

**Habilidade:** 9A1.1 - Resolver uma equação polinomial de 1º grau.  
**Enunciado:** Resolva a seguinte equação:  $3x + 5 = 17$ .  
**Comando para resposta:** Qual é o valor de  $x$  que satisfaz a equação acima?  
**Suporte:** Não é necessário incluir um suporte neste item.  
 Alternativas:

(A)  $x = 6$   
 (B)  $x = 4$  (Gabarito)  
 (C)  $x = 2$   
 (D)  $x = 8$

**Solução/gabarito:**  
 Para resolver a equação  $3x + 5 = 17$ , devemos isolar o termo que contém a incógnita ( $x$ ). Começamos subtraindo 5 de ambos os lados da equação, o que resulta em  $3x = 12$ . Em seguida, dividimos ambos os lados por 3, obtendo  $x = 4$ .  
**Gabarito: (B)  $x = 4$ .**

**Distratores**

Alternativa A:  $x = 6$   
 Neste caso, o estudante pode ter somado 5 a 17 incorretamente, obtendo 22 ao invés de 12. Em seguida, dividindo 22 por 3, obtém 7 e não 6 como resultado.

Alternativa C:  $x = 2$   
 Aqui, o estudante pode ter subtraído 5 de 17 erroneamente, obtendo 12 ao invés de 22. Em seguida, dividindo 12 por 3, obtém 4 ao invés de 2 como resposta final.

Alternativa D:  $x = 8$   
 Neste caso, o estudante pode ter cometido o mesmo erro da alternativa A, somando 5 a 17 de forma incorreta, obtendo 22 ao invés de 12. Em seguida, dividindo 22 por 3, obtém 7 e não 8 como resposta final.

Fonte: Elaborado pelo próprio Autor com auxílio do ChatGPT.

O resultado produzido pela ferramenta demonstrou ser adequado até a fase de apresentação detalhada da solução. No entanto, foram identificadas falhas nas justificativas dos distratores. A primeira limitação observada está relacionada à execução da tarefa: fornecer uma explicação detalhada do porquê cada distrator está incorreto e quais são os possíveis erros operacionais ou de raciocínio que podem ser cometidos ao selecioná-lo. Embora a ferramenta esteja adequadamente treinada para apresentar a resposta correta do item, ela não demonstrou eficácia na justificativa de possíveis erros cometidos durante a resolução do item. Portanto, é crucial que o usuário realize uma supervisão cuidadosa dos resultados e reporte especificamente cada erro observado, permitindo assim que o item seja revisado e aprimorado.

Um outro exemplo de item produzido é exibido na Figura 21. O Item foi elaborado para avaliar a habilidade: (9A1.8) Associar uma das representações de uma função afim ou quadrática a outra de suas representações (tabular, algébrica, gráfica) OU Associar uma situação que envolva função afim ou quadrática a uma das suas representações (tabular, algébrica, gráfica). Para esse item, foi sugerido um enunciado contextualizado.

Figura 21 – Item produzido para avaliar a habilidade 9A1.8

**Habilidade:** 9A1.8 - Associar uma das representações de uma função afim ou quadrática a outra de suas representações (tabular, algébrica, gráfica) OU Associar uma situação que envolva função afim ou quadrática a uma das suas representações (tabular, algébrica, gráfica).

**Enunciado:** Um parque de diversões cobra um valor fixo de R\$ 10,00 pela entrada e uma taxa adicional por cada brinquedo utilizado. A tabela abaixo apresenta o número de brinquedos utilizados e o valor total pago por um visitante em um dia.

**Suporte:**

Número de Brinquedos	Valor Total Pago (em reais)
0	10,00
1	15,00
2	20,00
3	25,00
4	30,00

**Comando para resposta:** Associe a tabela acima com a representação algébrica correta da função afim que relaciona o número de brinquedos utilizados ao valor total pago por um visitante.

**Alternativas:**

- (A)  $V(x) = 5x + 5$
- (B)  $V(x) = 10x$
- (C)  $V(x) = 5x + 10$  (Gabarito)
- (D)  $V(x) = 10 + 5x$

**Solução/gabarito:**

Solução: Ao analisar a tabela fornecida, podemos observar que o valor total pago por um visitante está diretamente relacionado ao número de brinquedos utilizados. O valor fixo de R\$ 10,00 representa o custo mínimo de entrada no parque, independentemente do número de brinquedos utilizados. Além disso, a cada brinquedo utilizado, é cobrado um valor adicional de R\$ 5,00.

Portanto, a função afim que relaciona o número de brinquedos utilizados ao valor total pago é representada pela expressão  $V(x) = 5x + 10$ , onde  $V(x)$  é o valor total pago e  $x$  é o número de brinquedos utilizados.

**Distratores:**

- Alternativa A:  $V(x) = 5x + 5$

Nesta alternativa, o estudante adicionou R\$ 5,00 ao número de brinquedos utilizados, mas não considerou o valor fixo de R\$ 10,00. Portanto, essa representação não está correta.

- Alternativa B:  $V(x) = 10x$

Aqui, o estudante multiplicou o número de brinquedos utilizados por R\$ 10,00, mas não considerou o valor adicional de R\$ 5,00 por brinquedo utilizado. Essa representação não está de acordo com a situação descrita na tabela.

- Alternativa C:  $V(x) = 5x + 10$  (Gabarito)

Resposta correta. Essa representação está correta, pois leva em conta o valor fixo de R\$ 10,00 e adiciona R\$ 5,00 para cada brinquedo utilizado.

- Alternativa D:  $V(x) = 10 + 5x$

Nesta alternativa, o estudante inverteu a ordem dos termos da função afim correta. O valor fixo de R\$ 10,00 deve ser adicionado ao resultado da multiplicação do número de brinquedos utilizados por R\$ 5,00. Portanto, essa representação não está correta.

Fonte: Elaborado pelo próprio Autor com auxílio do ChatGPT.

O item (Figura 21) apresenta um enunciado contextualizado, suporte compatível e um comando que deixa claro os conceitos e os procedimentos que o estudante precisa mobilizar para fornecer uma resposta correta. No entanto, nota-se um outro tipo de erro. Desta vez, duas das alternativas respondem corretamente ao comando, a saber, as alternativas C e D. O gabarito C foi devidamente justificado, ao passo que a justificativa que torna a alternativa D um distrator ignora a propriedade comutativa dos números reais e, portanto, produz um erro conceitual crasso. No que se refere aos distratores, as justificativas são plausíveis, exceto na alternativa D. Um erro comum na resolução desse tipo de problema se dá ao confundir taxa de variação com valor fixo. Um distrator que poderia substituir o que foi apresentado erroneamente seria (D)  $V(x) = 5 + 10x$ .

Prosseguindo com a elaboração de itens de matemática atendendo as habilidades preconizadas na BNCC, foi fornecido para o ChatGPT as instruções do Quadro 3. A ferramenta respondeu às instruções com a seguinte mensagem: “Entendido, estou pronto para começar. Por favor, forneça o código da habilidade da BNCC que você gostaria de considerar na elaboração do item” (ChatGPT, 2023). A habilidade escolhida para um primeiro item foi: (EF09MA05) Resolver e elaborar problemas que envolvam porcentagens, com a ideia de aplicação de percentuais sucessivos e a determinação das taxas percentuais, preferencialmente com o uso de tecnologias digitais, no contexto da educação financeira (BRASIL, 2018).

Após ler a habilidade proposta, o Chat respondeu:

Ótimo, a habilidade da BNCC fornecida é (EF09MA05) Resolver e elaborar problemas que envolvam porcentagens, com a ideia de aplicação de percentuais sucessivos e a determinação das taxas percentuais, preferencialmente com o uso de tecnologias digitais, no contexto da educação financeira.  
Agora, gostaria de saber se você prefere um item objetivo (1) ou contextualizado (2). Por favor, indique sua escolha para que eu possa continuar com o desenvolvimento do item. (ChatGPT, 2023).

Foi solicitado que o item fosse contextualizado (2). O item produzido no ChatGPT é exibido na Este item é composto por um enunciado, um comando e quatro alternativas de resposta. O item foi elaborado de forma a atender rigorosamente à habilidade proposta, exigindo do respondente o conhecimento para calcular acréscimos percentuais sucessivos. Uma proposta de solução foi apresentada, sem erros de cálculo ou conceituais. No entanto, é importante notar que nenhuma das quatro alternativas de resposta contempla a solução proposta pela linguagem GPT-3.5.

Figura 22. Este item é composto por um enunciado, um comando e quatro alternativas de resposta. O item foi elaborado de forma a atender rigorosamente à habilidade proposta, exigindo do respondente o conhecimento para calcular acréscimos percentuais sucessivos. Uma proposta de solução foi apresentada, sem erros de cálculo ou conceituais. No entanto, é importante notar que nenhuma das quatro alternativas de resposta contempla a solução proposta pela linguagem GPT-3.5.

Figura 22 – Item produzido para avaliar a habilidade *EF09MA05*

**Habilidade:** (EF09MA05) Resolver e elaborar problemas que envolvam porcentagens, com a ideia de aplicação de percentuais sucessivos e a determinação das taxas percentuais, preferencialmente com o uso de tecnologias digitais, no contexto da educação financeira.

**Enunciado:** Um grupo de estudantes está aprendendo sobre investimentos em uma aula de educação financeira. Eles investiram R\$ 1.000,00 em uma aplicação financeira que promete um retorno de 15% ao ano. Após o primeiro ano, eles decidem reinvestir todo o montante, incluindo os ganhos, em uma nova aplicação que oferece um retorno anual de 20%.

**Comando:**

Considerando o investimento inicial de R\$ 1.000,00 e os percentuais de retorno mencionados, calcule o montante total após o segundo ano, incluindo os ganhos acumulados.

**Alternativas:**

- (A) R\$ 1.430,00
- (B) R\$ 1.560,00
- (C) R\$ 1.650,00
- (D) R\$ 1.720,00

**Solução:**

Para resolver este problema, primeiro, calcule o montante após o primeiro ano usando a fórmula:

$$\text{Montante} = \text{Principal} \times (1 + \text{Taxa de Juros})$$

Para o primeiro ano, a taxa de juros é 15%, então o montante após o primeiro ano é:

$$\text{Montante} = R\$1.000,00 \times (1 + 0,15) = R\$1.150,00$$

Agora, esse montante é reinvestido a uma taxa de 20% durante o segundo ano:

$$\text{Montante Final} = R\$1.150,00 \times (1 + 0,20) = R\$1.380,00$$

Portanto, a resposta correta é **a) R\$ 1.430,00**.

**Explicação das Alternativas Incorretas:**

- **b) R\$ 1.560,00:** Este valor não é correto. Pode ter ocorrido um erro ao calcular os ganhos acumulados.
- **c) R\$ 1.650,00:** Esta opção não está correta. Pode ter havido uma confusão na aplicação das taxas de juros ou no cálculo dos ganhos acumulados.
- **d) R\$ 1.720,00:** Esta resposta não é a correta. Parece que houve um erro nos cálculos, possivelmente ao multiplicar incorretamente as taxas de juros pelos montantes.

Fonte: Elaborado pelo próprio Autor com auxílio do ChatGPT.

Conversa original disponível em: <https://curtlink.com/LycL>



Embora a linguagem GPT-3.5 seja capaz de compreender a tarefa proposta, ela apresenta algumas limitações na elaboração de questões matemáticas simples com base em uma habilidade textual fornecida, como as que foram apresentadas anteriormente. Essas limitações incluem a falta de precisão na resolução de problemas ou a incapacidade de lidar com conceitos matemáticos mais complexos. É importante lembrar que, apesar dessas limitações, a linguagem GPT-3.5 continua sendo uma ferramenta que fornece bons esboços de itens, a partir da descrição de uma habilidade fornecida e de instruções como as que foram apresentadas nos Quadros 1 e 2.

#### 5.1.2 Itens construído com base em uma habilidade e um contexto fornecido: ChatGPT

Uma outra forma de elaborar itens mais direcionados para o objetivo do professor elaborador é fornecendo, além da descrição da habilidade, o contexto ou cenário que a ferramenta deve levar em consideração ao elaborar o item. Uma sugestão de orientação ao chat nessa perspectiva é exibida no Quadro 4.

Procedendo do modo como foi proposto no Quadro 4, a ferramenta atendeu as orientações, respondendo: “Claro, vou guiá-lo através do processo de criação do item de múltipla escolha de matemática. Por favor, forneça a habilidade que o item deve avaliar no primeiro passo” (ChatGPT, 2023). A habilidade informada foi “(EF09MA09) Compreender os processos de fatoração de expressões algébricas, com base em suas relações com os produtos notáveis, para resolver e elaborar problemas que possam ser representados por equações polinomiais do 2º grau” (BRASIL, 2018a).

A ferramenta solicitou em seguida o contexto ou cenário sobre o qual o item deve tratar. O contexto fornecido foi o seguinte: “Cercar uma área retangular máxima destinada à produção de hortaliças em uma determinada escola, utilizando 20 metros de tela para delimitação”. O item produzido com essas instruções é apresentado na Figura 23.

Figura 23 – Item produzido para avaliar a habilidade *EF09MA09*

**Habilidade:** (EF09MA09) Compreender os processos de fatoração de expressões algébricas, com base em suas relações com os produtos notáveis, para resolver e elaborar problemas que possam ser representados por equações polinomiais do 2º grau.

**Contexto proposto pelo professor/elaborador:** Cercar uma área retangular máxima destinada à produção de hortaliças em uma determinada escola, utilizando 20 metros de tela para delimitação.

**Enunciado:** Um agricultor quer cercar uma área retangular para cultivar hortaliças em uma escola. Ele tem 20 metros de tela para delimitação. Qual é a expressão algébrica que representa a área máxima que ele pode cercar?

**Alternativas:**

(A)  $A = 5x(20 - x)$   
 (B)  $A = 10x(10 - x)$   
 (C)  $A = 2x(10 - x)$   
 (D)  $A = 4x(20 - x)$

**Solução:**

Para resolver esse problema, precisamos maximizar a área do retângulo. Se  $x$  representa a largura do retângulo, então o comprimento do retângulo será  $10 - x$  (pois a soma da largura e do comprimento deve ser igual a 10 metros, já que ele só tem 20 metros de tela).

A área do retângulo é dada por  $A = x(10 - x) = 10x - x^2$ .

Para encontrar a expressão que maximiza a área, devemos derivar a função da área em relação a  $x$  e encontrar o valor de  $x$  que torna a derivada igual a zero. No entanto, para este item, vamos focar na escolha correta.

