

<https://doi.org/10.18222/ae.v36.11415>

UMA ANÁLISE DA COMPLEXIDADE DE ITENS DE MATEMÁTICA POR MEIO DA TAXONOMIA SOLO

 SOLANGE MARIA MOL^I

 DANIEL ABUD SEABRA MATOS^{II}

^I Universidade Federal de Ouro Preto (UFOP), Mariana-MG, Brasil; solangemoli@gmail.com

^{II} Universidade Federal de Ouro Preto (UFOP), Mariana-MG, Brasil; danielmatos@ufop.edu.br

RESUMO

A pesquisa objetivou analisar a complexidade cognitiva de itens de matemática do 5º e 9º anos do ensino fundamental da Prova Brasil, a partir da taxonomia cognitiva denominada Structure of Observed Learning Outcomes (Solo). A consistência da análise qualitativa de 123 itens foi avaliada pelo método de concordância entre juízes e investigou-se a complexidade cognitiva dos itens por meio dos níveis da taxonomia Solo. Também foi verificada a relação entre os níveis da Solo e os parâmetros de discriminação e de dificuldade dos itens e os níveis de proficiência necessários para acerto. Encontramos uma expressiva proporção de itens que exigem aprendizagem superficial. O tipo de aprendizagem superficial prevaleceu inclusive nas faixas de dificuldade e discriminação altas e nos níveis de proficiência considerados adequados e avançados.

PALAVRAS-CHAVE AVALIAÇÃO EXTERNA • NÍVEIS DE COMPLEXIDADE • ANÁLISE DE ITEM.

COMO CITAR:

Mol, S. M., & Matos, D. A. S. (2025). Uma análise da complexidade de itens de matemática por meio da taxonomia Solo. *Estudos em Avaliação Educacional*, 36, Artigo e11415. <https://doi.org/10.18222/ae.v36.11415>

ANÁLISIS DE LA COMPLEJIDAD DE LOS ÍTEMS DE MATEMÁTICAS USANDO LA TAXONOMÍA SOLO

RESUMEN

Esta investigación tuvo como objetivo analizar la complejidad cognitiva de los ítems de matemáticas de 5º y 9º grado de primaria de la Prova Brasil, a partir de la taxonomía cognitiva denominada Structure of Observed Learning Outcomes (Solo). La consistencia del análisis cualitativo de 123 ítems fue evaluada mediante el método de acuerdo entre jueces y se investigó la complejidad cognitiva de los ítems por medio de los niveles de la taxonomía Solo. También fue verificada la relación entre los niveles de la Solo y los parámetros de discriminación y de dificultad de los ítems, así como los niveles de competencia requeridos para los aciertos. Encontramos una proporción significativa de ítems que exigen un aprendizaje superficial. El tipo de aprendizaje superficial prevaleció incluso en los niveles altos de dificultad y discriminación, y en los niveles de competencia considerados adecuados y avanzados.

PALABRAS CLAVE EVALUACIÓN EXTERNA • NIVELES DE COMPLEJIDAD • ANÁLISIS DE ÍTEMS.

AN ANALYSIS OF THE COMPLEXITY OF MATHEMATICS ITEMS USING SOLO TAXONOMY

ABSTRACT

This study aimed to analyze the cognitive complexity of mathematics items from the 5th and 9th grades of elementary education in the Prova Brasil assessment, based on the cognitive taxonomy called Structure of Observed Learning Outcomes (Solo). The consistency of the qualitative analysis of 123 items was evaluated using the inter-rater agreement method, and the cognitive complexity of the items was investigated using the levels of the Solo taxonomy. The relationship between Solo levels and the discrimination and difficulty parameters of the items and the proficiency levels required for correct answers was also examined. We found a significant proportion of items that require surface learning. Surface learning predominated even among items with high levels of difficulty and discrimination and at proficiency levels considered adequate and advanced.

KEYWORDS EXTERNAL ASSESSMENT • COMPLEXITY LEVELS • ITEM ANALYSIS.

Recebido em: 6 SETEMBRO 2024

Aprovado para publicação em: 21 MAIO 2025



Este é um artigo de acesso aberto distribuído nos termos da licença Creative Commons do tipo BY.

INTRODUÇÃO

Este estudo¹ investiga a complexidade cognitiva de itens de matemática por meio de uma taxonomia denominada Structure of Observed Learning Outcomes (Solo). Quando falamos em complexidade cognitiva, estamos mencionando as operações mentais que esse item requer dos alunos em sua resolução. A complexidade cognitiva está relacionada a uma avaliação qualitativa da aprendizagem, ou seja, à qualidade da resposta exigida em um item. É importante destacar ainda que a complexidade cognitiva de uma questão não determina suas respostas, pois podem surgir resoluções além ou aquém do que foi exigido. Entretanto a forma como a questão está estruturada requer do aluno uma demanda cognitiva específica pela qual ele será avaliado. Assim, a complexidade cognitiva do item é uma informação pedagógica a mais quando vamos interpretar os resultados dos estudantes.

Os itens analisados integram um conjunto de avaliações já aplicadas em testes de edições anteriores da Prova Brasil,² que estavam disponíveis em uma plataforma digital do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), intitulada Devolutivas Pedagógicas das Avaliações de Larga Escala. Essa plataforma disponibilizava a professores e escolas itens com comentários pedagógicos, com o intuito de facilitar a interpretação dos resultados das avaliações externas. Cada item era composto de várias informações, como a habilidade avaliada, um comentário pedagógico e dados estatísticos obtidos a partir das respostas dos alunos (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [Inep], 2015b).

A plataforma contava com itens comentados referentes às disciplinas de língua portuguesa e matemática, para o 5º e o 9º anos do ensino fundamental e o 3º ano do ensino médio. No projeto inicial, o intuito era acrescentar mais itens ao longo do tempo, para formar um grande banco de dados de domínio público. A plataforma foi uma iniciativa do Inep com a parceria de colaboradores.³ O projeto, lançado em 6 de agosto de 2015, foi criado com o intuito de envolver professores e gestores da educação básica, visando a aproximá-los da linguagem das avaliações externas. O site do Inep também apresentava, por exemplo, vídeos explicativos sobre a Teoria Clássica dos Testes e sobre a Teoria de Resposta ao Item (Inep, 2015b). Entretanto essa plataforma foi tirada do ar.

De acordo com Santos Fontanive et al. (2007), a compreensão e apropriação dos resultados das avaliações externas pelos professores, gestores e equipes de

1 Este texto é fruto da dissertação de mestrado da autora Solange Maria Mol (2019).

2 Apesar da recente mudança de nomenclatura do Sistema de Avaliação da Educação Básica (Saeb), optamos por manter o nome da avaliação da época na qual os itens foram criados e aplicados.

3 Foram colaboradores a ONG Todos pela Educação, a Associação Brasileira de Avaliação Educacional (Abave), a Fundação Lemann, o Itaú BBA e o Instituto Unibanco (Inep, 2015b).

secretarias de educação se configuram um desafio, visto que a forma como essas informações são apresentadas geralmente é pouco familiar a esses sujeitos. Para Gatti (2009), o grande desafio das avaliações externas é a apropriação pedagógica por parte das escolas, no sentido de orientar as atividades de ensino e tornar mais relevante a aprendizagem dos alunos. As avaliações externas têm muito a contribuir para o trabalho dos docentes na escola. Quando interpretados corretamente, seus resultados oferecem informações relevantes sobre o aprendizado dos alunos nos conteúdos básicos. Entretanto, quando o resultado dessas avaliações não é compreendido, fontes importantes de informações deixam de ser usadas em prol do desenvolvimento dos alunos.

Portanto este artigo é uma contribuição ao alargamento e visibilidade dos estudos sobre a complexidade cognitiva de itens. Seu objetivo é analisar os itens de matemática do 5º e 9º anos do ensino fundamental da Prova Brasil pertencentes à plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala, por meio dos níveis de complexidade cognitiva da taxonomia Solo. Nesse sentido, classificamos e analisamos os itens a partir da dificuldade (nível de proficiência e índice de dificuldade) e da discriminação (índice de discriminação). Também relacionamos e diferenciamos a dificuldade e a complexidade cognitiva dos itens, conceitos tratados separadamente pela Solo.

TAXONOMIA SOLO

Com o objetivo de aprimorar o processo de ensino e aprendizagem, pesquisadores vêm criando instrumentos como as taxonomias cognitivas. A mais conhecida no campo educacional é a taxonomia de Bloom (Anderson & Krathwohl, 2001), entretanto existem diversas taxonomias, como a de Marzano (Irvine, 2017); a Depth of Knowledge (Webb, 2002); taxonomias cognitivas do Programa Internacional de Avaliação dos Estudantes (Pisa) (Organisation for Economic Co-operation and Development [OECD], 2017); dentre outras.

