

AVALIAÇÃO DA EDUCAÇÃO BÁSICA E SEUS INSTRUMENTOS

CARLOS PALACIOS
LINA KÁTIA MESQUITA DE OLIVEIRA
Organizadores

AVALIAÇÃO DA EDUCAÇÃO BÁSICA E SEUS INSTRUMENTOS

Carlos Palacios

Lina Kátia Mesquita de Oliveira

Juiz de Fora, MG

2022

FICHA CATALOGRÁFICA

Avaliação da educação básica e seus instrumentos
[livro eletrônico] / organização Carlos Palácios, Lina Kátia Mesquita de Oliveira.
-- 1. ed. -- Juiz de Fora, MG : Ed. dos Autores, 2022. PDF.

Bibliografia.
ISBN 978-65-00-48785-5

1. Aprendizagem - Metodologia 2. Educação básica - Avaliação
3. Educação - Finalidades e objetivos 4. Prática pedagógica
5. Práticas educacionais I. Palácios, Carlos.
II. Oliveira, Lina Kátia Mesquita de.

SUMÁRIO

Prefácio	6
-----------------------	----------

Lina Kátia Mesquita de Oliveira

CAPÍTULO 1

Objetivos educacionais, matrizes de referência e a Base Nacional Comum Curricular	9
--	----------

Girlene Jesus

1. O que você vai aprender aqui?	9
2. Iniciando a construção do conhecimento	10
3. Construindo o conhecimento	12
3.1 O que são objetivos?	12
3.1.1 Como elaborar um objetivo?	12
3.1.2 Diferenciando os objetivos	13
3.1.3 O que os objetivos não são?	16
3.2 Matriz de referência	17
3.3 Taxonomia de Bloom revisada	25
4. O que você aprendeu até aqui?	30
5. Referências	32
6. Testando o que você aprendeu	34

CAPÍTULO 2

Elaboração de itens de resposta selecionada	37
--	-----------

Carlos Palacios, Glauker Amorim, Hilda Micarello, Josiane Toledo e Rosângela Veiga

1. O que você vai aprender aqui?	37
2. Iniciando a construção do conhecimento	37
3. Construindo o conhecimento	40
3.1 O que é um item de resposta selecionada?	40
3.2 Diferentes modelos de itens de resposta selecionada	41
3.3 Modelos de itens de resposta selecionada mais utilizados em avaliações em larga escala	49
3.3.1 Itens de resposta selecionada na etapa de alfabetização	50

3.3.2 Itens de resposta selecionada de língua portuguesa	58
3.3.3 Itens de resposta selecionada de matemática	64
3.4 Habilidades que podem ser avaliadas por itens de resposta selecionada	69
3.5 Etapas e critérios para elaborar itens de resposta selecionada	74
4. O que você aprendeu até aqui?	87
5. Referências	89
6. Testando o que você aprendeu	90

CAPÍTULO 3

Elaboração de itens de resposta construída	105
---	------------

Glauker Amorim, Hilda Micarello, Josiane Toledo e Rosângela Veiga

1. O que você vai aprender aqui?	105
2. Iniciando a construção do conhecimento	105
3. Construindo o conhecimento	107
3.1 O que é um item de resposta construída?	107
3.2 Diferentes modelos de itens de resposta construída	108
3.2.1 Itens de resposta construída com respostas objetivas	108
3.2.2 Itens de resposta construída com respostas subjetivas	113
3.3 Critérios para elaborar itens de resposta construída	117
3.4 Itens de resposta construída da etapa de alfabetização, de língua portuguesa e de matemática	127
3.4.1 Itens de resposta construída para a etapa de alfabetização e língua portuguesa	127
3.4.2 Itens de resposta construída de matemática	141
3.5 Considerações finais sobre o uso de itens de resposta construída	146
4. O que você aprendeu até aqui?	147
5. Referências	149
6. Testando o que você aprendeu	150

CAPÍTULO 4

Leitura e interpretação de resultados de avaliações em larga escala	159
--	------------

Luis Antônio Fajardo, Joaquim José Soares Neto, Manuel Palacios Cunha Melo e Wellington Silva

1. O que você vai aprender aqui?	159
2. Iniciando a construção do conhecimento	160
3. Construindo o conhecimento	161
3.1 A definição do propósito de um teste	161
3.2 Considerações sobre a construção dos testes educacionais	163
3.3 A Teoria Clássica dos Testes aplicada à avaliação educacional	166
3.3.1 Definição de percentis	168
3.3.2 Testes referenciados por norma e critério e dificuldade dos itens	171
3.3.3 Discriminação dos itens	174
3.4 A fidedignidade dos testes de proficiência	177
3.4.1 Distribuições bivariadas	178
3.4.2 Fatores que afetam a fidedignidade de um teste	182
3.4.3 A validade dos testes de proficiência	184
3.5 O uso da Teoria da Resposta ao Item na interpretação dos resultados das avaliações	185
3.5.1 Escalas cumulativas: o caso da escala de Guttman	185
3.5.2 Função degrau unitário: definição e relevância para a TRI	188
3.5.3 Modelo de um parâmetro ou de Rasch	190
3.5.4 Parâmetro de dificuldade dos itens no modelo de Rasch	194
3.5.5 A discriminação dos itens no modelo de Rasch	197
3.5.6 Modelo de dois parâmetros	201
3.5.7 Modelo de três parâmetros	202
3.5.8 Afinal, qual tipo de modelo empregar?	204
3.5.9 Pressupostos da TRI	205
3.6 Critérios de ancoragem de itens na construção de escalas de proficiência	206
3.7 Padrões de desempenho estudantil para nortear as interpretações pedagógicas das escalas de proficiência	211
4. O que você aprendeu até aqui?	213
5. Referências	215
6. Testando o que você aprendeu	216
Gabaritos e resoluções das atividades	223
Sobre os autores	247