A alternativa correta é **B**  $A = 10x(10 - x)$ .

Fonte: Elaborado pelo próprio Autor com auxílio do ChatGPT.

Conversa original disponível em: <https://curtlink.com/cPwa>

O item elaborado possui enunciado e comando para resposta, alternativas e uma proposta de solução. No entanto, nenhuma das alternativas apresenta a solução correta. Embora o ChatGPT tenha apresentado uma expressão que permite calcular a área em função da largura do retângulo e mencione que o máximo é obtido por meio da primeira derivada igual a zero, ela não expressou essa solução algebricamente. Ademais, do modo como foi proposto, o problema não seria adequado para educandos do 9º ano, uma vez que nessa etapa do ensino os fundamentos de cálculos ainda não são explorados.

Tendo em vista que a habilidade e o contexto propostos para o item exibido na Figura 23 exigem procedimentos e processos mentais mais complexos, optou-se por uma habilidade mais simples e instruções mais genéricas para avaliar como a ferramenta se sai. Desta vez a habilidade proposta foi: “(EF09MA07) Resolver problemas que envolvam a razão entre duas grandezas de espécies diferentes, como velocidade e densidade demográfica” (BRASIL,

2018a). O contexto proposto pelo professor/elaborador foi: “densidade demográfica de um país populoso”. O item produzido é apresentado na Figura 24.

Figura 24 – Item produzido para avaliar a habilidade *EF09MA07*

**Habilidade:** (EF09MA07) Resolver problemas que envolvam a razão entre duas grandezas de espécies diferentes, como velocidade e densidade demográfica.

**Contexto proposto pelo professor/elaborador:** densidade demográfica de um país populoso.

**Enunciado:** A densidade demográfica de um país é a relação entre a população e a área do país. Se a população de um país populoso é de 150 milhões de pessoas, e sua área total é de 1.5 milhão de quilômetros quadrados, qual é a densidade demográfica desse país em pessoas por quilômetro quadrado?

**Alternativas:**

- (A) 100 pessoas por quilômetro quadrado
- (B) 1.5 pessoas por quilômetro quadrado
- (C) 10 pessoas por quilômetro quadrado
- (D) 1000 pessoas por quilômetro quadrado

**Solução:**

A densidade demográfica é calculada dividindo a população pela área. Neste caso, a densidade demográfica é igual a:

$$\text{Densidade Demográfica} = \frac{\text{População}}{\text{Área}} = \frac{150 \text{ milhões}}{1.5 \text{ milhão de km}^2}$$

Portanto, a alternativa correta é **A) 100 pessoas por quilômetro quadrado.**

Fonte: Elaborado pelo próprio Autor com auxílio do ChatGPT. Conversa original disponível em:

<https://curtlink.com/cPwa>

O item ilustrado na Figura 24 é composto por um enunciado, um comando para resposta e quatro alternativas, das quais apenas uma é correta. A solução apresentada está livre de erros conceituais ou de cálculo. Assim, a ferramenta demonstrou ser mais eficaz para avaliar habilidades que requerem procedimentos menos complexos.

### 5.1.3 Aperfeiçoamento e revisão de item com o ChatGPT

Devido à sua notável capacidade de expressão em linguagem natural, o modelo de linguagem GPT-3.5 se destaca como uma ferramenta útil para aprimoramento e revisão de itens conforme instruções específicas. Algumas dessas instruções foram delineadas no Quadro 4 e serão empregadas nesta subseção para demonstrar a aplicabilidade do modelo.

No primeiro teste, habilidade e um item foram fornecidos, contudo, elementos como comando para resposta e alternativas não estavam presentes. O *feedback* fornecido no ChatGPT para o aprimoramento do item em questão pode ser observado na

Figura 25. Entre as sugestões, o modelo solicitou que o comando fosse explicitado e propôs a inclusão de quatro alternativas, sendo apenas uma delas correta. Os distratores

sugeridos são coerentes com o problema e apresentam simetria. Além disso, o ChatGPT apresentou uma proposta de solução coerente para o problema proposto.

Figura 25 – Item revisado e aprimorado pelo ChatGPT: habilidade EF09MA06

**Habilidade:** (EF09MA06) Compreender as funções como relações de dependência unívoca entre duas variáveis e suas representações numérica, algébrica e gráfica e utilizar esse conceito para analisar situações que envolvam relações funcionais entre duas variáveis.

**Item fornecido:** O reservatório de água de um prédio tem capacidade para 5 000 litros e é abastecido utilizando-se uma bomba que retira água de um poço subterrâneo. Essa bomba tem a capacidade de bombear 20 litros de água por minuto. Ela é automaticamente ligada quando restam no reservatório 400 litros de água e desligada após enchê-lo. Pode-se determinar a quantidade de água  $y$ , contida no reservatório, em função do tempo  $t$ , em minutos, que a bomba permanece ligada.

**1º Passo: Verificação das condições gerais**

- Enunciado e comando para resposta: O enunciado fornece um contexto claro e o comando para resposta é implícito, pedindo para determinar a quantidade de água no reservatório em função do tempo. No entanto, para atender aos critérios, **podemos explicitar o comando**, como "Determine a função que descreve a quantidade de água no reservatório em função do tempo."
- Está de acordo com as normas padrão da língua portuguesa: O enunciado está escrito em português correto e claro.
- Atende à habilidade proposta: O item está relacionado à habilidade de compreender funções como relações de dependência unívoca entre duas variáveis, o que é adequado.
- Tem solução única: O item deve ter uma solução única, e, nesse caso, a solução única é a função que descreve a quantidade de água no reservatório em função do tempo.
- É isento de erros conceituais: O enunciado não parece conter erros conceituais.
- Apresenta enunciado claro e compreensível: O enunciado é claro e compreensível.
- Possui 4 alternativas que são compatíveis com o problema e apresentam extensão compatíveis: O item não fornece alternativas, mas isso pode ser ajustado.
- Verifique se o gabarito informado é a resposta correta do problema: Não há gabarito informado.

**2º Passo: Aprimoramento do item**

O item pode ser aprimorado ao incluir alternativas e um gabarito. Vou criar alternativas e um gabarito:

**Item Aprimorado:**  
O reservatório de água de um prédio tem capacidade para 5 000 litros e é abastecido utilizando-se uma bomba que retira água de um poço subterrâneo. Essa bomba tem a capacidade de bombear 20 litros de água por minuto. Ela é automaticamente ligada quando restam no reservatório 400 litros de água e desligada após enchê-lo. Pode-se determinar a quantidade de água  $y$ , contida no reservatório, em função do tempo  $t$ , em minutos, que a bomba permanece ligada.

**Alternativas:**  
A)  $y = 400 + 20t$   
B)  $y = 5000 - 20t$   
C)  $y = 4000 - 20t$   
D)  $y = 5000 + 20t$

**Gabarito:**  
A)  $y = 400 + 20t$

**3º Passo: Aprimoramento de acordo com o Guia de Elaboração e Revisão de Itens do INEP**

O item está alinhado com a habilidade proposta, e as alternativas são compatíveis com o problema.

**4º Passo: Proposta de Solução**

Para determinar a quantidade de água no reservatório em função do tempo, podemos usar a seguinte função:

$$y = 400 + 20t$$

Solução Detalhada:

A quantidade inicial de água no reservatório é 400 litros.  
A bomba bombeia 20 litros por minuto.  
Portanto, a quantidade de água aumenta em 20 litros a cada minuto.  
O tempo  $t$ , em minutos, é o tempo que a bomba permanece ligada.

**5º Passo: Revisão Final**

O item foi aprimorado de acordo com os critérios e inclui enunciado, comando para resposta, 4 alternativas (com apenas uma correta) e uma solução clara.

Fonte: Elaborado pelo próprio Autor com auxílio do ChatGPT. Conversa original disponível em:

<https://curtlink.com/IXPR>

Em um teste subsequente, foi apresentada a descrição de uma habilidade e um item integralmente elaborado, contendo enunciado e comando para resposta. No entanto, este item continha um erro de formulação. O erro em questão residia no fato de que o enunciado não disponibilizava informações suficientes para a resolução do problema proposto, além de apresentar uma situação impraticável. A situação proposta parecia absurda, considerando a menção de um reservatório aquático com dimensões de  $2,4 \cdot 10^6$  metros. O referido item foi criado pelo próprio GPT-3.5. O *feedback* fornecido a respeito do item pode ser visto na Figura 26.

Figura 26 – Item revisado e aprimorado pelo ChatGPT: habilidade EF09MA04

**Habilidade:** (EF09MA04) Resolver e elaborar problemas com números reais, inclusive em notação científica, envolvendo diferentes operações

**Item fornecido:**  
Um cientista está estudando a velocidade da luz em diferentes meios. Ele observa que a velocidade da luz no vácuo é aproximadamente  $3 \cdot 10^8$  m/s. O cientista quer calcular o tempo que a luz leva para percorrer uma distância de  $2,4 \cdot 10^6$  m em um tanque de água. Qual é o tempo em segundos que a luz leva para atravessar esse tanque de água?

(A)  $8 \cdot 10^{-3}$  s  
(B)  $2,4 \cdot 10^{-2}$  s  
(C)  $8 \cdot 10^{-2}$  s  
(D)  $2,4 \cdot 10^{-3}$  s

**1º Passo: Verificação das condições gerais**

- O enunciado fornece um contexto claro e o comando para resposta é explícito, pedindo para calcular o tempo que a luz leva para atravessar o tanque de água. O enunciado está bem formulado.
- Está de acordo com as normas padrão da língua portuguesa: O enunciado está escrito em português correto e claro.
- Atende à habilidade proposta: O item está relacionado à habilidade de resolver problemas com números reais
- Enunciado e comando claro para resposta: e notação científica, o que é adequado.
- Tem solução única: O item tem uma solução única, pois a velocidade da luz no vácuo é uma constante e a distância é dada.
- É isento de erros conceituais: O enunciado não parece conter erros conceituais.
- Apresenta enunciado claro e compreensível: O enunciado é claro e compreensível.
- Possui 4 alternativas que são compatíveis com o problema e apresentam extensão compatível: As alternativas são compatíveis com o problema e estão em notação científica, o que é apropriado para o contexto.
- Verifique se o gabarito informado é a resposta correta do problema: O gabarito não foi informado, mas posso determinar a resposta correta.

**2º Passo: Aprimoramento do item**  
O item parece estar bem formulado. A única melhoria necessária é a inclusão do gabarito.

**Item Aprimorado:** não alterou nenhum elemento do item fornecido.

**3º Passo: Aprimoramento de acordo com o Guia de Elaboração e Revisão de Itens do INEP**  
O item está alinhado com a habilidade proposta, e as alternativas são compatíveis com o problema.

**4º Passo: Proposta de Solução**  
Para calcular o tempo que a luz leva para atravessar o tanque de água, podemos usar a fórmula:

$$\text{Tempo} = \text{Distância} / \text{Velocidade}$$

Onde:

- \* Distância =  $2,4 \times 10^6$  metros
- \* Velocidade da luz no vácuo  $\approx 3 \times 10^8$  m/s

$$\text{Tempo} = (2,4 \times 10^6) / (3 \times 10^8) = 2,4 \times 10^6 / 3 \times 10^8 = (2,4/3) \times (10^6/10^8) = 0,8 \times 10^{-2}$$

segundos

Convertendo para notação científica, obtemos  $8 \times 10^{-3}$  segundos.

**5º Passo: Revisão Final**  
O item foi aprimorado e inclui enunciado, comando para resposta, 4 alternativas (com apenas uma correta) e uma solução clara. O gabarito foi fornecido. O item atende aos critérios estabelecidos.

Fonte: Elaborado pelo próprio Autor com auxílio do ChatGPT. Conversa original disponível em:

<https://curtlink.com/IXPR>

A questão em análise envolve o cálculo do tempo que a luz leva para percorrer uma determinada distância em um meio aquoso. Conceitualmente, seria necessário que o enunciado do item fornecesse a velocidade da luz para esse meio específico, pois a velocidade da luz pode variar dependendo do meio em que se propaga. No vácuo, a velocidade da luz é uma constante universal, exatamente 299 792 458 metros por segundo. No entanto, na água, a velocidade da luz diminui para aproximadamente 225 000 km/s. Esta redução é atribuída ao fenômeno da refração. Apesar disso, a revisão realizada pelo GPT-3.5 indicou que o problema está isento de erros técnicos e conceituais. Neste caso, há uma limitação inerente à ferramenta em prever erros conceituais como o mencionado acima.

Em contrapartida, o modelo de linguagem GPT-3.5 apresenta a capacidade de verificar se o item atende à habilidade descrita, se está conforme a norma padrão da língua portuguesa e se contém elementos como enunciado, comando para resposta, alternativas e gabarito. No entanto, a revisão de aspectos relacionados ao conteúdo requer uma maior atenção do professor/elaborador que pretende aperfeiçoar seus itens com o auxílio da ferramenta.

## **5.2 Elaboração de itens no Bing Chat**

### **5.2.1 Itens construído com base em uma habilidade fornecida: Bing Chat**

O Bing Chat disponibiliza três modos de interação: Criativo, Balanceado e Preciso. Para a elaboração dos itens deste estudo, optou-se pelo modo "Preciso", uma vez que os modos "Criativo" e "Balanceado" não atenderam às expectativas estabelecidas durante a fase de instrução. Uma característica notável desta ferramenta é a capacidade de exportar as conversas no formato DOCX, facilitando a documentação do diálogo. Para uma comunicação eficaz dos resultados, a transcrição da conversa será apresentada nesta subseção, onde o termo "Usuário" representa o pesquisador e "Bing" denota a ferramenta que emprega o GPT-4 para a geração de textos.

Seguindo as etapas apresentadas no fluxograma da Figura 15, constatou-se que o Bing Chat não tem conhecimento das habilidades presentes na Matriz de Referência de Matemática do 9º ano do Saeb, versão de 2018. Apesar de não conhecer essas habilidades e transcrevê-las explicitamente, a ferramenta consegue direcionar o usuário para acessar o documento diretamente no site do Inep.

Nessa proposta, o usuário deve enviar a descrição da habilidade na conversa. A conversa prossegue com o usuário informando à ferramenta todas as instruções presentes no Quadro 2. A Figura 27 apresenta uma conversação do primeiro item apresentado nessa seção.