A taxonomia utilizada nesta pesquisa é a Solo, que foi criada pelos australianos Jhon Biggs e Kevin Collis e publicada em 1982. Ela pode ser usada de diversas formas, como para analisar a estrutura das respostas de alunos em tarefas escolares, na criação de itens ou atividades, na avaliação de estudantes, na formação de docentes, como método de pesquisa, na abordagem sobre estratégias de aprendizagem, em devolutivas pedagógicas, dentre outras. É uma taxonomia de fácil entendimento, que pode ser empregada em contextos variados. De acordo com Wells (2015), ao fazer um diagnóstico do nível cognitivo em que um estudante se encontra em determinado conhecimento, a Solo permite que o professor faça uma adaptação do processo de ensino e dos materiais pedagógicos para promover desafios apropriados

a ele. No campo da avaliação educacional, a Solo pode ser usada tanto na avaliação externa quanto na avaliação de sala de aula. Neste trabalho, analisamos itens de avaliações externas.

A taxonomia Solo se caracteriza por um sistema de categorias que tem o intuito de analisar a estrutura de respostas de alunos, visando a identificar o tipo de pensamento demonstrado (Biggs & Collis, 1982). De acordo com Irvine (2021), diferentemente de outras taxonomias cognitivas, a Solo não associa a qualidade da aprendizagem com a aquisição de conhecimentos generalizados, mas sim com a informação que se tem como evidência a partir de como uma resposta ou pergunta está estruturada. Nesse sentido, o alvo é a resposta que foi fornecida em um dado momento. Portanto o objetivo da Solo não é descobrir como um aluno opera de forma mais ampla (estágio de desenvolvimento cognitivo), mas sim de que forma ele operou em uma tarefa específica, sem fazer maiores generalizações sobre o seu funcionamento cognitivo.

Biggs e Collis (1982) organizaram a Solo em cinco níveis hierárquicos, que crescem em complexidade cognitiva:

- Pré-estrutural – a resposta elaborada é inadequada, o sujeito não atinge o mínimo esperado na tarefa (ideia incorreta);
- Uniestrutural – um elemento relevante da tarefa é apontado ou entendido em série, de forma independente (uma ideia);
- Multiestrutural – dois ou mais elementos relevantes da tarefa são apontados, entretanto esses aspectos não são relacionados, mas tratados de forma independente (múltiplas ideias);
- Relacional – vários elementos relevantes da tarefa são apontados, avaliados e relacionados, o que forma uma estrutura coerente (relacionamento entre ideias);
- Abstrato estendido – ocorre a generalização de uma estrutura coerente para um patamar mais abstrato, no qual a resolução (de uma tarefa) vai além das informações fornecidas previamente, podendo ser aplicada em outros contextos (extensão das ideias).

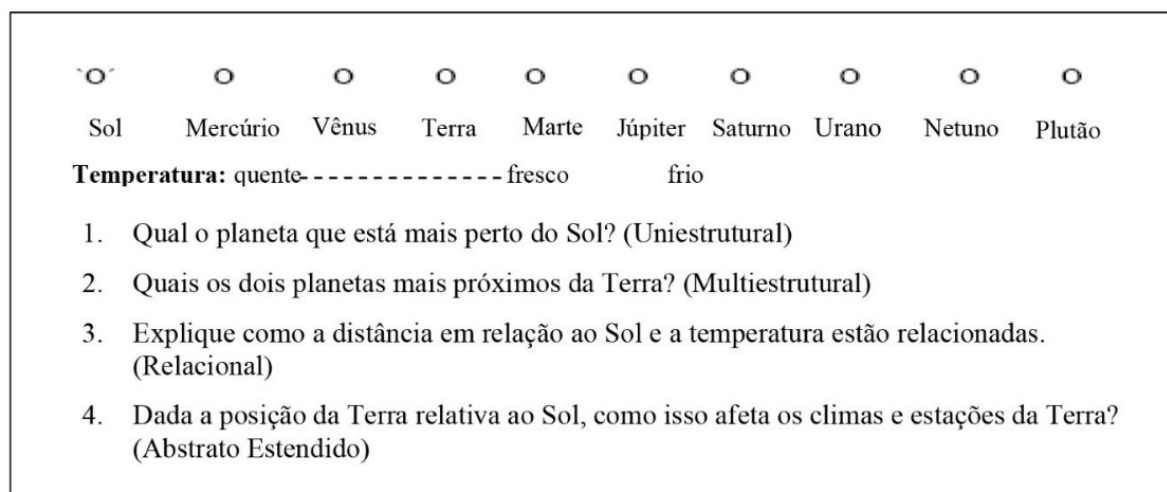
Adicionalmente, os níveis de complexidade cognitiva são divididos em duas categorias de aprendizagem: superficial e profunda. A aprendizagem superficial engloba os níveis uniestrutural e multiestrutural, que são considerados mais simples e não exigem relações ou abstrações. Nesse tipo de aprendizagem, os níveis mudam apenas de forma quantitativa, aumentando o número de informações: nível uniestrutural (uma ideia) e multiestrutural (várias ideias não relacionadas). Já a aprendizagem profunda é composta dos níveis relacional e abstrato estendido.

Nessa aprendizagem acontecem mudanças qualitativas, pois é necessário um entendimento mais complexo, capaz de fazer relação entre as informações disponíveis, bem como realizar teorizações e generalizações. O conteúdo precisa ser pensado em um nível cognitivo mais complexo. O nível pré-estrutural não faz parte dessas categorias. Como ele envolve uma ideia incorreta, não indica nenhuma construção de aprendizagem (Hattie & Brown, 2004). Nesse sentido, a Solo possui uma estrutura de categorias que permitem avaliar desde elementos quantitativos básicos até pensamentos profundos de complexidade superior (Irvine, 2021).

Vale ainda destacar um ponto importante: é necessário um equilíbrio entre os dois tipos de aprendizagem. Trata-se de um equívoco pensar que a única aprendizagem que importa é a profunda e que a aprendizagem superficial é simples e pode ser ignorada. Para operar em níveis de complexidade cognitiva mais elevados (relações e abstrações), precisamos, antes, aprender elementos e conceitos básicos por meio dos níveis uni e multiestrutural. Quando desenvolvemos somente uma aprendizagem superficial, a consequência é um conhecimento fácil e raso. Por outro lado, se fomentamos apenas a profundidade, o resultado pode ser uma aprendizagem muito difícil para o sujeito. Portanto os dois tipos de aprendizagem são importantes e se complementam (Hattie & Brown, 2004).

A Figura 1 exemplifica perguntas classificadas pelos níveis de complexidade cognitiva da Solo.

FIGURA 1
Taxonomia Solo aplicada a perguntas



Fonte: Hattie e Brown (2004, p. 15, tradução própria).

Na Figura 1, a questão uniestrutural exige que o estudante identifique uma informação fornecida na representação ordenada dos planetas. A multiestrutural somente aumenta o número de informações solicitadas (dois planetas), mas sem

ainda estabelecer relações. A questão relacional solicita que sejam relacionadas duas informações: a posição dos planetas em relação ao Sol e a temperatura. Já a questão de nível abstrato estendido solicita que o estudante encontre o princípio geral que explica como a posição da Terra em relação ao Sol afeta os climas e estações. Nesse sentido, essa questão demanda que o estudante vá além das informações fornecidas no enunciado.

Outro ponto bastante relevante na taxonomia Solo é tratar os conceitos de profundidade/complexidade e dificuldade de forma separada. Embora seja comum que perguntas aumentem em dificuldade à medida que aumenta sua complexidade, isso não é uma regra absoluta. Nesse sentido, na Solo, um conteúdo difícil não é determinante para um pensamento profundo (complexo). A complexidade está relacionada com as demandas cognitivas necessárias para realizar uma dada tarefa, o raciocínio. Dessa forma, pode haver itens superficiais (pouco complexos cognitivamente) e que, ao mesmo tempo, são difíceis (Hattie & Brown, 2004).

Os níveis relacional e abstrato estendido não são necessariamente os que apresentam maior dificuldade. Assim, existe a possibilidade de um item uniestrutural ser mais difícil de resolver do que um item relacional, dependendo da forma como é elaborado. A dificuldade é um conceito mais relacionado com a análise empírica, enquanto a profundidade/complexidade está mais associada à análise teórica. Essa distinção entre tais conceitos pode ser considerada uma contribuição importante da Solo, pois a confusão entre eles pode levar a interpretações errôneas sobre tarefas (Hattie & Brown, 2004).

O Pisa indica que, não raro, a dificuldade empírica é confundida com a demanda cognitiva. Enquanto a dificuldade empírica pode ser obtida, por exemplo, pelo percentual de acertos nos itens, a demanda cognitiva envolve o tipo de processo mental necessário para responder ao item. Portanto a dificuldade dos itens não nos informa necessariamente sobre a sua complexidade. Dessa forma, para termos uma visão mais significativa do tipo de aprendizagem promovida, é necessário, além de discriminar itens fáceis e difíceis por meio de indicadores empíricos, também compreender a demanda desses itens em termos de complexidade cognitiva. O Pisa, além das habilidades e competências descritas em cada nível da sua escala de proficiência, inclui também informações sobre os processos cognitivos que os estudantes são capazes de realizar (OECD, 2017).

O uso da taxonomia Solo para analisar itens de avaliações externas oferece a possibilidade de olhar para esses instrumentos de um ponto de vista mais profundo. Alves e Soares (2013) defendem que o uso de testes de larga escala para aferição de sistemas educacionais foi um passo importante no sentido de associar essa qualidade a diagnósticos de aprendizagem, e não mais apenas em acesso ou disfunção no

rendimento/fluxo, como em outros momentos (Oliveira & Araujo, 2005). Contudo a qualidade desses testes é um elemento importante para garantir que os resultados sejam ainda mais coerentes com a realidade. Por isso a complexidade cognitiva não deve ser ignorada.

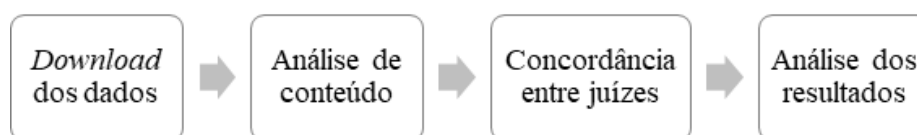
Em síntese, no caso da nossa pesquisa, a taxonomia Solo traz contribuições no trabalho com as avaliações externas. Ela pode auxiliar na elaboração dos itens, permitindo que os formuladores dos testes equilibrem questões superficiais e profundas, e também pode ser usada na análise dos resultados (interpretação pedagógica). Hattie e Brown (2004) apresentam um exemplo de aplicação da Solo nas avaliações externas. O Assessment Tools for Teaching and Learning (Asttle) é um *software* de avaliação usado na Nova Zelândia, que utiliza a taxonomia Solo na classificação de itens. É uma ferramenta disponível a todas as escolas públicas do país que permite a elaboração de testes. Dentre outras informações, os itens são classificados em aprendizagem superficial e profunda.

Assim, a principal contribuição da Solo, para os professores, se dá pela sugestão de um trabalho que progride em profundidade de conhecimento. Para isso, o professor deve levar seus alunos a avançar pelos níveis de complexidade cognitiva, partindo de uma aprendizagem superficial rumo a uma mais profunda.

MÉTODO

Nosso objeto de estudo é um conjunto de itens que já fizeram parte da Prova Brasil e que receberam tratamento estatístico e análise pedagógica dos resultados. Os itens foram obtidos na plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala, que tinha 467 itens distribuídos entre língua portuguesa e matemática. Nessa pesquisa, analisamos 123 itens de matemática do ensino fundamental, sendo 57 referentes ao 5º ano e 66 ao 9º ano. A Figura 2 indica as etapas da pesquisa.

FIGURA 2
Etapas seguidas na pesquisa



Fonte: Elaboração dos autores.

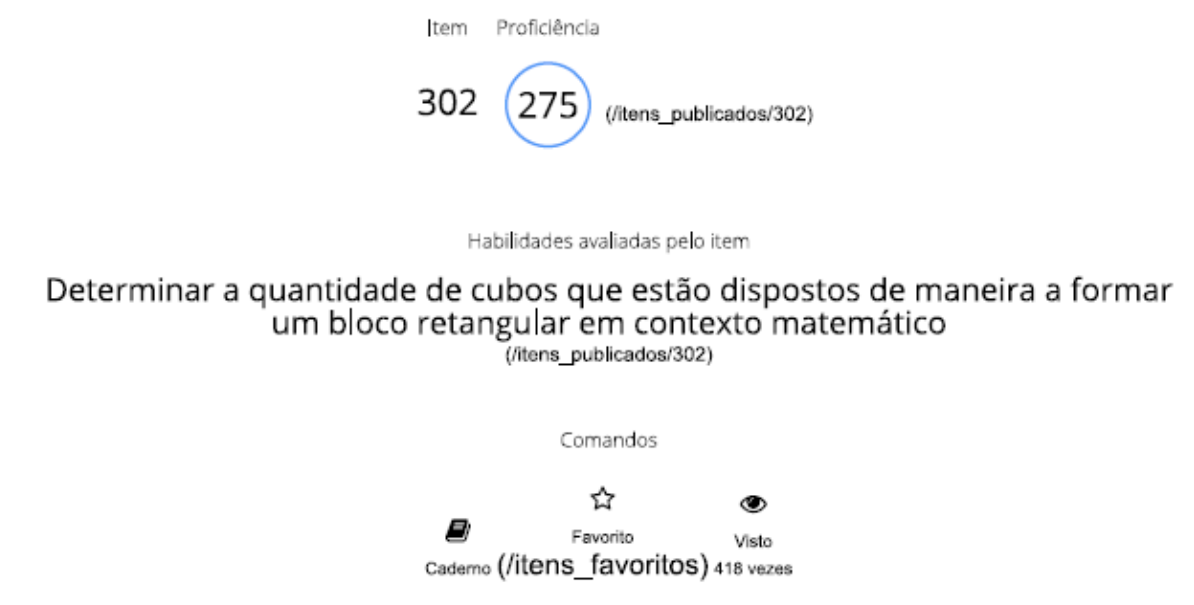
Os itens da plataforma foram baixados em formato PDF, individualmente, e organizados por nível de proficiência. Cada um desses documentos apresenta a habilidade que o item exige, a questão tal como foi aplicada, o comentário pedagógico

e as informações estatísticas. Vale ressaltar que os 123 itens que compõem o *corpus* da pesquisa correspondem à totalidade de itens de matemática do ensino fundamental que estavam disponíveis na plataforma. A opção por tal área do conhecimento se deu por uma necessidade de recorte metodológico devido ao período de tempo disponível para a realização da pesquisa.

Para a análise teórica sobre a taxonomia Solo, iniciamos um grupo de estudos que, além dos dois autores deste trabalho, contou com outra pesquisadora. As atividades realizadas no grupo foram usadas para um aperfeiçoamento da análise de questões. Os integrantes analisaram e discutiram itens diversos. É importante ressaltar que os itens da plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala não foram usados nesse momento. Selecionamos questões de avaliação de sala de aula e exemplos de itens da Prova Brasil encontrados no *site* do Inep.

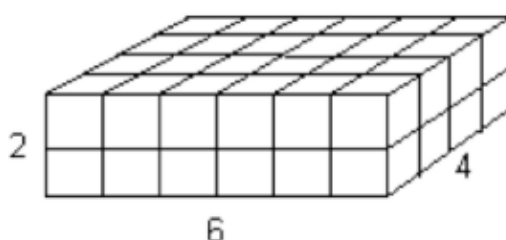
Para investigar os itens da plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala, inicialmente, usamos a análise de conteúdo como estratégia metodológica, uma vez que os itens seriam classificados em um dos níveis de complexidade cognitiva da taxonomia Solo. Assim, as categorias de análise são os quatro níveis de complexidade cognitiva da Solo: uniestrutural; multiestrutural; relacional; e abstrato estendido. É importante destacar que a decisão sobre a categoria atribuída a cada item foi tomada com base em todas as informações disponíveis do item que já apresentamos previamente: pergunta, resposta e comentário pedagógico. Em outras palavras, não se trata de uma análise apenas da questão em si, mas desse conjunto de informações, a fim de decidir sobre cada categoria.

A seguir, apresentamos a estrutura dos itens por meio de um exemplo retirado da plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala. Cada item comentado é dividido em quatro partes: sentença descritora; questão/item; comentário pedagógico; e estatísticas. A Figura 3 apresenta um exemplo de item de matemática do 9º ano.

FIGURA 3**Exemplo de item da plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala**

Item

Letícia guardou numa caixa 2 camadas de cubos, cada camada com 4 fileiras de 6 cubos.