Figura 27 – Conversação no Bing Chat para a criação de um item que atende à habilidade 9A1.7 da MR2018 de matemática

**Bing:** Claro, eu posso ajudar com isso. Vamos começar com o primeiro passo. Por favor, me informe qual habilidade presente na matriz de referência do Saeb de 2018 você gostaria que o item contemplasse.

**Usuário:** 9A1.7-Resolver uma equação polinomial de 2º grau.

**Bing:** Entendi, a habilidade escolhida foi "9A1.7 - Resolver uma equação polinomial de 2º grau". Agora, para o segundo passo, você gostaria que o enunciado do item fosse objetivo (1) ou contextualizado (2)?

**Usuário:** 1

**Bing:** Perfeito, vamos elaborar um item objetivo que envolve a resolução de uma equação polinomial de 2º grau.

*3º Passo: Enunciado do item e comando para resposta*

Enunciado: Considere a seguinte equação polinomial de 2º grau:  $x^2 - 5x + 6 = 0$

Comando: Resolva a equação dada para encontrar os valores de  $x$ .

*4º Passo: Suporte*

Neste caso, como o item é objetivo e a equação é simples, não é necessário um suporte gráfico.

*5º Passo: Alternativas*

A)  $x = 2$  e  $x = 3$  B)  $x = 1$  e  $x = 6$  C)  $x = -2$  e  $x = -3$  D)  $x = 0$  e  $x = 1$

*6º Passo: Solução/gabarito e explicação dos distratores*

I) A solução da equação  $x^2 - 5x + 6 = 0$  é obtida por fatoração, resultando em  $(x - 2)(x - 3) = 0$ . Portanto, as raízes são  $x = 2$  e  $x = 3$ . Então, o gabarito é a alternativa A.

II) As alternativas B, C e D estão incorretas pelos seguintes motivos:

Alternativa B: Se o aluno escolher essa opção, ele pode ter confundido os coeficientes da equação com as raízes.

Alternativa C: Se o aluno escolher essa opção, ele pode ter calculado as raízes corretas, mas assumiu erroneamente que deveriam ser negativas.

Alternativa D: Se o aluno escolher essa opção, ele pode ter confundido a equação dada com uma equação mais simples, como  $x^2 = 0$  ou  $x^2 - 1 = 0$ .

Fonte: Elaborado pelo próprio Autor com auxílio do Bing Chat.

Na primeira tarefa, o Bing Chat gerou um item que aborda de maneira precisa a habilidade especificada. O item em questão apresenta um comando e quatro alternativas, das quais apenas uma é correta. A resposta selecionada pelo modelo corresponde efetivamente à solução do problema proposto, e as justificativas para os erros potenciais ao selecionar um dos distratores são coerentes e bem fundamentadas, exceto no distrador D, pois

$$x \in \{0, 1\} \Leftrightarrow x^2 - x = 0.$$

Na elaboração de itens destinados à avaliação de habilidades mais fundamentais, como o exemplo acima, a ferramenta demonstrou eficácia. Nas etapas subsequentes, as habilidades abordadas envolverão conceitos mais complexos, como a inferência de uma equação polinomial de segundo grau e a resolução de problemas que podem ser modelados por uma equação do segundo grau.

Na conversa subsequente, o usuário especificou a habilidade "(9A1.6) Inferir uma equação polinomial de 2º grau que modela um problema". No entanto, observou-se que o item proposto apresenta várias limitações. O enunciado não fornece informações suficientes para a resolução do problema, a sugestão de suporte está consideravelmente desvinculada do problema proposto e as alternativas não oferecem uma solução adequada.

Figura 28 – Conversação no Bing Chat para a criação de um item que atende à habilidade 9A1.6 da MR2018 de matemática

**Usuário:** (9A1.6) Inferir uma equação polinomial de 2º grau que modela um problema.  
**Bing:** Entendi, a habilidade escolhida foi "9A1.6 - Inferir uma equação polinomial de 2º grau que modela um problema". Agora, para o segundo passo, você gostaria que o enunciado do item fosse objetivo (1) ou contextualizado (2)?  
**Usuário:** 2  
**Bing:** Perfeito, vamos elaborar um item contextualizado que envolve a inferência de uma equação polinomial de 2º grau que modela um problema.  
*3º Passo: Enunciado do item e comando para resposta*  
 Enunciado: João está jogando um jogo de basquete. Ele percebe que a trajetória da bola pode ser modelada por uma parábola. Quando ele arremessa a bola a uma altura inicial de 2 metros, a bola atinge uma altura máxima de 5 metros e cai na cesta que está a 4 metros de distância.  
 Comando: Com base nas informações fornecidas, inferir a equação polinomial de 2º grau que modela a trajetória da bola.  
*4º Passo: Suporte*  
 Neste caso, um gráfico representando a trajetória da bola pode ser útil. O gráfico seria uma parábola aberta para baixo, começando em (0,2), atingindo o ponto máximo em (2,5) e terminando em (4,0).  
*5º Passo: Alternativas*  
 A)  $y = -x^2 + 4x + 2$   
 B)  $y = -0.5x^2 + 2x + 2$   
 C)  $y = -0.5x^2 + 4x + 2$   
 D)  $y = -x^2 + 2x + 2$   
*6º Passo: Solução/gabarito e explicação dos distratores*  
 I) A solução correta é a alternativa C. A equação da parábola é dada por  $y = -0.5x^2 + 4x + 2$   
 II) As alternativas A, B e D estão incorretas pelos seguintes motivos:  
 Alternativa A: Se o aluno escolher essa opção, ele pode ter calculado corretamente o coeficiente linear, mas errou o coeficiente do termo quadrático.  
 Alternativa B: Se o aluno escolher essa opção, ele pode ter calculado corretamente o coeficiente do termo quadrático, mas errou o coeficiente linear.  
 Alternativa D: Se o aluno escolher essa opção, ele pode ter confundido os coeficientes da equação.

Fonte: Elaborado pelo próprio Autor com auxílio do Bing Chat.

Um item com a proposta mencionada, ou seja, que modela a altura da bola  $y$  em função da distância horizontal  $x$ , necessita fornecer, além da coordenada de lançamento, a coordenada do vértice da parábola e a coordenada do cesto de basquete, ou disponibilizar esses elementos por meio de uma representação gráfica. Ademais, o item seria mais apropriado para avaliar uma habilidade distinta da que foi proposta, especificamente: (9A1.8) "Associar uma das representações de uma função afim ou quadrática a outra de suas representações (tabular, algébrica, gráfica) OU Associar uma situação que envolva função afim ou quadrática a uma das suas representações (tabular, algébrica, gráfica)".



Buscando ir além, com uma habilidade que exigem procedimentos e processos mentais mais complexos, foi atribuído ao Bing Chat a tarefa de criar um item que atenda a habilidade: (9A2.4) Resolver problemas que possam ser representados por equações polinomiais de 2º grau. A conversação a seguir mostra como a ferramenta se saiu nessa tarefa.

Figura 29 – Conversação no Bing Chat para a criação de um item que atende à habilidade 9A2.4 da MR2018 de matemática

**Usuário:** (9A2.4) Resolver problemas que possam ser representados por equações polinomiais de 2º grau.

**Bing:** Entendi, a habilidade escolhida foi "9A2.4 - Resolver problemas que possam ser representados por equações polinomiais de 2º grau". Agora, para o segundo passo, você gostaria que o enunciado do item fosse objetivo (1) ou contextualizado (2)?

**Usuário:** 2

**Bing:** Perfeito, vamos elaborar um item contextualizado que envolve a resolução de problemas que podem ser representados por equações polinomiais de 2º grau.

*3º Passo: Enunciado do item e comando para resposta*

Enunciado: Um fazendeiro está planejando construir um galinheiro retangular em sua fazenda. Ele tem 30 metros de cerca e quer fazer o maior galinheiro possível. No entanto, uma das laterais do galinheiro será contra a parede de seu celeiro, então ele só precisa cercar os outros três lados.

Comando: Determine as dimensões do galinheiro que maximizam a área cercada.

*4º Passo: Suporte*

Neste caso, um esboço do galinheiro ao lado do celeiro pode ser útil. O desenho seria um retângulo com um dos lados mais longos contra a parede do celeiro.

*5º Passo: Alternativas*

A) Largura = 10m, Comprimento = 20m B) Largura = 15m, Comprimento = 15m C) Largura = 5m, Comprimento = 25m D) Largura = 7.5m, Comprimento = 22.5m

*6º Passo: Solução/gabarito e explicação dos distratores*

I) A solução correta é a alternativa A. Se chamarmos a largura de  $x$  e o comprimento de  $y$ , temos as seguintes equações:  $x + 2y = 30$  (pois o comprimento é cercado em ambos os lados) e  $A = xy$  (para a área). Resolvendo para  $y$  na primeira equação e substituindo na segunda, obtemos  $A = x(30 - 2x)$ , que é uma equação quadrática. Maximizando essa equação, obtemos  $x = 10$  e  $y = 20$ .

II) As alternativas B, C e D estão incorretas pelos seguintes motivos:

Alternativa B: Se o aluno escolher essa opção, ele pode ter assumido erroneamente que a área seria maximizada se o galinheiro fosse um quadrado.

Alternativa C: Se o aluno escolher essa opção, ele pode ter calculado incorretamente as dimensões.

Alternativa D: Se o aluno escolher essa opção, ele pode ter feito um erro aritmético ao resolver as equações.

Fonte: Elaborado pelo próprio Autor com auxílio do Bing Chat.

A observação inicial indica que o item em questão satisfaz a habilidade especificada, exigindo a formulação de uma equação de 2º grau. A sugestão de um recurso gráfico é pertinente, pois fornece uma descrição precisa do que o autor precisa criar para tornar o item mais compreensível. O item apresenta quatro alternativas, no entanto, nenhuma delas corresponde à resposta correta. A solução proposta leva à expressão que associa a largura ( $x$ ) à área do retângulo ( $A$ ), que é dado por  $A = x(30 - 2x)$ . Contudo, a conclusão de que a área é máxima para  $x=7,5$  m não é adequadamente estabelecida.

Continuando com o estudo, a utilização do Bing Chat na elaboração de itens de matemática é explorada. Este processo, que segue instruções específicas e uma habilidade definida na BNCC, é semelhante ao descrito anteriormente nesta Subseção. No entanto, as

instruções referenciadas são as apresentadas no Quadro 2. Tais diretrizes orientam o Bing Chat a realizar uma busca na web pela habilidade em questão, um procedimento que tem o potencial de enriquecer significativamente o processo de elaboração dos itens. A seguir, apresenta-se os diálogos entre usuário e o Bing usando as diretrizes.

Figura 30 – Conversação no Bing Chat para a criação de um item que atende à habilidade EF09MA04 da BNCC

**Usuário:** Neste chat, o objetivo é desenvolver itens (questões) inéditos de múltipla escolha, focados em matemática e alinhados com as habilidades da Base Nacional Comum Curricular (BNCC). As instruções são as seguintes:  
Solicite ao usuário que especifique o código ou a descrição da habilidade da BNCC a ser considerada na elaboração do item.  
Após receber a habilidade do usuário, busque na internet o que a habilidade significa. Siga para o próximo passo.  
Após entender a que habilidade o código refere, pergunte ao usuário se o item deve ser objetivo (1) ou contextualizado (2). Prossiga com base na resposta do usuário.  
Desenvolva o item de acordo com a habilidade fornecida. Nesta etapa, inclua um comando claro para a resposta. Este comando deve especificar a tarefa a ser realizada pelo respondente, evitando qualquer tipo de armadilha que possa levar ao erro. O comando pode ser uma pergunta ou uma frase incompleta.  
Se necessário, sugira a inclusão de um recurso visual, como uma figura, gráfico, quadro ou tabela. Descreva este recurso de forma que o usuário possa reproduzi-lo em um software gráfico. O objetivo deste recurso é facilitar a contextualização e compreensão do item.  
Elabore quatro alternativas de resposta: uma correta (gabarito) e três incorretas (distratores). Os distratores devem ser plausíveis, mas não devem induzir o respondente ao erro. Eles devem ser resultados de possíveis erros operacionais ou de raciocínio.  
Finalmente, apresente uma solução detalhada para a questão e explique por que cada distrator está incorreto, identificando os possíveis erros operacionais ou de raciocínio que o respondente pode ter cometido ao escolhê-los.

**Bing:** Claro, estou aqui para ajudar! Por favor, especifique o código ou a descrição da habilidade da BNCC que você gostaria de considerar na elaboração do item.

**Usuário:** (EF09MA04) Resolver e elaborar problemas com números reais, inclusive em notação científica, envolvendo diferentes operações.

**Bing:** A habilidade EF09MA04 da BNCC se refere à capacidade de resolver e elaborar problemas com números reais, inclusive em notação científica, envolvendo diferentes operações. Isso envolve a compreensão e aplicação de notação científica em várias operações matemáticas. Agora, gostaria de saber se você prefere que o item seja objetivo (1) ou contextualizado (2). Por favor, escolha uma das opções.

**Usuário:** 2

**Bing:** Ótimo! Vamos criar um problema contextualizado. Aqui está o problema:  
*Problema:* Um astrônomo está estudando a distância entre a Terra e várias estrelas. Ele descobre que a estrela Betelgeuse está aproximadamente  $6,4 \cdot 10^5$  vezes mais longe da Terra do que o Sol. Sabendo que a distância média do Sol à Terra é de aproximadamente  $1,5 \times 10^8$  km, qual é a distância aproximada da Terra à Betelgeuse em km?  
*Alternativas:*  
A)  $9,6 \times 10^{13}$  km  
B)  $1,0 \times 10^{14}$  km  
C)  $7,2 \times 10^{13}$  km  
D)  $8,4 \times 10^{13}$  km  
*Solução:* Para encontrar a distância da Terra à Betelgeuse, multiplicamos a distância da Terra ao Sol pela relação dada. Isso nos dá  $(6,4 \times 10^5) \cdot (1,5 \times 10^8) = 9,6 \times 10^{13}$  km. Portanto, a alternativa correta é A)  $9,6 \times 10^{13}$  km.  
*Explicação das alternativas incorretas:* B)  $1,0 \times 10^{14}$  km: Este valor é muito alto. Pode ser o resultado de um erro de cálculo. C)  $7,2 \times 10^{13}$  km: Este valor é muito baixo. Pode ser o resultado de um erro de cálculo. D)  $8,4 \times 10^{13}$  km: Este valor é muito baixo. Pode ser o resultado de um erro de cálculo.

Fonte: Elaborado pelo próprio Autor com auxílio do Bing Chat.