Quantos cubos Letícia guardou nessa caixa?

- (A) 12
- (B) 24
- (C) 44
- (D) 48

Comentário pedagógico

O item aborda a concepção inicial de volume com base na contagem do número de cubos idênticos dispostos organizadamente em 2 camadas de 4 fileiras de 6 cubos. O enunciado solicita o número de cubos existentes na formação apresentada na figura. Há diferentes estratégias de contagem, 1) Contar o número de cubos que há em cada camada, fazendo $4 \times 6 = 24$ para, em seguida, duplicar essa quantidade, 2) Contar os cubos de cada pilha de fileira ($6 \times 2 = 12$ cubos) e, em seguida, multiplicar essa quantidade pelo número de fileiras, fazendo $4 \times 12 = 48$ cubos, 3) Contar as carreiras formadas por $4 \times 2 = 8$ cubos para depois multiplicar essa quantidade por 6 carreiras, obtendo $6 \times 8 = 48$ cubos. Alunos que tenham a noção sobre volume já consolidada podem tomar um desses cubos como unidade de volume, reconhecer que a formação dada na figura corresponde a um bloco retangular e calcular seu volume por meio do produto de suas três dimensões ($6 \times 4 \times 2 = 48$ cubos).

(continua)

(continuação)

O item foi acertado por metade dos participantes, sendo que o gabarito (D) se tornou a alternativa preferencial para aqueles situados no nível 225 da escala de proficiência. Foi alta a capacidade de discriminação do item, tendo em vista que acertaram a resposta 86% dos alunos do grupo de maior desempenho e somente 21% dos que estão no grupo de menor desempenho. Em relação aos distratores, vale destacar que o B obteve 24% das marcações, principalmente entre os estudantes com proficiência até 250. Eles, provavelmente, contaram apenas os cubos da camada superior, sem reconhecer que há uma inferior com outros 24 cubos, que não estão totalmente visíveis na figura. O distrator A atraiu 14% dos respondentes que, possivelmente, contaram apenas os 12 cubos frontais, cometendo um equívoco semelhante ao dos que escolheram a alternativa B. Já o distrator C obteve 8% das escolhas, e só foi capaz de atrair significativamente alunos de baixa proficiência, que teriam contado não os cubos, mas sim os quadrados que estão visíveis na figura apresentada.

Estatísticas

Nível do item na escala: 275

Posição do item na escala: 250

Coeficientes estatísticos clássicos do item

ÍNDICES						PROPORÇÕES DE RESPOSTA			
GAB	DIFI	DISCR	ABAI	ACIM	BISSE	A	B	C	D
D	0.5	0.66	0.21	0.86	0.63	0.14	0.24	0.08	0.5

COEFICIENTES BISSERIAIS			
A	B	C	D
-0,39	-0,3	-0,26	0,63

Gráficos de Teoria de Resposta ao Item (TRI)



Proficiência	150	175	200	225	250	275	300	325	350	375	400
Proporções de acertos empíricos	0.07	0.09	0.22	0.45	0.57	0.71	0.86	0.93	0.99	1.0	0.99

Fonte: Plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala (Inep, 2015a).

Quanto à concordância entre juízes, usamos o termo juiz como sinônimo de avaliador. Nesse sentido, “em situações nas quais mais de um juiz (avaliador) participa de um processo avaliativo, existem estratégias para verificar qual o grau de concordância desses juízes” (Matos, 2014, p. 301). São estratégias que permitem identificar em que proporção esses juízes concordam em suas avaliações. Tal artifício possibilita ao pesquisador mais segurança em seus resultados, dando mais confiabilidade ao estudo.

Nesse sentido, duas pesquisadoras classificaram de forma independente os itens da plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala nos níveis da taxonomia Solo (a primeira autora desta pesquisa e uma professora, mestra em Educação). Após essa primeira análise, calculamos a concordância entre juízes. Para verificar a concordância das classificações, calculamos dois coeficientes: o coeficiente Kappa de Cohen – K (Cohen, 1960) e o coeficiente Alfa de Krippendorff (Hayes & Krippendorff, 2007).

Após o cálculo dos coeficientes de concordância entre juízes, os resultados da análise de conteúdo (níveis de complexidade da Solo) foram discutidos e, quando necessário, revistos. Em seguida, foi verificada a relação entre os níveis da Solo e os parâmetros de discriminação e dificuldade dos itens, bem como a relação com os níveis de proficiência necessários para os acertos dos itens.

Vale destacar que classificamos os níveis de proficiência do Sistema de Avaliação da Educação Básica (Saeb) em uma estrutura de quatro categorias, elaborada por Soares (2009). No Saeb, os itens são alocados em uma escala de proficiência que mede o desempenho dos alunos por meio de parâmetros calculados com base na teoria de resposta ao item. A escala do Saeb é estruturada em um intervalo de 0 a 500, com níveis de dificuldade que variam de acordo com a etapa de ensino e a disciplina avaliada. A escala do 5º ano é dividida em 10 níveis de proficiência e a do 9º ano, em 9.

Entretanto a interpretação da escala do Saeb é de pouca familiaridade dos professores, pois apresenta muitos níveis e, apesar de indicar as habilidades associadas a cada um, sua apresentação é pouco clara. De acordo com Soares (2009), nessa escala, ao identificar o nível de proficiência dos alunos, é possível interpretar o que foi e o que não foi consolidado, porém não sabemos se essa proficiência é adequada. Assim, com o intuito de tornar essa escala mais compreensível, Soares (2009) criou intervalos aos quais atribuiu significados pedagógicos e normativos. A escala apresentada pelo autor é composta de quatro níveis: abaixo do básico; básico; adequado; e avançado. Nesse sentido, fizemos a transposição das escalas do Saeb de matemática do ensino fundamental (5º e 9º anos) para essa escala normativa. A Tabela 1 apresenta os pontos de corte usados.

TABELA 1**Níveis pedagógicos de desempenho em matemática por ano**

NÍVEL DE DESEMPENHO	LEGENDA	5º ANO	9º ANO
Abaixo do básico	Os estudantes neste nível apresentam domínio elementar da competência. É necessária a recuperação de conteúdo.	0 a 174 pontos	0 a 224 pontos
Básico	Os alunos neste nível demonstram domínio parcial da competência, precisam melhorar. Sugerem-se atividades de reforço.	175 a 224 pontos	225 a 299 pontos
Adequado	Os alunos neste nível demonstram dominar os conteúdos e as habilidades esperados para sua etapa escolar. Encontram-se preparados para continuar os estudos.	225 a 274 pontos	300 a 349 pontos
Avançado	Os alunos neste nível dominam a competência esperada em sua etapa. O aprendizado demonstrado vai além da expectativa.	Igual ou maior que 275 pontos	Igual ou maior que 350 pontos

Fonte: Soares (2009).

Além da diminuição do número de níveis, a proposta de Soares (2009) possibilita uma interpretação pedagógica mais clara dos resultados. Por sua vez, a interpretação clara dos resultados permite que ela seja mais útil aos professores na sala de aula.

RESULTADOS

Os itens foram analisados pelo nível de complexidade cognitiva da taxonomia Solo: uniestrutural, multiestrutural, relacional e abstrato estendido (variável denominada “níveis Solo” nas análises subsequentes). Posteriormente, os níveis da taxonomia Solo foram reagrupados nos tipos de aprendizagem: aprendizagem superficial (uniestrutural ou multiestrutural) e aprendizagem profunda (relacional ou abstrato estendido) (variável “tipo Solo” nas análises subsequentes).

Análise da concordância entre juízes

Os itens foram analisados por duas pesquisadoras independentes, com o intuito de verificar a concordância quanto à classificação nos níveis de complexidade cognitiva da Solo e, conseqüentemente, do tipo de aprendizagem. As duas especialistas que fizeram a classificação inicial tiveram maior dificuldade de concordância na classificação dos itens do 9º ano do que nos do 5º ano. Na sequência, todos os itens com classificação discordante foram discutidos e classificados em conjunto por um comitê de especialistas composto de quatro integrantes: além das duas juízas (a autora do trabalho e uma mestra em Educação), participaram o segundo autor deste estudo e uma pesquisadora que também trabalha com este tema. O grupo se

reuniu e discutiu todos os itens discordantes, 8 referentes ao 5º ano e 22 ao 9º ano. Nesse momento, as juízas originais apresentaram os argumentos para a classificação de cada item e os demais integrantes se posicionaram a respeito do item e das justificativas. A classificação final de cada questão só era feita quando todos estavam de acordo. Vale ressaltar que, mesmo que os resultados da concordância entre juízes fossem considerados excelentes, a revisão final do comitê também seria uma etapa necessária para todos os itens discordantes.