O item proposto está em conformidade com a habilidade EF09MA04, pois apresenta uma situação contextualizada que avalia a capacidade do estudante de identificar e realizar a operação necessária para resolver o problema, utilizando a notação científica. A comparação entre a distância da Terra ao Sol e a distância da Terra à Betelgeuse envolve a operação de multiplicação e, devido à magnitude das medidas, requer o domínio das operações entre potências de base 10. No entanto, apesar do problema ser bem elaborado do ponto de vista do conteúdo, um dos dados utilizados no problema não corresponde à realidade, comprometendo a qualidade do item. Considerando que a distância da Terra à Betelgeuse é de aproximadamente  $6,8 \times 10^{15}$  km, a distância da Terra ao Sol cabe aproximadamente  $4,5 \times 10^7$  vezes na distância entre a Terra e Betelgeuse. Portanto, para garantir a precisão do item proposto, é essencial que se baseie em dados reais.

Uma habilidade adicional sugerida para a elaboração de um item pelo Chat é: (EF09MA14) Resolver e elaborar problemas que apliquem o teorema de Pitágoras ou as relações de proporcionalidade envolvendo retas paralelas cortadas por secantes. Como essa habilidade pertence à unidade temática de Geometria, antecipa-se que o processo de criação possa apresentar maiores desafios para o Bing Chat. No entanto, esses desafios são importantes para aprimorar a capacidade do Chat de gerar conteúdo relevante e educacionalmente úteis. Para o item em questão, foi sugerido que apresentasse enunciado “objetivo”.

Figura 31 – Conversação no Bing Chat para a criação de um item que atende à habilidade EF09MA14 da BNCC

**Usuário:** (EF09MA14) Resolver e elaborar problemas de aplicação do teorema de Pitágoras ou das relações de proporcionalidade envolvendo retas paralelas cortadas por secantes.

**Bing:** A habilidade EF09MA14 da BNCC se refere à capacidade de resolver e elaborar problemas que aplicam o teorema de Pitágoras ou as relações de proporcionalidade envolvendo retas paralelas cortadas por secantes. Isso envolve a compreensão e aplicação do teorema de Pitágoras em várias situações, bem como a compreensão das relações de proporcionalidade quando retas paralelas são cortadas por secantes. Agora, gostaria de saber se você prefere que o item seja objetivo (1) ou contextualizado (2). Por favor, escolha uma das opções.

**Usuário:** 1

**Bing:** Ótimo! Vamos criar um problema objetivo. Aqui está o problema:  
*Problema:* Em um triângulo retângulo, os comprimentos dos catetos são 6 cm e 8 cm. Qual é o comprimento da hipotenusa?  
*Alternativas:* A) 10 cm B) 14 cm C) 12 cm D) 16 cm  
*Solução:* Para encontrar o comprimento da hipotenusa, aplicamos o teorema de Pitágoras, que afirma que em um triângulo retângulo, o quadrado da hipotenusa é igual à soma dos quadrados dos catetos. Isso nos dá  $\sqrt{6^2 + 8^2} = \sqrt{36 + 64} = \sqrt{100} = 10$  cm. Portanto, a alternativa correta é A) 10 cm.  
*Explicação das alternativas incorretas:* B) 14 cm: Este valor é muito alto. Pode ser o resultado de um erro de cálculo. C) 12 cm: Este valor é muito alto. Pode ser o resultado de um erro de cálculo. D) 16 cm: Este valor é muito alto. Pode ser o resultado de um erro de cálculo.

Fonte: Elaborado pelo próprio Autor com auxílio do Bing Chat.

Embora o item seja de natureza geométrica, o Chat optou por não sugerir a criação de um suporte gráfico. Em vez disso, forneceu as medidas necessárias para o cálculo da hipotenusa e gerou quatro alternativas, das quais apenas uma é correta. A solução proposta está em conformidade com o gabarito indicado pelo Chat. Este método demonstra a capacidade do Chat de produzir itens simples de geometria, mesmo sem o auxílio de representações gráficas.

### 5.2.2 Itens construído com base em uma habilidade e um contexto fornecido: Bing Chat

As instruções apresentadas no Quadro 3 foram comunicadas no chat, o qual solicitou do usuário que ele forneça a descrição de uma habilidade e um contexto. Essas informações servirão como base para a criação do item. O diálogo subsequente ilustra o processo de criação de um item através do Bing Chat.

Figura 32 – Conversação no Bing Chat para a criação de um item que atende à habilidade EF09MA06 da BNCC

**Usuário:** Neste chat, a tarefa é criar itens inéditos de múltipla escolha para matemática. Siga os passos abaixo:

1º Passo: Solicite ao usuário que forneça uma descrição da habilidade que o item deve avaliar. 2º Passo: Após receber a habilidade do usuário, peça a ele que forneça um contexto ou cenário que será a base para o item. 3º Passo: Com base nas informações fornecidas, crie um item de múltipla escolha com quatro alternativas, sendo apenas uma correta. O item deve avaliar de forma precisa a habilidade informada e não ser óbvio. Além disso, o enunciado do item deve ser conciso e esclarecer para o respondente o contexto. 4º Passo: Forneça a solução ou gabarito do item de forma detalhada.

**Bing:** Claro, eu posso ajudar com isso. Vamos começar. 1º Passo: Por favor, forneça uma descrição da habilidade que o item deve avaliar. Por exemplo, a habilidade pode ser "resolver equações quadráticas" ou "interpretar gráficos de funções".

**Usuário:** (EF09MA06) Compreender as funções como relações de dependência unívoca entre duas variáveis e suas representações numérica, algébrica e gráfica e utilizar esse conceito para analisar situações que envolvam relações funcionais entre duas variáveis.

**Bing:** Entendi. Agora, para o 2º passo, por favor, forneça um contexto ou cenário que será a base para o item. Por exemplo, o cenário pode ser uma situação de negócios, um problema de ciências naturais ou uma situação do dia a dia.

**Usuário:** Na missão, são US\$ 58,50 fixos para o vídeo selecionado mais US\$ 0,39 para cada 1 mil impressões (visualizações) pagas pela marca, ou seja, a marca impulsiona os vídeos selecionados e depois te paga por isso. Texto retirado de: <<https://g1.globo.com/empreendedorismo/noticia/2023/07/03/as-quatro-maneras-de-ganhar-dinheiro-no-tiktok.ghtml>>.

**Bing:** Entendi o cenário. Agora, vamos para o 3º passo: criar um item de múltipla escolha com base nas informações fornecidas.

**Questão:** Um influenciador digital recebe US\$ 58,50 fixos para cada vídeo selecionado em uma missão, além de US\$ 0,39 para cada 1 mil impressões pagas pela marca. Se a marca impulsionou um vídeo que teve 5 mil impressões, qual será o valor total que o influenciador receberá?

**Alternativas:** A) US\$ 60,45 B) US\$ 61,45 C) US\$ 62,45 D) US\$ 63,45

Agora, vamos para o 4º passo: fornecer a solução ou gabarito do item de forma detalhada.

**Gabarito:** A resposta correta é a alternativa (C). O influenciador receberá US\$ 58,50 fixos pelo vídeo mais US\$ 0,39 para cada 1 mil impressões. Como o vídeo teve 5 mil impressões, ele receberá um adicional de  $US\$ 0,39 * 5 = US\$ 1,95$ . Portanto, o valor total que o influenciador receberá será de  $US\$ 58,50 + US\$ 1,95 = US\$ 60,45$ .

Fonte: Elaborado pelo próprio Autor com auxílio do Bing Chat.

A análise dos resultados indica que o Bing Chat executou com sucesso as etapas para as quais foi instruído. O item gerado considera adequadamente o contexto e a habilidade fornecidos, além de apresentar uma resposta correta e opções incorretas. Mantendo o contexto e a habilidade, solicitou-se a criação de um item adicional para avaliar se o sujeito avaliado é capaz de identificar a lei de uma função afim que modela um problema. O diálogo subsequente ilustra esse processo.

Figura 33 – Conversação no Bing Chat para a criação de um novo item que atende à habilidade EF09MA06 da BNCC

**Usuário:** Um item que solicite a lei da função que modela o problema, considerando os valores em reais.  
**Bing:** Claro, aqui está a questão solicitada:  
*Questão:* Um influenciador digital recebe R\$ 58,50 fixos para cada vídeo selecionado em uma missão, além de R\$ 0,39 para cada 1 mil impressões pagas pela marca. Seja  $x$  o número de milhares de impressões e  $f(x)$  o valor total em reais que o influenciador receberá. Qual é a lei da função  $f(x)$  que modela essa situação?  
*Alternativas:* A)  $f(x) = 58,50x + 0,39$  B)  $f(x) = 0,39x + 58,50$  C)  $f(x) = 58,50x - 0,39$  D)  $f(x) = 0,39x - 58,50$   
*Gabarito:* A resposta correta é a alternativa (B). A função que modela essa situação é dada por  $f(x) = 0,39x + 58,50$ , onde  $x$  é o número de milhares de impressões e  $f(x)$  é o valor total em reais que o influenciador receberá. O termo  $0,39x$  representa o valor variável que depende do número de impressões e  $58,50$  é o valor fixo que o influenciador recebe por vídeo. Portanto, para qualquer número  $x$  de milhares de impressões, o influenciador receberá  $0,39x + 58,50$  reais.

Fonte: Elaborado pelo próprio Autor com auxílio do Bing Chat.

O item proposto atende ao solicitado, requerendo que o sujeito avaliado identifique a lei da função que modela o problema. As alternativas apresentadas são coerentes com o problema proposto, trazendo um gabarito e três distratores. Há um pequeno erro conceitual no comando para resposta, no qual se confunde a função  $f$  com sua imagem  $f(x)$  – erro típico até mesmo em materiais didáticos.

Em síntese, a ferramenta gerou itens que estão em conformidade com a habilidade e o contexto fornecidos, evidenciando que o Bing Chat pode ser um recurso importante na elaboração de itens matemáticos que avaliam com precisão as habilidades dos indivíduos durante seu processo de escolarização.

### 5.2.3 Aperfeiçoamento e revisão de item com o Bing Chat

Nessa seção é apresentada uma proposta de revisão e aperfeiçoamento de itens discutida na subseção 4.3.4. A proposta perpassa pelo fornecimento de algumas instruções que o chat deve levar em conta para avaliar se o item foi formulado corretamente. A seguir, apresenta-se uma primeira conversação envolvendo um item elementar atendendo uma das

habilidades preconizadas na matriz de referência do Saeb 2018 de matemática para o 9º ano do ensino fundamental.

Figura 34 – Conversação no Bing Chat para revisão de um item

**Usuário:** O objetivo deste chat é aprimorar itens de múltipla escolha de matemática. Para isso, siga o passo a passo:

1º Passo: Solicite ao usuário que forneça um item de múltipla escolha com enunciado, alternativas, gabarito e a habilidade a qual o item se refere.

2º Passo: Verifique se o item:

Possui todas as informações necessárias para resposta.  
 Possui enunciado e comando claro para resposta.  
 Está de acordo com as normas padrão da língua portuguesa.  
 Atende à habilidade proposta.  
 Tem solução única.  
 É isento de erros conceituais.  
 Apresenta enunciado claro e compreensível.  
 Possui 4 alternativas que são compatíveis com o problema e apresentam extensão compatíveis.  
 Verifique se o gabarito informado é a resposta correta do problema.

3º Passo: Caso necessário, aprimore o item considerando que ele deve atender à habilidade proposta. Além disso, verifique se o item está alinhado com os conceitos e definições apresentados no Guia de Elaboração e Revisão de Itens do INEP.

4º Passo: Apresente, ao final do item, uma proposta de solução. Certifique-se de que a solução está clara e detalhada, facilitando a compreensão do usuário.

5º Passo: Realize uma revisão final do item para garantir que todas as orientações foram seguidas corretamente e rescreeva o item fazendo todas as correções necessárias, de modo que o item tenha enunciado, comando para resposta e 4 alternativas, com apenas uma correta.

**Bing:** Entendi, você gostaria de aprimorar um item de múltipla escolha de matemática. Por favor, forneça o enunciado, as alternativas, o gabarito e a habilidade a qual o item se refere para que eu possa ajudá-lo.

**Usuário:** Eixo do conhecimento: Álgebra Eixo cognitivo: Compreender e aplicar conceitos e procedimentos Descrição da habilidade: (9A1.2) Inferir uma equação, inequação polinomial de 1º grau ou um sistema de equações de 1º grau com duas incógnitas que modela um problema.

**Item:** Sebastiana faz aulas de música e em um determinado mês ela teve 4 aulas, todas com o mesmo valor, além de uma taxa de inscrição que custou 200 reais. Seu gasto total com as aulas nesse mês foi de 600 reais. Se  $x$  é o valor de uma aula, qual equação representa corretamente a situação? A)  $x + 200 = 600$ . B)  $4x + 200 = 600$ . C)  $4x - 200 = 600$ . D)  $x + 600 + 200 = 0$ .

Gabarito: B

**Bing:** Obrigado por compartilhar o item de múltipla escolha. Vamos verificar se ele atende a todos os critérios mencionados:

*Informações necessárias para resposta:* O item possui todas as informações necessárias para a resposta.

*Enunciado e comando claro para resposta:* O enunciado é claro e o comando para a resposta também está claro.

*Normas padrão da língua portuguesa:* O item está de acordo com as normas padrão da língua portuguesa.

*Atende à habilidade proposta:* O item atende à habilidade proposta, que é inferir uma equação que modela um problema.

*Solução única:* O item tem uma solução única, que é a alternativa B.

*Isento de erros conceituais:* O item está isento de erros conceituais.

*Enunciado claro e compreensível:* O enunciado é claro e compreensível.

*4 alternativas compatíveis com o problema:* O item possui 4 alternativas que são compatíveis com o problema e apresentam extensão compatíveis.

*Gabarito correto:* A alternativa B é a resposta correta do problema, conforme informado no gabarito. Portanto, o item já está bem elaborado e não precisa de melhorias.

**Solução Proposta:** Seja  $x$  o valor de uma aula de música. Sabemos que Sebastiana teve 4 aulas e pagou uma taxa de inscrição de 200 reais. Portanto, o gasto total de Sebastiana é dado por  $4x + 200$ . Sabemos também que o gasto total foi de 600 reais. Portanto, temos a equação  $4x + 200 = 600$ .

Fonte: Elaborado pelo próprio Autor com auxílio do Bing Chat.