Análise da complexidade cognitiva dos itens

A Tabela 2 apresenta a frequência na qual os itens foram classificados em cada um dos níveis de complexidade e em cada tipo de aprendizagem da Solo.

TABELA 2
Frequência dos níveis e tipos de aprendizagem Solo entre os itens

NÍVEL SOLO	FREQUÊNCIA	PERCENTUAL	TIPO SOLO	FREQUÊNCIA	PERCENTUAL
5º ano					
Uniestrutural	44	77,2	Aprendizagem superficial	57	100,0
Multiestrutural	13	22,8			
Relacional	0	0,0	Aprendizagem profunda	0	0,0
Abstrato estendido	0	0,0			
Total	57	100,0	Total	57	100,0
9º ano					
Uniestrutural	21	31,8	Aprendizagem superficial	62	93,9
Multiestrutural	41	62,1			
Relacional	4	6,1	Aprendizagem profunda	4	6,1
Abstrato estendido	0	0,0			
Total	66	100,0	Total	66	100,0

Fonte: Elaboração dos autores com base nos dados da pesquisa.

No 5º ano, encontramos apenas itens uni e multiestruturais. O nível uniestrutural apareceu em proporção bem expressiva (77,2%) em relação ao multiestrutural (22,8%). Esses dados indicam que os itens da Prova Brasil de Matemática do 5º ano do ensino fundamental, coletados na plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala, exigem dos alunos a resolução de tarefas simples, que requerem apenas o trato de uma (na maioria dos casos) ou mais informações isoladamente, sem necessidade de fazer relações de ideias para encontrar os resultados.

Quanto ao 9º ano, a maior parte dos itens foi classificada como multiestrutural (62,1%). No nível uniestrutural, encontramos 21 itens (31,8%), enquanto no

relacional foram classificados apenas 4 (6,1%). Chamou a atenção a baixa frequência do nível relacional, visto que relacionar informações é uma habilidade necessária a várias situações cotidianas, sendo essencial para um aprendizado de qualidade. Os resultados encontrados para o 9º ano foram convergentes com os obtidos na pesquisa de Filipe (2011), que analisou itens de matemática do 9º ano de avaliações externas em Portugal. O nível mais recorrente na pesquisa da autora também foi o multiestrutural, sendo identificados poucos itens relacionais e nenhum no nível abstrato estendido.

A frequência dos tipos de aprendizagem da taxonomia Solo também é exibida na Tabela 3. Todos os itens do 5º ano foram classificados no tipo de aprendizagem superficial, pois, nessa série, só encontramos itens nos níveis uni e multiestrutural. Já para o 9º ano, dos 66 itens analisados, 62 (93,9%) foram classificados como aprendizagem superficial e apenas 4 (6,1%) como aprendizagem profunda. Ou seja, apesar de existirem itens de aprendizagem profunda, eles aparecem em uma proporção muito pequena. Se compararmos o 5º e o 9º anos, podemos afirmar que os itens do 9º ano apresentam maior complexidade cognitiva. Primeiro, porque o 9º ano apresenta itens de aprendizagem profunda, ainda que em pouca quantidade. Além disso, o nível que mais apareceu no 9º ano foi o multiestrutural, enquanto no 5º prevaleceu o uniestrutural (nível mais elementar). Possíveis motivos para essas diferenças são as etapas de ensino distintas e a faixa etária dos alunos. No entanto isso não pode ser considerado uma justificativa para a ausência de itens de aprendizagem profunda no 5º ano e a baixa frequência deles no 9º ano.

De forma geral, esses resultados indicam que os itens analisados exigem aprendizagem de baixa complexidade cognitiva em quase a sua totalidade. Hattie e Brown (2004) defendem um equilíbrio entre os dois tipos de aprendizagem. Portanto a baixa frequência de questões de matemática que requerem aprendizagem profunda sugere que os testes nacionais dessa disciplina do 5º e do 9º ano do ensino fundamental estão exigindo dos alunos resoluções de tarefas pouco complexas cognitivamente. Isso precisaria ser mais bem investigado, uma vez que não sabemos se as questões da plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala são representativas da Prova Brasil como um todo. No entanto elas parecem ser, pois estão representados itens de todos os níveis de proficiência.

Na pesquisa de Filipe (2011), que investigou itens de matemática do 9º ano de avaliações externas em Portugal, foram encontrados resultados semelhantes aos nossos: 8 itens de aprendizagem profunda e 124 de aprendizagem superficial.

Análise por faixa de dificuldade

As faixas de dificuldade foram obtidas por meio do índice de dificuldade do item (Difi), que representa o percentual de acertos empíricos da questão. Assim, os

itens que possuem Difi acima de 65% são considerados fáceis, os que têm Difi de 30% a 65% são medianos (médios) e aqueles com Difi abaixo de 30% são difíceis. Nesse sentido, separamos a dificuldade dos itens em três faixas: difícil, médio e fácil. A Tabela 3 apresenta a distribuição dos níveis de complexidade cognitiva da taxonomia Solo e dos tipos de aprendizagem a partir da faixa de dificuldade.

TABELA 3

Distribuição dos níveis e tipos de aprendizagem Solo por faixa de dificuldade

DIFICULDADE	N. DE ITENS	UNI- ESTRUTURAL (N. - %)	MULTI- ESTRUTURAL (N. - %)	RELACIONAL (N. - %)	APRENDIZAGEM SUPERFICIAL (N. - %)	APRENDIZAGEM PROFUNDA (N. - %)
5º ano						
Difícil	14	11 - 78,6	3 - 21,4	0 - 0,0	14 - 100,0	0 - 0,0
Médio	29	21 - 72,4	8 - 27,6	0 - 0,0	29 - 100,0	0 - 0,0
Fácil	14	12 - 85,7	2 - 14,3	0 - 0,0	14 - 100,0	0 - 0,0
9º ano						
Difícil	22	5 - 22,7	14 - 63,6	3 - 13,6	19 - 86,4	3 - 13,6
Médio	31	10 - 32,3	20 - 64,5	1 - 3,2	30 - 96,8	1 - 3,2
Fácil	13	6 - 46,2	7 - 53,8	0 - 0,0	13 - 100,0	0 - 0,0

Fonte: Elaboração dos autores com base nos dados da pesquisa.

Na Tabela 3, alguns grupos de classificações podem ser considerados inesperados. Mesmo sabendo que os conceitos de complexidade e de dificuldade são diferentes, existe uma expectativa/tendência de que os itens difíceis também sejam mais complexos. Entretanto, no 5º ano, todos os itens analisados foram classificados como superficiais. Na faixa difícil, de um total de 14 itens, 11 foram classificados no nível uniestrutural. Esse é um resultado inesperado, pois foram questões pouco acertadas pelos alunos, mas que possuem a demanda cognitiva mais baixa possível.

O 9º ano revelou mais itens na faixa média (31), entretanto o número de itens na faixa difícil foi expressivo (22). Dos itens da faixa difícil, 86,4% (19) exigiram aprendizagem superficial, sendo 5 classificados como uniestruturais e 14 como multiestruturais. Assim como no 5º ano, observamos um considerável número de itens com dificuldade empírica alta, mas que exigem aprendizagem superficial. A Figura 4 exibe um exemplo de item difícil e com baixa demanda cognitiva para o 9º ano.

FIGURA 4
Item 361 (9º ano)

Marta foi comprar um pedaço de queijo. No supermercado ela encontrou 4 pedaços de queijo com as seguintes etiquetas:

0,545 kg	0,6 kg	0,38 kg	0,375 kg
(A)	(B)	(C)	(D)

Qual desses pedaços de queijo tem o menor peso?

(A) A
(B) B
(C) C
(D) D

Fonte: Plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala (Inep, 2015a).

O item 361 está localizado no nível de desempenho adequado. O Difi do item é 0,2, considerado difícil. Ele foi classificado como uniestrutural, visto que exige apenas identificar uma informação dada (peso), o que, consequentemente, envolve aprendizagem superficial. Esse é um exemplo claro em que um item é difícil sem ser complexo. Por um lado, a dificuldade se encontra relacionada ao número de casas decimais, visto que 61% dos alunos marcaram a opção B (provavelmente, eles imaginaram ser o menor peso, por apresentar apenas uma casa após a vírgula). Por outro lado, não se trata de um item complexo, pois, tendo claro o funcionamento da forma decimal, basta que o aluno identifique o menor peso (alternativa D).

Vale destacar que quatro itens foram classificados no nível relacional no 9º ano: três na faixa difícil e um na faixa média. Segundo Hattie e Brown (2004), apesar de os conceitos de profundidade/complexidade e dificuldade não serem iguais, eles costumam estar associados (existe uma tendência nesse sentido). O fato de nenhum item relacional ser classificado na faixa fácil e de três itens estarem na faixa difícil indica essa tendência de que questões classificadas nos níveis de maior complexidade da Solo têm maior probabilidade de serem também mais difíceis. No entanto isso não muda o fato de que, nos nossos resultados, a quantidade de itens de aprendizagem profunda é muito baixa.

Os resultados indicam claramente que os conceitos de profundidade/complexidade e de dificuldade são diferentes. Se partíssemos do pressuposto de que os dois conceitos são sinônimos, seria esperado que os itens da faixa difícil fossem classificados, em sua maioria, nos níveis de complexidade correspondentes à aprendizagem profunda. Entretanto os dados mostram que não houve nenhum item de aprendizagem profunda para o 5º ano e que, para o 9º ano, apenas 13,6% dos itens

difíceis foram classificados em um nível de complexidade cognitiva alta. Portanto uma questão pode ser difícil sem ser profunda/complexa e vice-versa.

Análise por nível de proficiência

Utilizamos os níveis normativos propostos por Soares (2009): abaixo do básico; básico; adequado; e avançado. A distribuição do número de itens, do nível de complexidade cognitiva da Solo e do tipo de aprendizagem por nível de proficiência é apresentada na Tabela 4.

TABELA 4

Distribuição dos níveis e tipos de aprendizagem Solo por nível de proficiência

NÍVEL DE PROFICIÊNCIA	N. DE ITENS	UNI-ESTRUTURAL (N. - %)	MULTI-ESTRUTURAL (N. - %)	RELACIONAL (N. - %)	APRENDIZAGEM SUPERFICIAL (N. - %)	APRENDIZAGEM PROFUNDA (N. - %)
5º ano						
Abaixo do básico	9	9 - 100,0	0 - 0,0	0 - 0,0	9 - 100,0	0 - 0,0
Básico	18	13 - 72,2	5 - 27,8	0 - 0,0	18 - 100,0	0 - 0,0
Adequado	17	10 - 58,8	7 - 41,2	0 - 0,0	17 - 100,0	0 - 0,0
Avançado	13	12 - 92,3	1 - 7,7	0 - 0,0	13 - 100,0	0 - 0,0
9º ano						
Abaixo do básico	3	3 - 100,0	0 - 0,0	0 - 0,0	3 - 100,0	0 - 0,0
Básico	26	9 - 34,6	17 - 65,4	0 - 0,0	26 - 100,0	0 - 0,0
Adequado	17	6 - 35,3	9 - 52,9	2 - 11,8	15 - 88,2	2 - 11,8
Avançado	20	3 - 15	15 - 75	2 - 10,0	18 - 90,0	2 - 10,0

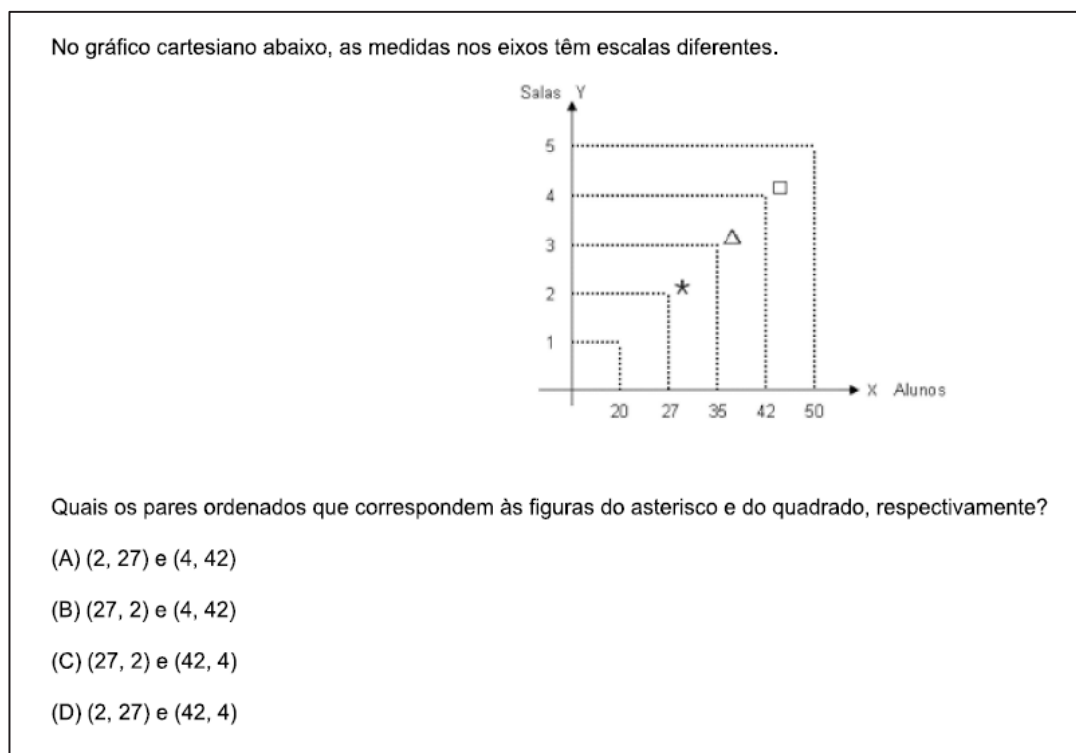
Fonte: Elaboração dos autores com base nos dados da pesquisa.

Para o 5º ano, todos os níveis de proficiência estão relacionados à aprendizagem superficial. O maior número de itens está concentrado nos níveis básico (18) e adequado (17). Um dado importante é o número de itens posicionados no nível abaixo do básico (9), pois se trata de um nível em que os estudantes apresentam um domínio muito elementar de uma competência. Para o 5º ano, somando-se os itens desse nível com os do nível básico (os dois que estão abaixo do considerado adequado), encontramos 27 itens. Isso revela que quase metade do total de itens analisados avalia competências abaixo do adequado para essa etapa. Por outro lado, os 13 itens do nível avançado chamam ainda mais a atenção, pois todos envolvem aprendizagem superficial e 12 foram classificados como uniestruturais. Isso significa que mesmo os itens considerados avançados (aqueles que exigem aprendizagem além da expectativa) são superficiais e requerem o uso de uma informação dada ou a realização de uma única operação. Ou seja: são itens com uma demanda cognitiva baixa. Esse é um resultado inesperado.

Já no 9º ano, o maior número de itens está posicionado no nível básico (26), seguido pelo avançado (20), enquanto apenas 3 estão no abaixo do básico. Os níveis adequado e avançado concentram 37 itens. O 9º ano apresenta itens superficiais e profundos, sendo que os profundos estão posicionados nos níveis adequado e avançado, mais especificamente 2 itens em cada um. Assim como nas faixas de dificuldade, quando comparamos os itens classificados como relacionais e os níveis de proficiência, observamos que o nível relacional da Solo só aparece nos níveis de proficiência superiores. Isso, mais uma vez, nos permite dizer que os itens que exigem o tipo de aprendizagem profunda, de acordo com a taxonomia Solo, têm uma maior probabilidade de aparecer nos níveis de proficiência (dificuldade) mais elevados. Apesar dessa tendência, em todos os níveis de proficiência, há predominância de aprendizagem superficial.

O grupo de itens do nível avançado do 9º ano também merece destaque. De um total de 20 itens nesse nível, 18 são superficiais (3 uniestruturais e 15 multiestruturais). Isso significa que, no 9º ano, o nível de proficiência mais avançado possível está associado a questões que exigem baixa demanda cognitiva, em sua maioria, multiestruturais. Ou seja, requerem o uso de duas ou mais informações de forma independente, sem fazer relacionamento entre elas. Esse também é um resultado inesperado. A Figura 5 mostra um exemplo de item desse grupo.

FIGURA 5
Item 189 (9º ano)



Fonte: Plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala (Inep, 2015a).

O item 189 está situado no nível avançado do 9º ano. Foi classificado como multiestrutural, de acordo com a taxonomia Solo, pois exige que os estudantes identifiquem duas informações dadas sem fazer relação entre elas. Para resolver é preciso localizar dois pares ordenados específicos em um gráfico cartesiano (alternativa C).