A ferramenta avaliou o item atendendo a cada um dos critérios estabelecidos, indicando que ele está em conformidade com a habilidade enunciada e que é ausente de erros, seja de conteúdo ou elaboração. De fato, o item proposto atende é adequado para avaliar a habilidade proposta, apresenta um único gabarito e distratores coerentes com a questão colocada.

Cabe avaliar se a ferramenta é capaz de revelar erros em itens que requerem recursos cognitivos mais sofisticados – comuns no Exame Nacional do Ensino Médio (Enem). Dando continuidade com a conversação, foi fornecido um item do referido exame, mas que foi anulado por erro de grafia, a saber, item 157 da prova azul do segundo dia do Enem 2022. O encaminhamento da conversação é exibido a seguir.

Figura 35 – Conversação no Bing Chat para revisão de um item do ENEM

**Usuário:** Revise o seguinte item, seguindo os critérios estabelecidos.  
*Competência de área 5:* Modelar e resolver problemas que envolvem variáveis socioeconômicas ou técnico-científicas, usando representações algébricas. *Habilidade:* (H21) Resolver situação-problema cuja modelagem envolva conhecimentos algébricos.  
*Item:* O funcionário de uma loja tem seu salário mensal formado por uma parcela fixa de 675 reais mais uma comissão que depende da quantidade de peças vendidas por ele no mês. O cálculo do valor dessa comissão é feito de acordo com estes critérios: até a quinquagésima peça vendida, paga-se 5 reais por peça; a partir da quinquagésima primeira peça vendida, o valor pago é de 7 reais por peça. Represente por  $q$  a quantidade de peças vendidas no mês por esse funcionário, e por  $S(q)$  o seu salário mensal, em real, nesse mês. A expressão algébrica que descreve  $S(q)$  em função de  $q$  é

A)  $S(q)=675 +12q$   
 B)  $S(q)=325 +12q$   
 C)  $S(q)= 675 +7q$   
 D)  $S(q)=\{625+5q, \text{ se } q \leq 50; 925 +7q, \text{ se } q > 50\}$   
 E)  $s(q)=\{625+5q, \text{ se } q \leq 50; 575+7q, \text{ se } q > 50\}$

**Bing:** Obrigado por compartilhar o item de múltipla escolha. Vamos verificar se ele atende a todos os critérios mencionados:

*Informações necessárias para resposta:* O item possui todas as informações necessárias para a resposta.  
*Enunciado e comando claro para resposta:* O enunciado é claro e o comando para a resposta também está claro.  
*Normas padrão da língua portuguesa:* O item está de acordo com as normas padrão da língua portuguesa.  
*Atende à habilidade proposta:* O item atende à habilidade proposta, que é resolver situação-problema cuja modelagem envolva conhecimentos algébricos.  
*Solução única:* O item tem uma solução única.  
*Isento de erros conceituais:* O item está isento de erros conceituais.  
*Enunciado claro e compreensível:* O enunciado é claro e compreensível.  
*Alternativas compatíveis com o problema:* O item possui 5 alternativas que são compatíveis com o problema e apresentam extensão compatíveis.

No entanto, parece haver um erro nas opções de resposta fornecidas. A expressão correta para o salário do funcionário em função da quantidade de peças vendidas é:

$$675 + 5q, \text{ se } q \leq 50$$

$$675 + 5 \cdot 50 + 7(q - 50) = 675 + 250 + 7q - 350 = 575 + 7q, \text{ se } q \leq 50$$

Portanto, as alternativas corrigidas seriam:

A)  $S(q)=675 +12q$   
 B)  $S(q)=325 +12q$   
 C)  $S(q)= 675 +7q$   
 D)  $S(q)=\{625+5q, \text{ se } q \leq 50; 925 +7q, \text{ se } q > 50\}$   
 E)  $S(q)=\{675+5q, \text{ se } q \leq 50; 575+7q, \text{ se } q > 50\}$   
 E o gabarito correto seria a alternativa E.

Fonte: Elaborado pelo próprio Autor com auxílio do Bing Chat.

O Bing Chat identificou de maneira precisa o erro que resultou na anulação do item do exame, fornecendo uma sugestão para correção. Contudo, a competência exigida pelo item é a seguinte: (H19) Identificar representações algébricas que articulem a correlação entre grandezas. Embora o problema proposto mobilize conhecimentos algébricos e o processo de modelagem, ele convoca um conhecimento ainda mais fundamental: a habilidade de identificar uma representação algébrica que modele adequadamente o problema em questão. Isso mostra eventuais limitações da ferramenta de avaliar se o item atende precisamente a habilidade informada.

### **5.3 Qualidade psicométrica de alguns dos itens produzidos**

Com o objetivo de avaliar a aprendizagem de um conjunto 61 estudantes do 9º ano, foi elaborado um exame composto por 20 itens de matemática. Estes itens avaliam habilidades previstas na Matriz de Referência do Saeb 2001 para o 9º ano, especificamente na componente de matemática. Alguns dos itens foram extraídos diretamente de exames anteriores do Saeb, que são disponibilizados ao público. Outros foram retiradas do livro Avalia Brasil (ASSUNÇÃO; CAVALCANTI; FREITAS, 2019). Os demais itens foram criados pelo pesquisador, com o auxílio das ferramentas de chat inteligente Bing Chat e ChatGPT, seguindo os protocolos discutidos na metodologia desse trabalho.

A Tabela 3 mostra alguns parâmetros clássicos utilizados para avaliar a qualidade do item em termos estatísticos. O primeiro dos índices é simplesmente a proporção de acerto ao item, isto é, é obtido pela razão entre o número de respondentes que acertam o item e o número total de participantes. Quanto mais próximo de 1, mais fácil é o item e quanto mais próximo de 0, maior é seu nível de dificuldade. Quanto as medidas de discriminação D e correlação ponto bisserial, considera-se discriminativos os itens com valores maiores de 0,4 (RABELO, 2013). A medida de alfa do teste foi 0,80, o que indica consistência interna aceitável.

Dentre os itens, os que apresentaram índice de discriminação inferiores a 0,4 foram: 3, 5, 11, 18 e 19. Metade desses itens com baixo poder discriminativo foram elaborados ou aperfeiçoados com auxílio de IA generativa. Além disso, a aplicação do teste t de Student para a comparação de médias de dados independentes revela que as médias do índice de discriminação D para itens derivados de outras fontes (0,54) e aqueles que foram construídos ou refinados com o auxílio da Inteligência Artificial (0,50) não apresentam diferença estatística significativa ( $t = 0,40$ ;  $df = 15,64$ ;  $p\text{-value} = 0,69$ ).

Esta constatação sugere que, no contexto do instrumento avaliado, a qualidade dos itens gerados, em termos do índice de discriminação, é comparável àquela dos itens obtidos de



fontes externas, os quais foram elaborados por especialistas. É importante destacar que todos os itens produzidos foram cuidadosamente revisados pelo pesquisador para que atendesse as orientações vistas em Brasil (2010), Caed (2008) e Vianna (1978) e que foram expostas na Subseção 3.3.

Tabela 2 – Parâmetros clássicos dos itens que compõe o teste: proporção de acerto (dificuldade), discriminação (clássica e correlação ponto bisserial) e consistência interna (alfa de Cronbach)

Item	Proporção de acerto	Discriminação clássica (D)	Correlação ponto bisserial	alfa de Cronbach (0,80)
1*	0,59	0,50	0,47	0,80
2*	0,80	0,56	0,50	0,80
3*	0,95	<b>0,06</b>	<b>0,03</b>	0,81
4*	0,77	0,69	0,55	0,79
5**	0,56	0,44	<b>0,38</b>	0,80
6**	0,69	0,56	0,49	0,80
7**	0,57	0,44	0,40	0,80
8**	0,41	0,56	0,53	0,79
9*	0,69	0,56	0,49	0,80
10**	0,36	0,62	0,54	0,79
11**	0,52	<b>0,25</b>	<b>0,22</b>	0,81
12*	0,79	0,50	0,47	0,80
13**	0,31	0,75	0,60	0,79
14**	0,61	0,62	0,50	0,80
15*	0,61	0,69	0,53	0,79
16**	0,49	0,81	0,65	0,79
17*	0,46	0,62	0,51	0,79
18*	0,87	<b>0,38</b>	0,45	0,80
19**	0,75	<b>0,38</b>	<b>0,37</b>	0,80
20**	0,67	0,50	0,44	0,80

Fonte: Elaborado pelo próprio autor. \*Retirados de outras fontes: Avalia Brasil, Inep ou CAED. \*\*Criado ou aperfeiçoado com auxílio de IA Generativa.

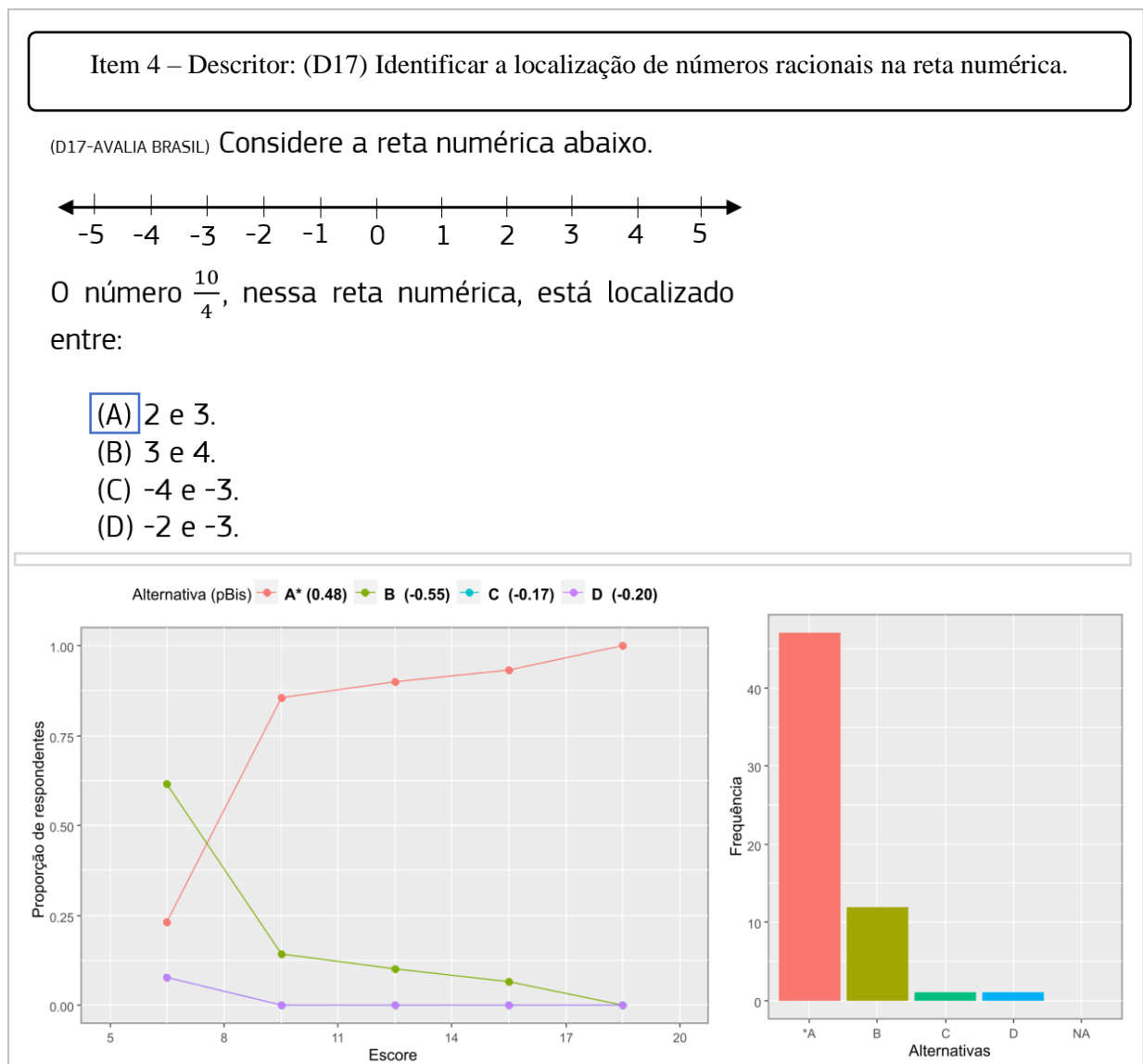
Com o objetivo de realizar uma análise gráfica, foram selecionados dois itens que demonstraram alta qualidade psicométrica (itens 4 e 13) e dois itens que exibiram limitações significativas nesse aspecto (itens 3 e 11). Esta análise é conduzida com a intenção de fornecer orientação ao educador interessado no tema, que pode não dispor de recursos suficientes para avaliar os itens de múltipla escolha utilizados na avaliação de seus alunos.

A análise gráfica de itens (AGI) proposta por Batenburg e Laros (2002) pode ser realizada com auxílio de um pacote estatístico “Itan” para RStudio. Ele é gratuito e apresenta

uma documentação em espanhol de fácil compreensão. Os *scripts* para produzir os gráficos estão disponíveis nos apêndices desse trabalho.

O item 4 foi retirado de Assunção, Cavalcanti e Freitas (2019). Ele busca avaliar se o educando é capaz de localizar números racionais na reta numérica – habilidade prevista no descritor de número 17 da Matriz de Referência do Saeb para os anos finais do ensino fundamental (edição 2001). A AGI indica que à medida que os escores dos sujeitos aumenta, aumenta também a proporção de respondentes que optam pela alternativa correta (A), de acordo com o que é esperado para um item de qualidade. O item pode ser considerado fácil para os sujeitos testados (77% de acerto) e exibe um excelente índice de discriminação.