No nível de proficiência avançado, os alunos dominam a competência esperada em sua etapa e o aprendizado demonstrado vai além da expectativa. Entretanto a demanda cognitiva dos itens desse nível foi baixa. Mais uma vez, ficou clara a distinção entre os conceitos de profundidade/complexidade e dificuldade. Quando comparamos as duas etapas, percebemos que, provavelmente, como o 9º ano é uma série mais avançada, há um maior número de itens posicionados nos níveis mais avançados e apenas três no abaixo do básico.

Outros estudos apresentam resultados similares. Ao analisar a competência matemática dos alunos brasileiros no Pisa 2003, Nunes (2013) descobriu que as habilidades consolidadas estão relacionadas à reprodução de informações e que itens desse tipo estão posicionados nos níveis mais baixos da escala de proficiência. Em outras palavras, os resultados mostraram que os estudantes brasileiros só conseguiram acertar os itens mais simples e os níveis mais baixos da escala. Os resultados do Pisa de 2015 não divergem muito. De acordo com o Relatório Nacional do Pisa de 2015, a nota média dos estudantes brasileiros em matemática foi 377 pontos, muito inferior à dos estudantes dos países da OECD (490 pontos). Além disso, 70,3% dos estudantes estão abaixo do nível 2 na escala do Pisa, patamar considerado pela OECD como necessário para os jovens exercerem plenamente sua cidadania (Inep, 2020).

Ao comparar o Pisa e o Saeb, Nunes (2013) encontrou discrepâncias entre essas avaliações. Os itens dos testes do Pisa exigem dos estudantes a aplicação de conhecimentos, habilidades e competências necessários no cotidiano, enquanto o foco do Saeb está mais voltado para o domínio do conteúdo. Outra diferença é o formato dos testes: o Saeb é constituído apenas por itens de múltipla escolha, enquanto o Pisa inclui outros tipos de itens, inclusive de resposta construída aberta. Em suma, a comparação feita pela autora mostra, por meio do Saeb, que o ensino em nosso país está organizado em bases diferentes do que é esperado pelo Pisa.

Muri (2017), ao comparar o letramento científico de estudantes brasileiros e japoneses por meio dos resultados do Pisa, assinala que o bom desempenho do Japão no Pisa está relacionado à existência de um currículo nacional comum, à formação continuada de docentes e às reformas educacionais impulsionadas pelos resultados do Pisa. Por outro lado, o desempenho baixo do Brasil no Pisa pode ser associado ao despreparo dos estudantes brasileiros, à precária formação de professores, à pouca familiaridade dos estudantes com o teste e à limitada utilização das evidências fornecidas pelas avaliações externas.

Como destacado na pesquisa de Muri (2017), a falta de familiaridade com o teste do Pisa é um dos fatores relacionados ao baixo desempenho dos estudantes brasileiros, que estão acostumados com questões de múltipla escolha. O Pisa tem um padrão de exigência alto, se comparado ao Saeb. Tomados em conjunto, nossos resultados e os de outros estudos dão indícios de que os alunos brasileiros estão sendo pouco exigidos do ponto de vista da complexidade cognitiva.

Análise por discriminação do item

A variável discriminação foi obtida por meio do índice de discriminação (Discr). Ele é calculado a partir da diferença entre o percentual de acerto dos 27% de estudantes que tiveram melhor desempenho no teste e os 27% que tiveram pior desempenho. Santos Fontanive et al. (2007) consideram que um item possui baixa discriminação quando o valor do Discr é menor que 0,25, ou seja, significa que ele não separou adequadamente os estudantes de melhor e pior desempenho. Quando o Discr é superior a 0,25, significa que o item fez boa discriminação entre os alunos. Vale destacar que, neste artigo, optamos por tal critério de corte, que é o utilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Entretanto há outras formas de classificar a variável. A Tabela 5 apresenta os resultados de acordo com a discriminação.

TABELA 5
Distribuição dos níveis e tipos Solo por faixa de discriminação

DISCRIMINAÇÃO	N. DE ITENS	UNI- ESTRUTURAL (N. - %)	MULTI- ESTRUTURAL (N. - %)	RELACIONAL (N. - %)	APRENDIZAGEM SUPERFICIAL (N. - %)	APRENDIZAGEM PROFUNDA (N. - %)
5º ano						
Baixa discriminação	4	4 - 100,0	0 - 0,0	0 - 0,0	4 - 100,0	0 - 0,0
Alta discriminação	53	40 - 75,5	13 - 24,5	0 - 0,0	53 - 100,0	0 - 0,0
9º ano						
Baixa discriminação	7	2 - 28,6	4 - 57,1	1 - 14,3	6 - 85,7	1 - 14,3
Alta discriminação	59	19 - 32,2	37 - 62,7	3 - 5,1	56 - 94,9	3 - 5,1

Fonte: Elaboração dos autores com base nos dados da pesquisa.

Pelo critério adotado neste trabalho, houve predominância expressiva de itens de alta discriminação, tanto no 5º ano quanto no 9º ano. Isso nos revela que uma parte considerável dos itens analisados fez uma discriminação adequada entre os alunos de melhor e pior desempenho. No 5º ano, os itens de baixa discriminação correspondem apenas ao nível uniestrutural, enquanto nos de alta dis-

criminação 24,5% deles são multiestruturais. No 9º ano, foram encontrados itens dos três níveis de complexidade (uni, multi e relacional) nas faixas de alta e baixa discriminação. Destacamos que três dos quatro itens relacionais estão presentes nos itens de alta discriminação. Entretanto observamos também um item relacional (aprendizagem profunda) na faixa de baixa discriminação. A Figura 6 exhibe o referido item.

FIGURA 6**Item 587 (9º ano)**

A idade de Luís é o triplo da idade de seu filho. A soma das duas idades é 40 anos. O sistema que representa essa situação é

(A) $\begin{cases} x + 3 = y \\ x + y = 40 \end{cases}$

(B) $\begin{cases} x + 3x = y \\ x + y = 40 \end{cases}$

(C) $\begin{cases} x = 3y \\ x + y = 40 \end{cases}$

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

Fonte: Plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala (Inep, 2015a).

O item 587 está localizado no nível de desempenho avançado e foi classificado como relacional. O índice de discriminação da questão é 0,22, considerado baixo, ou seja, significa que o item não conseguiu separar adequadamente os estudantes de melhor e pior desempenho. Além disso, se trata de um item difícil, pois só 26% dos alunos o acertaram. A questão exige que o estudante descubra o sistema de equação a partir de uma situação apresentada. Para resolver é necessário relacionar as informações do enunciado (referentes às idades de Luís e do seu filho) e representá-las em forma de equação. Vale ressaltar que as duas equações do sistema são dependentes entre si (alternativa C). Nesse caso, houve mais acertos somente a partir do grupo de alunos com proficiência acima de 325 (avançado).

CONSIDERAÇÕES FINAIS

Nossos resultados sugerem que os itens de matemática do 5º e do 9º ano do ensino fundamental, aplicados em testes anteriores da Prova Brasil, exigem dos estudantes a realização de tarefas predominantemente superficiais e seu conjunto tem pouco

equilíbrio de complexidade cognitiva. Entretanto isso precisa ser mais bem investigado, pois não sabemos se a amostra da plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala é representativa. Apesar disso, outros estudos com avaliações externas encontraram resultados semelhantes aos nossos.

A partir da proporção de itens por nível de complexidade cognitiva e tipo de aprendizagem Solo, analisamos os resultados em relação a três características dos itens: faixa de dificuldade do item; nível de proficiência; e faixa de discriminação do item. O tipo de aprendizagem superficial prevaleceu, o que implica que esse tipo foi dominante, inclusive na faixa de dificuldade alta, nos níveis de proficiência considerados adequado e avançado, e na faixa alta de discriminação do item. A partir dessas constatações, ficou clara a distinção entre os conceitos de complexidade e de dificuldade. Constatamos que um número expressivo de alunos errou uma gama de itens de baixa complexidade cognitiva. Ou seja, de um lado, os testes de matemática do ensino fundamental oferecem aos estudantes altas proporções de itens superficiais para serem respondidos e, de outro, as análises estatísticas mostram que até itens que necessitam de níveis adequados ou mesmo avançados de proficiência para serem respondidos corretamente são de baixa complexidade cognitiva.

A expressiva proporção de itens que exigem aprendizagem superficial indica um alerta importante: apesar do avanço na qualidade da aprendizagem dos alunos brasileiros nos últimos anos, é preciso cautela com relação a essa evolução, pois ela pode estar associada a um padrão de exigência baixo. Considerados em conjunto, nossos resultados e os de outras pesquisas indicam que os estudantes brasileiros estão sendo pouco exigidos quanto à complexidade cognitiva.