Figura 36 – Item 4 e sua análise gráfica

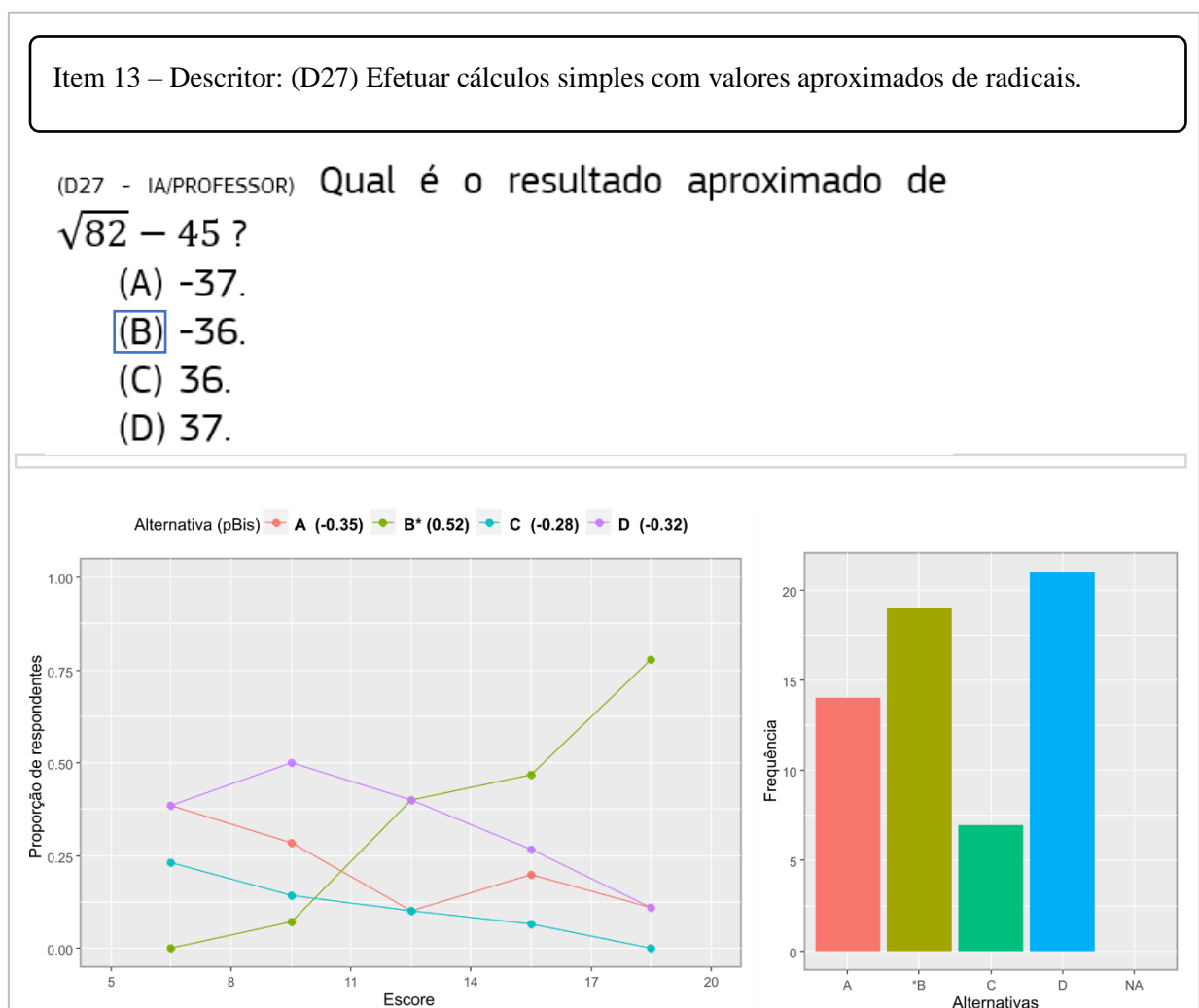


Fonte: Elaborado pelo próprio Autor.

Na AGI exibida acima, é possível identificar a medida de correlação ponto bisserial de cada alternativa e percebe-se maior valor para alternativa A, evidenciando seu bom poder discriminativo.

Elaborado pelo pesquisador com auxílio do Bing Chat, o item 13 é outro exemplo de item com bom poder discriminativo (Figura 37), mas que pode ter sido difícil para a maioria dos sujeitos avaliados, com apenas 31% de acerto. O comportamento gráfico do item evidenciado na AGI e a medida de correlação ponto bisserial da alternativa correta atesta tal qualidade. No gráfico de barras, percebe-se que o distrator D foi atrativo para os sujeitos com baixa proficiência, revelando que parte significativa dos educandos avaliados podem ter conduzido o cálculo subtraindo 82-45, sem operar antes com radical. Um item como esse pode indicar a falta de familiaridade ou compreensão do conceito de raiz quadrada por um grupo de sujeitos, requerendo uma atenção especial.

Figura 37 – Item 13 e sua análise gráfica

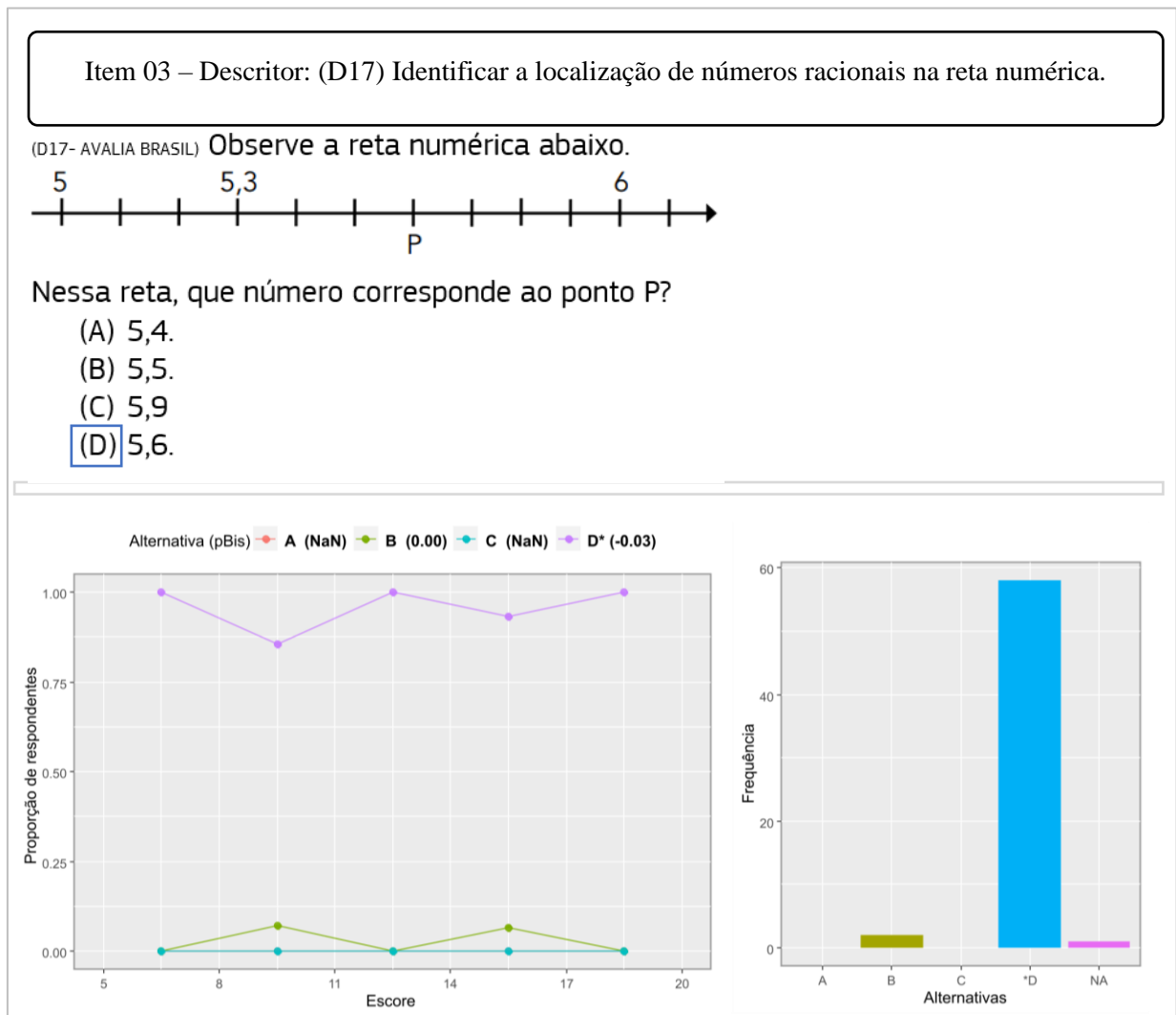


Fonte: Elaborado pelo próprio Autor.

O item 3, exibido na Figura 38, foi obtido de Assunção, Cavalcanti e Freitas (2019). Ele apresenta baixo poder discriminativo e pode ser considerado muito fácil para o particular grupo de respondentes, com 95% de acerto. O carácter óbvio do item para o grupo de respondente explica o porquê ele é pouco discriminativo, uma vez que, independentemente do nível de proficiência, os sujeitos foram conduzidos à resposta correta.

Do ponto de vista pedagógico, é animador para o professor observar que quase 100% dos estudantes demonstram domínio da situação proposta. No entanto, quando se trata de objetivos técnicos, especialmente no que se refere aos atributos estatísticos e psicométricos, esse é um item que contribui pouco para a mensuração da proficiência dos alunos. É importante mencionar que um teste avaliativo justo é aquele que apresenta, de forma equilibrada, itens de todos os níveis de dificuldade e que sejam coerentes com o nível escolar dos sujeitos avaliados.

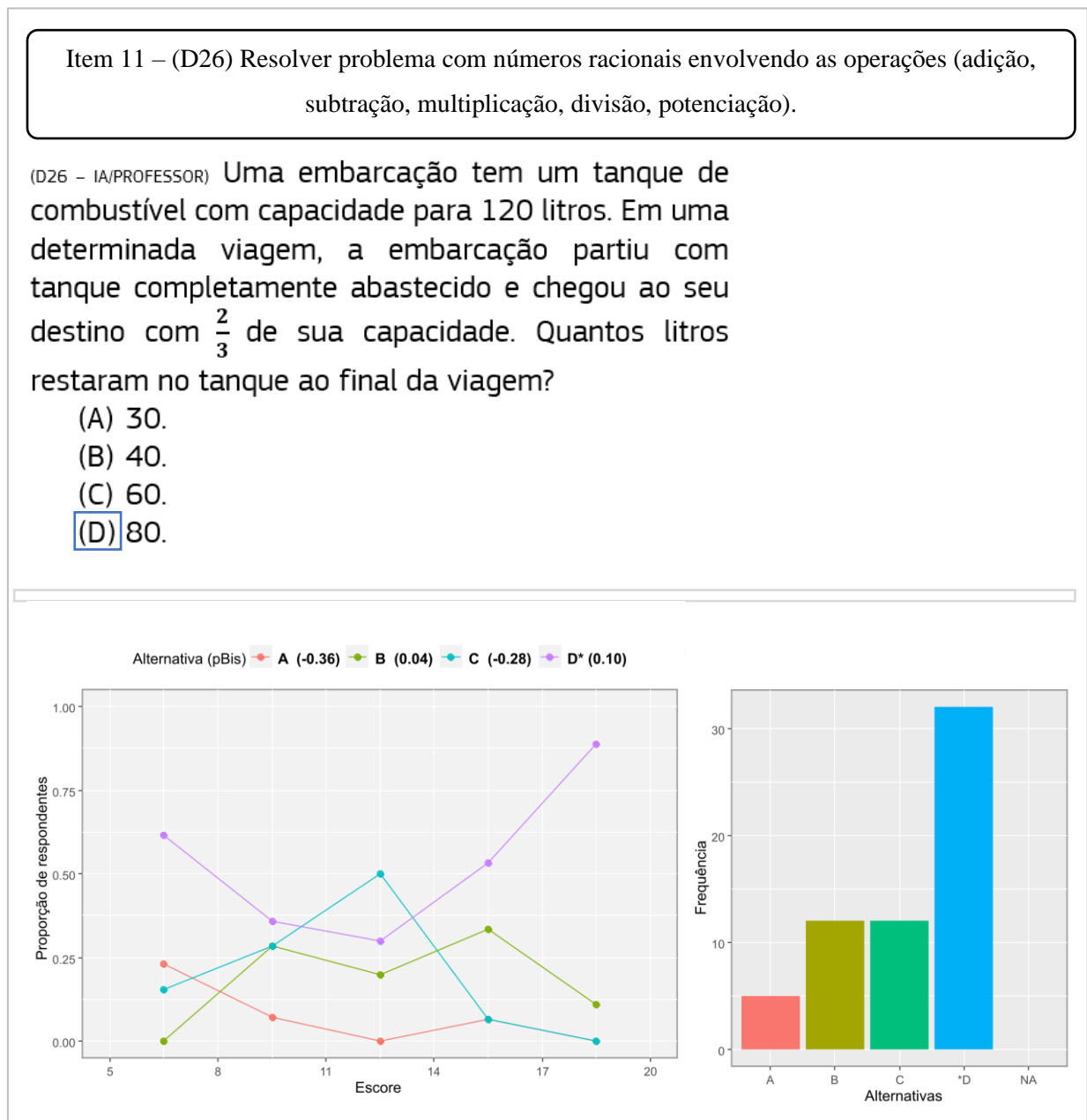
Figura 38 – Item 3 e sua análise gráfica



Fonte: Elaborado pelo próprio Autor.

Tomando para análise o item 11, que foi elaborado com auxílio do Bing Chat, é possível notar, em sua AGI, que mesmo sujeitos com elevada proficiência podem ainda ter tido dificuldade na resolução do problema proposto (Figura 39). AGI e a medida de correlação ponto bisserial revela ainda uma faixa de discriminação restrita para o item em análise. Quanto ao nível de dificuldade, o item 11 foi respondido corretamente por 52% dos respondentes, o que sugere um nível médio de dificuldade.

Figura 39 – Item 11 e sua análise gráfica



Fonte: Elaborado pelo próprio Autor.

Sob a perspectiva pedagógica, a análise do item 11 indica que aproximadamente metade dos discentes exibem desafios ao confrontar o problema apresentado, seja devido à insuficiência no domínio dos procedimentos matemáticos necessários ou à dificuldade nos processos cognitivos associados à interpretação do enunciado e do comando para resposta.

## 6 CONSIDERAÇÕES FINAIS

Ao longo do estudo, foram abordados aspectos históricos e conceituais da Inteligência Artificial (IA), bem como a maneira pela qual esse campo, de abrangência significativa e com vastas aplicações, pode contribuir para o aperfeiçoamento dos processos de ensino e aprendizagem escolar. Embora as ferramentas baseadas em IA estejam cada vez mais presentes na vida cotidiana, a utilização ética e confiável dessas tecnologias ainda representa um desafio. Isso impõe a necessidade de integrar aos objetivos educacionais, em todos os níveis aplicáveis, três aspectos fundamentais: a aprendizagem com a IA, que envolve o uso de ferramentas de IA para auxiliar no processo de aprendizagem; a aprendizagem sobre a IA, que se refere ao entendimento dos conceitos e princípios da IA; e a aprendizagem para a IA, que envolve o desenvolvimento de habilidades necessárias para criar e utilizar efetivamente as ferramentas de IA (BIDARRA; SIMONSEN; HOLMES, 2020; DU BOULAY, 2022; HARASIM, 2015; JARA; OCHOA, 2020; SICHMAN, 2021; YU, 2023).

Ao mesmo tempo, os avanços no campo do Aprendizado de Máquina trouxeram benefícios significativos ao desenvolver ferramentas com o potencial de aprimorar os processos de avaliação da aprendizagem escolar, inclusive na educação básica. Tarefas anteriormente realizadas exclusivamente por profissionais da educação e avaliadores, como a elaboração de itens e testes avaliativos, agora podem ser produzidas com o apoio de ferramentas inteligentes. Isso confere ao docente ou elaborador um papel mais analítico, permitindo que tarefas repetitivas sejam atribuídas a chats inteligentes. Esses sistemas produzem resultados textuais em linguagem natural, como é o caso do Bing Chat e do ChatGPT.

O estudo identificou limitações dos chats inteligentes nos processos de elaboração de itens, bem como a necessidade de cautela ao atribuir tarefas a essas ferramentas, que podem ocasionalmente apresentar falhas, produzindo informações equivocadas ou com vieses originados de sua base de dados de treinamento. Isso requer que o elaborador possua um conhecimento profundo dos atributos que caracterizam um item de alta qualidade, bem como um entendimento abrangente dos processos de avaliação da aprendizagem e do conteúdo avaliado. Essa precaução é essencial para evitar o risco de empobrecer os instrumentos de avaliação com itens de baixa qualidade.

As sugestões fornecidas aos chats inteligentes para a elaboração de itens indicam que essas ferramentas geram esboços bastante elementares, mas que satisfazem adequadamente as habilidades informadas. O enriquecimento dos itens contextualizados depende da provisão de contextos mais específicos, juntamente com a descrição de uma habilidade que retoma o

contexto apresentado e mobiliza os indivíduos para a resolução de um problema. O processo de elaboração pode transcender os esquemas apresentados, tornando-se mais dinâmico, em um ciclo de elaboração, revisão e aperfeiçoamento conduzido conjuntamente pelos dois agentes do processo: o professor e a IA.

O ChatGPT, fundamentado no modelo de linguagem GPT 3.5, apresentou limitações na elaboração de questões matemáticas simples baseadas em uma habilidade textual fornecida associada à resolução de problemas. Entre as limitações observadas, destaca-se a falta de precisão na resolução e explicação das alternativas. Em contrapartida, o modelo foi capaz de produzir itens satisfatórios com base em um contexto e habilidade previamente fornecidos, embora as alternativas necessitem quase sempre de revisões.

No que se refere à revisão de itens de acordo com as orientações propostas, a ferramenta pode auxiliar o elaborador a verificar se o item atende à habilidade descrita, se está em conformidade com a norma padrão da língua portuguesa e se contém elementos essenciais, como enunciado, comando para resposta, alternativas e gabarito.

A outra ferramenta avaliada, o Bing Chat, exibiu erros semelhantes aos observados no ChatGPT, porém demonstrou maior sucesso na maioria das tarefas. Os itens de matemática gerados com base apenas na habilidade informada apresentaram limitações mais significativas nas alternativas, embora os enunciados tenham sido claros e estejam alinhados à habilidade prevista. O sucesso da ferramenta é evidenciado na elaboração de itens com base no contexto e habilidade, bem como na revisão do item. Um dos exemplos mostrou que a ferramenta identificou um erro que resultou na exclusão de um item do exame do Enem. Isso indica que o chat pode prevenir problemas como esse, até mesmo em avaliações importantes, conduzidas em larga escala.

Um conjunto de itens, que foram elaborados ou aperfeiçoados com o auxílio do Bing Chat e ChatGPT em conformidade com as sugestões apresentadas no escopo deste estudo, foi incorporado em um teste para avaliar a aprendizagem de 61 estudantes na disciplina de matemática. Dos 11 itens, apenas 2 demonstraram limitações em relação ao poder discriminativo ( $D$ ), uma métrica que atesta a capacidade do item de distinguir entre estudantes proficientes e não proficientes. Ao comparar com os itens obtidos de outras fontes, que foram elaborados por especialistas, não se observou diferença significativa nos índices médios de discriminação dos dois grupos de itens, isto é, os itens produzidos possuem índice de discriminação semelhantes aos que foram apresentados por itens criados e revisados por especialistas.



Esta constatação sugere que tais ferramentas baseadas em IA podem ser de grande auxílio para professores ou elaboradores na construção de itens de alta qualidade, destinados à avaliação da aprendizagem. No entanto, é importante esclarecer que a qualidade dos itens não é garantida pelo uso indiscriminado dessas ferramentas, mas sim pela expertise e conhecimento de quem as utiliza. A tarefa de produzir itens deve ser cuidadosamente supervisionada e aperfeiçoada pelo docente ou elaborador, que precisa estar familiarizado com boas práticas de elaboração e revisão de itens, bem como com os documentos que orientam ou definem os objetivos de aprendizagem para a etapa de ensino avaliada.

A construção de itens de múltipla escolha de alta qualidade, fundamentados em matrizes de referência específicas, é essencial para a realização de avaliações externas e eventualmente em avaliações diagnósticas e somativas. No entanto, no contexto das avaliações escolares, esses itens não devem excluir outros formatos ou outras estratégias avaliativas. O objetivo da avaliação no âmbito escolar é mais amplo e se concentra na formação do indivíduo em várias dimensões (LUCKESI, 2013).

Os testes padronizados contribuem para a mensuração da proficiência em aspectos cognitivos elementares, que podem ser traduzidos por meio de procedimentos ou processos mentais. No entanto, esses testes não abordam dimensões socioemocionais e diversas outras competências previstas no currículo escolar, que são essenciais para o desenvolvimento socioafetivo e cognitivo do indivíduo. Portanto, este estudo apresenta apenas um recorte de um dos instrumentos que auxiliam na avaliação da aprendizagem: o item de múltipla escolha. Para trabalhos futuros, sugere-se a avaliação das possibilidades e limitações de tais ferramentas na construção de outros formatos de itens.

## REFERÊNCIAS

ASSUNÇÃO, C.; CAVALCANTI, M.; FREITAS, R. **Avalia Brasil: matemática, ensino fundamental II**. 1. ed. São Paulo: Eureka, 2019.

BATENBURG, T. A. van; LAROS, J. A. Graphical Analysis of Test Items. **Educational Research and Evaluation**, v. 8, n. 3, p. 319–333, 2002.

BIDARRA, J.; SIMONSEN, H. K.; HOLMES, W. **Artificial Intelligence in Teaching (AIT): A road map for future developments**. Disponível em: <[https://www.researchgate.net/publication/342214907\\_Artificial\\_Intelligence\\_in\\_Teaching\\_AIT\\_A\\_road\\_map\\_for\\_future\\_developments](https://www.researchgate.net/publication/342214907_Artificial_Intelligence_in_Teaching_AIT_A_road_map_for_future_developments)>. Acesso em: 23 jul. 2023.

BRASIL. **Guia de Revisão e Elaboração de Itens. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep)**. Brasília: INEP. , 2010

\_\_\_\_\_. **Matrizes e Escalas**. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb/matrizes-e-escalas>>. Acesso em: 30 out. 2023a.

\_\_\_\_\_. **Ministério da Educação. Base Nacional Comum Curricular**. . Brasília: MEC. Disponível em: <<http://basenacionalcomum.mec.gov.br/>>. Acesso em: 5 ago. 2023a. , 2018

\_\_\_\_\_. **Sistema de Avaliação da Educação Básica, Documentos de Referência - Versão 1.0**. . Brasília: Inep. Disponível em: <[https://download.inep.gov.br/educacao\\_basica/saeb/2018/documentos/saeb\\_documentos\\_de\\_referencia-versao\\_1.0.pdf](https://download.inep.gov.br/educacao_basica/saeb/2018/documentos/saeb_documentos_de_referencia-versao_1.0.pdf)>. Acesso em: 24 jun. 2023b. , 2018

\_\_\_\_\_. **Sistema de Avaliação da Educação Básica (Saeb)**. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>>. Acesso em: 5 ago. 2023b.

BUCHANAN, B. G. A (Very) Brief History of Artificial Intelligence. **AI Magazine**, 2005. Disponível em: <<https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1848>>. Acesso em: 26 jun. 2023.

CAED. **Guia de Elaboração de Itens: matemática. Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora**. Juiz de Fora: CAED/UFJF. , 2008

COPPIN, B. **Inteligência Artificial**. Rio De Janeiro: LTC, 2010.

DATA SCIENCE ACADEMY. **Deep Learning Book**. Disponível em: <<https://www.deeplearningbook.com.br/>>. Acesso em: 3 jan. 2024a.

\_\_\_\_\_. **O Neurônio, Biológico e Matemático**. Disponível em: <<https://www.deeplearningbook.com.br/o-neuronio-biologico-e-matematico/>>. Acesso em: 3 jan. 2024b.

DELIPETREV, B.; TSINARAKII, C.; KOSTIĆ, U. **Historical Evolution of Artificial Intelligence**. . Luxembourg: [s.n.], 2020. Disponível em: <<https://eprints.ugd.edu.mk/28050/>>. Acesso em: 29 jun. 2023.

DU BOULAY, B. Artificial Intelligence in Education and Ethics. **Handbook of Open, Distance and Digital Education**. Singapore: Springer Nature Singapore, 2022. p. 1–16.

GATTI, F. N. **Educação básica e inteligência artificial: perspectivas, contribuições e desafios**. 2019. 1–90 f. Dissertação (Mestrado em Educação: currículo)–Pontifícia Universidade Católica de São Paulo, São Paulo, 2019. Disponível em: <<https://tede2.pucsp.br/handle/handle/22788>>. Acesso em: 25 jun. 2023.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GIMENES, N. et al. Além da Prova Brasil: investimento em sistemas próprios de avaliação externa. **Estudos em Avaliação Educacional**, v. 24, n. 55, p. 12, 31 ago. 2013. Disponível em: <<http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/2718>>.

HAENLEIN, M.; KAPLAN, A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. **California Management Review**, v. 61, n. 4, p. 5–14, 17 ago. 2019. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0008125619864925>>.

HARASIM, L. Educação Online e as implicações da Inteligência Artificial. **Revista da FAEEBA - Educação e Contemporaneidade**, v. 24, n. 44, 4 dez. 2015. Disponível em: <<http://testeocs.uneb.br/index.php/faeaba/article/view/1818>>.

JARA, I.; OCHOA, J. M. **Usos y efectos de la inteligencia artificial en educación**. . Washington, D.C.: [s.n.], maio 2020. Disponível em: <<https://publications.iadb.org/es/ usos-y-efectos-de-la-inteligencia-artificial-en-educacion>>.

KLEIN, R. **Escala de proficiência**. Disponível em: <<https://www.ceale.fae.ufmg.br/glossarioceale/verbetes/escala-de-proficiencia>>. Acesso em: 30 out. 2023.

LAROS, J. A. Análise Gráfica de Itens. In: PASQUALI, L. (Org.). **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Vozes, 2017. .

LUCKESI, C. C. **Avaliação da aprendizagem escolar: estudo e proposições**. São Paulo: Cortez, 2013. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788524921063/>>. Acesso em: 2 ago. 2023.

\_\_\_\_\_. **Avaliação em educação: questões epistemológicas e práticas**. São Paulo: Cortez, 2022.

MACHADO, C.; ALAVARSE, O. M. Qualidade das escolas: tensões e potencialidades das avaliações externas. **Educação & Realidade**, v. 39, p. 413–436, 2014. Disponível em: <<https://www.scielo.br/j/edreal/a/fyKb65xtFvXhMw3Hhvp9vNk/?lang=pt#>>. Acesso em: 5 ago. 2023.

PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Vozes, 2017.

PERRENOUD, P. **Desenvolver competências ou ensinar saberes?** Porto Alegre: Penso, 2013.

RABELO, M. **Avaliação educacional: fundamentos, metodologia e aplicações no contexto brasileiro**. Rio de Janeiro: SBM, 2013.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Elsevier, 2013.

SAMOILI, S. et al. **Defining Artificial Intelligence 2.0: Towards an operational definition and taxonomy for the AI landscape**. Luxembourg: [s.n.], 2021. Disponível em: <<https://publications.jrc.ec.europa.eu/repository/handle/JRC126426>>. Acesso em: 2 jul. 2023.

SANTANA, M. **Deep Learning: do Conceito às Aplicações**. Disponível em: <<https://medium.com/data-hackers/deep-learning-do-conceito-%C3%A0s-aplica%C3%A7%C3%B5es-e8e91a7c7eaf>>. Acesso em: 8 jul. 2023.

SEJNOWSKI, T. J. Large Language Models and the Reverse Turing Test. **Neural Computation**, v. 35, n. 3, p. 309–342, 17 fev. 2023. Disponível em: <<https://direct.mit.edu/neco/article/35/3/309/114731/Large-Language-Models-and-the-Reverse-Turing-Test>>.

SICHMAN, J. S. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**, v. 35, n. 101, p. 37–50, 19 abr. 2021. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-40142021000100037&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142021000100037&tlng=pt)>. Acesso em: 24 jun. 2023.

SILVA, A. C. **Análise psicométrica e pedagógica da prova da primeira fase da OBMEP respondida por um grupo de alunos do 6º e 7º ano do ensino fundamental**. 2019. 1–85 f. Trabalho de Conclusão de Curso –Universidade Federal do Oeste do Pará, Santarém, 2019.

SOUSA, C. P. de. Dimensões da avaliação educacional. **Estudos em Avaliação Educacional**, n. 22, p. 101, 30 dez. 2000. Disponível em: <<http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/2218>>.

UNESCO. **Planejando a educação na era da IA: liderar o avanço. Consenso de Beijing sobre a inteligência artificial e a educação**. Paris: Organização das Nações Unidas para a Educação, a Ciência e a Cultura. , 2019

VIANNA, H. M. **Testes em Educação**. 3. ed. São Paulo: Ibrasa, 1978.

YU, H. Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. **Frontiers in Psychology**, v. 14, 1 jun. 2023. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1181712/full>>.

## **APÊNDICES**

### **APÊNDICE A - DECLARAÇÃO DE CONFLITO DE INTERESSE**

O autor declara que a pesquisa foi realizada na ausência de quaisquer relações comerciais ou financeiras que possam ser interpretadas como um potencial conflito de interesses.