Com base no exposto, nosso trabalho traz três contribuições principais. Primeiramente, os resultados apresentados fornecem indícios importantes de que os itens de matemática das avaliações externas do ensino fundamental requerem dos estudantes aprendizagem predominantemente superficial. Assim, nosso estudo contribui para entender o tipo de aprendizagem em matemática que estamos demandando dos nossos alunos, o que revela que é preciso avançar em políticas educacionais que contribuam para a melhoria do processo de ensino e aprendizagem como um todo, incluindo a formação de professores. Em relação às avaliações externas, elas também precisam ser repensadas. Gestores educacionais devem ficar atentos à forma como esses itens são construídos. Sugerimos a utilização da demanda cognitiva como parte integrante da construção dos itens, pois é uma informação importante que permite analisar os resultados com base também em sua complexidade.

A segunda contribuição é a apresentação da taxonomia Solo como uma ferramenta que tem muito a colaborar na avaliação educacional. Ela possibilita vários usos, para além do que foi feito nesta pesquisa. Sua principal contribuição para os professores é no sentido de levar os alunos a atingirem níveis cada vez mais

profundos de aprendizagem. Nesse sentido, os conhecimentos sobre a Solo (complexidade cognitiva) podem auxiliar o professor em diversas dimensões da sua prática pedagógica: o que ensinar, como ensinar e como avaliar. Com um referencial desse tipo à sua disposição, o docente pode preparar suas aulas de uma forma mais consciente. Por exemplo: preparar uma aula com foco em aprendizagem superficial exige do professor uma postura diferente do que preparar uma aula com intenção de desenvolver uma aprendizagem profunda.

A terceira contribuição consiste na utilização de uma escala de proficiência adaptada a partir da plataforma Devolutivas Pedagógicas das Avaliações de Larga Escala. Fizemos uma transposição das escalas do Saeb de matemática para os quatro níveis de desempenho apresentados por Soares (2009): abaixo do básico, básico, adequado e avançado. Essa escala é composta de exemplos reais de itens associados às suas respectivas sentenças descritoras, para fornecer um material útil e com uma interpretação pedagógica mais clara para os docentes na sala de aula.

Uma limitação do trabalho diz respeito ao tipo de item analisado, de múltipla escolha. No momento da classificação, algumas vezes, encontramos dificuldade para especificar o tipo de processo cognitivo que o item requeria dos alunos. O item de múltipla escolha oferece diferentes possibilidades de resposta e um aluno pode chegar a ela usando um caminho mais simples ou mais complexo. Além disso, em muitos momentos, classificamos os itens pensando, ao mesmo tempo, na questão e na resposta. Esse procedimento nos ajudou a identificar os processos cognitivos envolvidos. Entretanto o acesso mais direto ao raciocínio utilizado por estudantes para responder às questões envolve procedimentos complexos, como as chamadas entrevistas cognitivas, em que se pergunta diretamente para o aluno quais estratégias e passos ele usou para responder à questão. Biggs e Collis (1982) chegaram a usar procedimentos desse tipo.

Por fim, para ampliar a possibilidade de considerações mais consistentes sobre o nível de complexidade cognitiva dos itens do Saeb, pesquisas complementares são necessárias. Sugerimos também o uso da taxonomia Solo para analisar a complexidade cognitiva de outras avaliações externas e níveis de ensino.

AGRADECIMENTOS

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela concessão da bolsa de mestrado.

REFERÊNCIAS

- Alves, M. T. G., & Soares, J. F. (2013). Contexto escolar e indicadores educacionais: Condições desiguais para a efetivação de uma política de avaliação educacional. *Educação e Pesquisa*, 39(1), 177-194. <https://doi.org/10.1590/S1517-97022013000100012>

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. Addison Wesley Longman.
- Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO Taxonomy*. Academic Press. <https://doi.org/10.1016/C2013-0-10375-3>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Filipe, M. A. E. R. (2011). *A taxonomia SOLO nos exames nacionais de matemática – 9º Ano* [Dissertação de mestrado, Universidade Nova de Lisboa]. Repositório da Universidade Nova. <https://run.unl.pt/handle/10362/8173>
- Gatti, B. A. (2009). Avaliação de sistemas educacionais no Brasil. *Sísifo – Revista de Ciências da Educação*, (9), 7-18. <http://sisifo.ie.ulisboa.pt/index.php/sisifo/article/view/144>
- Hattie, J. A. C., & Brown, G. T. L. (2004). *Cognitive processes in asTTle: The SOLO taxonomy* [Technical Report, 43]. University of Auckland; Ministry of Education.
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77-89. <https://doi.org/10.1080/19312450709336664>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2015a). *Devolutivas Pedagógicas das Avaliações de Larga Escala*. Recuperado em 14 de junho de 2025, de <https://devolutivas.inep.gov.br>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2015b). *Fundamentação teórica e metodológica. Devolutivas Pedagógicas das Avaliações de Larga Escala*. Recuperado em 14 de junho de 2025, de <https://devolutivas.inep.gov.br>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2020). Histórico: Relatório Nacional do PISA 2015. *Inep*. Recuperado em 29 de outubro de 2025, de <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa/historico>
- Irvine, J. (2017). A comparison of revised Bloom and Marzano's new taxonomy of learning. *Research in Higher Education Journal*, 33. <https://eric.ed.gov/?id=EJ1161486>
- Irvine, J. (2021). Taxonomies in education: Overview, comparison, and future directions. *Journal of Education and Development*, 5(2), 1-25. <https://doi.org/10.20849/JED.V5I2.898>
- Matos, D. A. S. (2014). Confiabilidade e concordância entre juízes: Aplicações na área educacional. *Estudos em Avaliação Educacional*, 25(59), 298-324. <https://doi.org/10.18222/aeae255920142750>
- Mol, S. M. (2019). *Prova Brasil: Uma análise da complexidade cognitiva de itens de matemática por meio da taxonomia SOLO* [Dissertação de mestrado, Universidade Federal de Ouro Preto]. Repositório da UFOP. <https://www.repositorio.ufop.br/items/05387b45-da1f-4266-b74f-f8d193581ec1>
- Muri, A. F. (2017). *Letramento científico no Brasil e no Japão a partir dos resultados do PISA* [Tese de doutorado, Pontifícia Universidade Católica do Rio de Janeiro]. Divisão de Bibliotecas e Documentação. https://www.dbd.puc-rio.br/pergamum/tesesabertas/1311525_2017_completo.pdf
- Nunes, S. M. L. (2013). *A proficiência matemática dos alunos brasileiros no PISA 2003: Uma análise dos itens de incerteza* [Tese de doutorado, Universidade Federal de Minas Gerais]. Repositório Institucional UFMG. <https://repositorio.ufmg.br/handle/1843/BUOS-9AZGWC>
- Oliveira, R. P., & Araujo, G. C. (2005). Qualidade do ensino: Uma nova dimensão da luta pelo direito à educação. *Revista Brasileira de Educação*, (28), 5-23. <https://doi.org/10.1590/S1413-24782005000100002>

- Organisation for Economic Co-operation and Development (OECD). (2017). *PISA 2015 assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving*. OECD. https://www.oecd.org/en/publications/pisa-2015-assessment-and-analytical-framework_9789264281820-en.html
- Santos Fontanive, N., Gomes Elliot, L., & Klein, R. (2007). Os desafios da apresentação dos resultados da avaliação de sistemas escolares a diferentes públicos. *REICE – Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 5(2e), 262-273. <https://www.redalyc.org/articulo.oa?id=55150219>
- Soares, J. F. (2009). Índice de Desenvolvimento da Educação de São Paulo – Idesp: Bases metodológicas. *São Paulo em Perspectiva*, 23(1), 29-41. http://produtos.seade.gov.br/produtos/spp/v23n01/v23n01_03.pdf
- Webb, N. L. (2002). *Depth-of-knowledge levels for four content areas* [Unpublished paper].
- Wells, C. (2015). The structure of observed learning outcomes (SOLO) taxonomy model: How effective is it? *Journal of Initial Teacher Inquiry*, 1, 37-40. <http://dx.doi.org/10.26021/845>

NOTA: As contribuições de cada autor para o desenvolvimento do artigo foram as seguintes: Solange Maria Mol – conceitualização; curadoria e análise de dados; pesquisa; metodologia; redação do manuscrito original, revisão e aprovação da versão final do trabalho. Daniel Abud Seabra Matos – conceitualização; curadoria e análise de dados; pesquisa; metodologia; supervisão; redação do manuscrito original, revisão e aprovação da versão final do trabalho.