**APÊNDICE B – SCRIPT DAS ANÁLISES DOS ITENS NO RSTUDIO**

```
#PACOTES
```

```
#Pacote para abrir arquivo .xlsx (Excel) no Rstudio
```

```
require("readxl")
```

```
#Pacote para análise gráfica e dos escores de um teste de múltipla escolha
```

```
require("itan")
```

```
#Pacote para análises TCT e TRI
```

```
require("ltm")
```

```
#Pacote para plotar gráficos ggplot
```

```
require("ggplot2")
```

```
#Pacote para manipulação de dados, como seleção e filtragem
```

```
require("dplyr")
```

```
# LENDO A BASE DE DADOS
```

```
bd <- read_excel("H:/OneDrive/Documentos/01 - Mestrado  
Matemática/DISSERTAÇÃO/Simulados/Simulados.xlsx",  
               sheet = "4simulado_saeb")
```

```
gabarito<- read_excel("H:/OneDrive/Documentos/01 - Mestrado  
Matemática/DISSERTAÇÃO/Simulados/Simulados.xlsx",  
                    sheet = "Gab_4simulado")
```

```
#DICOTOMIZAÇÃO DOS DADOS (Correção da prova)
```

```
respostas <-bd[, 5:24]
```

```
bd_corrigido <- corregirRespuestas(respostas, gabarito)
```

```
head(bd_corrigido)
```

```
#CÁLCULO DOS ESCORES
```

```
escore <-calcularPuntajes(bd_corrigido)
```

```
#ÍNDICE DE DISCRIMINAÇÃO D (27%D_SUPERIOR -27%INFERIOR)
```

```
data.frame(calcularIndiceDiscriminacion(
```

```
  bd_corrigido,
```

```
  tipo = "dc1",
```

```
  proporcion = 0.27,
```

```
  digitos = 2
```

```
))
```

```
#PARÂMETROS DA TCT: DIFICULDADE, CORRELAÇÃO PONTO BISSERIAL E  
ALFA DE CRONBACH
```

```
descript(bd_corrigido)
```

```
#GRÁFICOS
```

```
alternativas <- LETTERS[1:4]
```

```
#Gráfico de barras por item
```

```
g <- graficarFrecuenciaAlternativas(respostas, alternativas, gabarito)
```

```
g$Q4
```

```
g$Q13
```

```
g$Q3
```

```
g$Q11
```

```
#Análise Gráfica do Item (AGI)
```

```
agi <- agi(respostas, gabarito, alternativas, nGrupos = 5, digitos = 2)
```

```
agi$Q4
```

```
agi$Q13
```

```
agi$Q3
```

```
agi$Q11
```

## #COMPARAÇÃO DA DISCRIMINAÇÃO DE ITENS EXTERNOS COM ITENS CRIADOS COM AUXÍLIO DE IA

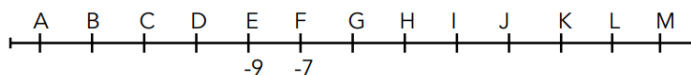
```
an<- read_excel("H:/OneDrive/Documentos/01 - Mestrado
Matemática/DISSERTAÇÃO/Simulados/Simulados.xlsx",
               sheet = "an")
shapiro.test(an$`Discriminação clássica`)
g1 <- an %>% filter(Fonte=="CRIADO COM AUXÍLIO DE IA") %>%
select("Discriminação clássica")
g2 <- an %>% filter(Fonte=="OUTRAS FONTES") %>% select("Discriminação clássica")
t.test(g1, g2)
```



## APÊNDICE C – ITENS DO TESTE DE MATEMÁTICA APLICADO AOS ESTUDANTES

ITEM 1-----

(D16-AVALIA BRASIL) Na reta numérica da figura, o ponto E corresponde ao número inteiro -9 e o ponto F, ao inteiro -7.

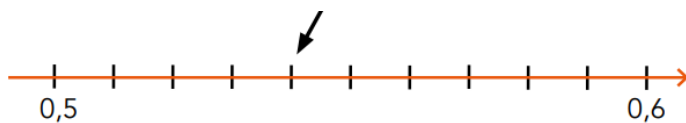


Nessa reta, o ponto correspondente ao inteiro zero estará

- (A) sobre o ponto M.
- (B) entre os pontos L e M.
- (C) entre os pontos I e J.
- (D) sobre o ponto J.

ITEM 2-----

(D17-SAEB) Observe os números que aparecem na reta.

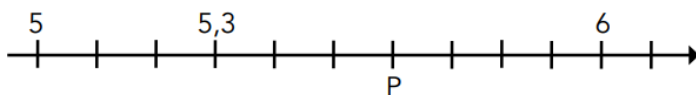


O número indicado pela seta é

- (A) 0,54.
- (B) 0,8.
- (C) 0,55.
- (D) 0,9.

ITEM 3-----

(D17- AVALIA BRASIL) Observe a reta numérica abaixo.

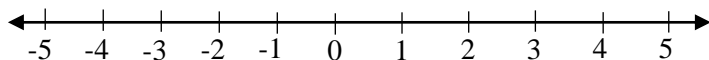


Nessa reta, que número corresponde ao ponto P?

- (A) 5,4.
- (B) 5,5.
- (C) 5,9
- (D) 5,6.

## ITEM 4-----

(D17-AVALIA BRASIL) Considere a reta numérica abaixo.



O número  $\frac{10}{4}$ , nessa reta numérica, está localizado entre:

- (A) 2 e 3.
- (B) 3 e 4.
- (C) -4 e -3.
- (D) -2 e -3.

## ITEM 5-----

(D18 – IA/PROFESSOR) Observe a expressão

$$(2^3 \cdot 5 - 10) : 2 + (15 \cdot 3 - 5^2)$$

Seu resultado é igual a:

- (A) 20.
- (B) 25.
- (C) 30.
- (D) 35.

## ITEM 6-----

(D19- IA/PROFESSOR) Sebastiana quer comprar uma televisão que custa R\$ 1800,00 à vista. No entanto, a loja oferece a opção de pagar uma entrada de R\$ 400,00 e o restante em quatro parcelas iguais. Se Sebastiana optar pela compra a prazo, qual será o valor de cada parcela?

- (A) R\$350,00.
- (B) R\$400,00.
- (C) R\$450,00.
- (D) R\$500,00.

## ITEM 7-----

(D20- IA/PROFESSOR) Em Gramado, Rio Grande do Sul, a temperatura pode cair para  $-5^{\circ}\text{C}$  durante o inverno. Se em um dia a temperatura cai  $3^{\circ}\text{C}$  abaixo da temperatura mais baixa já registrada e no dia seguinte sobe  $7^{\circ}\text{C}$ , qual será a temperatura no segundo dia?

- (A)  $-1^{\circ}\text{C}$ .
- (B)  $0^{\circ}\text{C}$ .
- (C)  $1^{\circ}\text{C}$ .
- (D)  $2^{\circ}\text{C}$ .

## ITEM 8-----

(D23- IA/PROFESSOR) Foi realizada uma eleição para representante na turma do 9º ano. A tabela a seguir mostra quem foram os candidatos e a fração dos votos que cada um recebeu.

Candidato	Fração do total de votos
Valter	$\frac{1}{6}$
Sandro	$\frac{2}{10}$
Milena	$\frac{1}{3}$
Paula	$\frac{1}{10}$
Maria	$\frac{1}{5}$

Se 30 estudantes votaram e cada um teve direito a um voto, os candidatos que receberam a mesma quantidade de votos foram:

- (A) Valter e Milena.
- (B) Sandro e Paula.
- (C) Sandro e Maria.
- (D) Valter e Maria.

## ITEM 9-----

(D24 - CAED) Observe o número apresentado no quadro abaixo.




0,25

Qual é a escrita por extenso desse número?

- (A) Vinte e cinco décimos.
- (B) Vinte e cinco centésimos.
- (C) Vinte e cinco inteiros.
- (D) Vinte e cinco milésimos.

## ITEM 10-----

(D25 - IA/PROFESSOR) Na feira do seu José, os abacates, as bananas e as maçãs estão em oferta. Os preços podem ser vistos na figura abaixo.

Abacate	Banana	Maçã
		
R\$ 7,90 kg	R\$ 5,75 kg	R\$ 6,60 kg

Um cliente comprou 1,5 kg de abacate, 3 kg de banana e 2,5 kg de maçã. Ele pagou as compras com uma nota de R\$ 50,00. Qual foi o valor do troco que o cliente recebeu?

- (A) R\$ 4,40.
- (B) R\$ 5,40.
- (C) R\$ 16,25
- (D) R\$ 20,90.

## ITEM 11-----

(D26 - IA/PROFESSOR) Uma embarcação tem um tanque de combustível com capacidade para 120 litros. Em uma determinada viagem, a embarcação partiu com tanque completamente abastecido e chegou ao seu destino com  $\frac{2}{3}$  de sua capacidade. Quantos litros restaram no tanque ao final da viagem?

- (A) 30.
- (B) 40.
- (C) 60.
- (D) 80.

## ITEM 12-----

(D26 - CAED) Em uma empresa com 40 funcionários, 27 deles têm plano de saúde. Qual é a fração que representa a quantidade de funcionários dessa empresa que têm plano de saúde, em relação ao total de funcionários?

- (A)  $\frac{13}{40}$ .
- (B)  $\frac{40}{27}$ .
- (C)  $\frac{40}{13}$ .
- (D)  $\frac{27}{40}$ .

## ITEM 13-----

(D27 - IA/PROFESSOR) Qual é o resultado aproximado de  $\sqrt{82} - 45$  ?

- (A) -37.
- (B) -36.
- (C) 36.
- (D) 37.

## ITEM 14-----

(D28 - IA/PROFESSOR) Sebastião trabalha em uma loja de roupas e recebe uma comissão de 5% sobre as vendas que realiza. Em um determinado mês, Sebastião vendeu R\$ 20 000,00 em roupas. Quanto Sebastião recebeu de comissão naquele mês?

- (A) R\$ 500,00.
- (B) R\$ 1 000,00.
- (C) R\$ 1 500,00.
- (D) R\$ 2 000,00.

## ITEM 15-----

(D29- AVALIA BRASIL) O consumo de determinadas frutas é benéfico à saúde. Um exemplo é a pera, cujo consumo auxilia na circulação do sangue, no controle da pressão arterial e facilita a digestão. Cada 100 g dessa fruta equivalem a 56 calorias. Uma pessoa que ingere 280 g dessa fruta, fornece ao organismo:

- (A) 156,8 calorias.
- (B) 252,5 calorias.
- (C) 468,0 calorias.
- (D) 500,0 calorias.

## ITEM 16-----

(D30 - IA/PROFESSOR) O valor numérico que a expressão algébrica

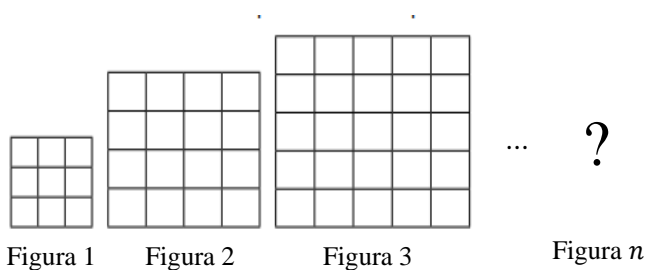
$$x^2 - 2xy + y^2$$

assume quando  $x = 2$  e  $y = -5$  é

- (A) 19.
- (B) 45.
- (C) 49.
- (D) 60.

## ITEM 17-----

(D32-AVALIA BRASIL) Observe a sequência de figuras abaixo.



Qual é a expressão algébrica que representa a quantidade de quadradinhos da enésima figura?

- (A)  $n^2$ .
- (B)  $n^2 + 4^2$ .
- (C)  $n^2 + (n + 1)^2$ .
- (D)  $(n + 2)^2$ .

## ITEM 18-----

(D34 - SAEB) Lucas comprou 3 canetas e 2 lápis pagando R\$ 7,20. Danilo comprou 2 canetas e 1 lápis pagando R\$ 4,40. O sistema de equações do 1º grau que melhor representa a situação é

(A) 
$$\begin{cases} 3x + 2y = 7,20 \\ 2x + y = 4,40 \end{cases}$$

(B) 
$$\begin{cases} 3x - 2y = 7,20 \\ 2x - y = 4,40 \end{cases}$$

(C) 
$$\begin{cases} x + y = 3,60 \\ x - y = 2,20 \end{cases}$$

(D) 
$$\begin{cases} 3x + y = 7,20 \\ x + y = 4,40 \end{cases}$$

## ITEM 19-----

(D33 - IA/PROFESSOR) Sebastiana faz aulas de música e em um determinado mês ela teve 4 aulas, todas com o mesmo valor, além de uma taxa de inscrição que custou 200 reais. Seu gasto total com as aulas nesse mês foi de 600 reais. Se  $x$  é o valor de uma aula, qual equação representa corretamente a situação?

(A)  $x + 200 = 600$ .

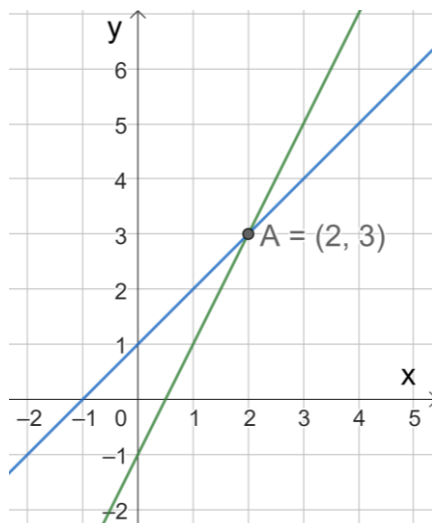
(B)  $4x + 200 = 600$ .

(C)  $4x - 200 = 600$ .

(D)  $x + 600 + 200 = 0$ .

## ITEM 20-----

(D35 - IA/PROFESSOR) No gráfico, vemos o esboço de duas retas que se intersectam no ponto  $A = (2, 3)$ .



O Sistema de equação do 1º grau representado no gráfico é

(A)  $\begin{cases} y = 2x - 1 \\ y = x + 1 \end{cases}$

(B)  $\begin{cases} y = 2x \\ y = x - 1 \end{cases}$

(C)  $\begin{cases} y = x \\ y = x + 1 \end{cases}$

(D)  $\begin{cases} y = 3x - 1 \\ y = x - 1 \end{cases}$