Prefácio

Investir na avaliação educacional é acreditar que o futuro de nossas crianças e de nossos jovens não está relegado à condição social e à sorte de cada um. Afinal, não se avalia o desenvolvimento de um estudante senão com o objetivo principal de identificar os obstáculos que ainda existem em seu caminho, de modo a se criar condições para que consiga vencê-los. Isso implica, por sua vez, que os profissionais e as instituições envolvidas confiem na escola como um espaço capaz de promover a superação desses desafios.

A proposta da educação básica sob a ótica da Base Nacional Comum Curricular (BNCC) é de uma aprendizagem progressiva. Habilidades e conhecimentos abrem as portas para a aquisição de outras aprendizagens, que se tornam cada vez mais complexas à medida que a criança vai se transformando em um jovem e se aproxima da vida adulta. Nesse percurso, quando alguma habilidade essencial escapa ao estudante, o caminho se torna mais tortuoso, até o momento em que ele poderá encontrar uma barreira e não conseguir ultrapassá-la.

A avaliação permite que gestores e professores identifiquem obstáculos na implementação do currículo e atuem para que barreiras como essa não venham a surgir e impedir a aprendizagem de seus estudantes. Esse processo pode se dar de diferentes formas, seja por meio da avaliação somativa, realizada ao final de um ciclo escolar, seja por meio da avaliação formativa, aplicada em diferentes momentos ao longo do ano letivo. A depender dos objetivos educacionais, dos recursos disponíveis e das estratégias mobilizadas, os instrumentos que compõem essas avaliações também podem variar, o que influenciará, por sua vez, na produção dos resultados e na sua apropriação.

Para que uma avaliação contribua de forma efetiva para a gestão do currículo e a promoção do direito que todo estudante tem de aprender, é fundamental conhecer seus conceitos e suas etapas inerentes. Isso diz respeito tanto a uma avaliação em larga escala, que envolve profissionais de diferentes instâncias, quanto a uma avaliação interna, aplicada pelo professor ou pela professora no nível de uma turma. Em ambos os casos, a avaliação precisa ser concebida com cuidado e técnicas que possibilitem a construção de um diagnóstico que venha a embasar tomadas de decisão voltadas à melhoria da aprendizagem.

Este livro busca cobrir os principais conceitos e etapas que compreendem o processo avaliativo. No Capítulo 1, tratamos do momento inicial, que envolve a construção dos objetivos que devem pautar a escolha das habilidades, com base no currículo, que compõem as matrizes de referência dos sistemas de avaliação. No Capítulo 2, iniciamos a abordagem sobre a construção dos instrumentos, por meio dos quais serão avaliadas as habilidades selecionadas. Aqui, nos concentramos nos itens de resposta selecionada, bastante frequentes em avaliações externas, dada a sua objetividade para correção e cálculo dos resultados.

Por sua vez, no Capítulo 3, nosso foco se direciona aos itens de resposta construída, comuns em atividades internas em sala de aula, mas que têm ganhado cada vez mais espaço em avaliações em larga escala. Um dos grandes diferenciais desse instrumento é a possibilidade de avaliar habilidades mais complexas do currículo, que exigem, por exemplo, raciocínio crítico do estudante. Finalmente, o Capítulo 4 trata de medidas produzidas a partir das avaliações em larga escala, com destaque à Teoria Clássica dos Testes (TCT) e à Teoria da Resposta ao Item (TRI), levando em conta a forma como podem ser calculadas e interpretadas, bem como as vantagens de cada uma e suas variações.

Ao final de cada capítulo, um conjunto de atividades revisa os principais pontos trabalhados, de modo a reforçar alguns tópicos fundamentais. Na última seção do livro, por sua vez, estão disponíveis os gabaritos e as resoluções dessas atividades.

Como podem perceber, este livro busca contribuir à formação de diferentes profissionais, desde gestores e técnicos de redes e sistemas de ensino, responsáveis por elaborar e conduzir políticas públicas de forma abrangente, até gestores escolares e professores, que se encontram na linha de frente da educação. Além disso, os capítulos deste livro também servem a pesquisadores da área, cujas reflexões podem vir a contribuir ao aperfeiçoamento das práticas de avaliação.

Entendemos que não basta que um ou outro ator educacional tenha domínio do processo avaliativo, sendo necessário que toda a rede atue em sintonia e com os mesmos objetivos. Se em nosso país a avaliação já começa a ser encarada como imprescindível para a gestão do currículo, resta, portanto, que faça parte da agenda de cada um de nós, envolvidos na missão de promover uma educação de qualidade a todos.

Lina Kátia Mesquita de Oliveira

OBJETIVOS EDUCACIONAIS, MATRIZES DE REFERÊNCIA E A BASE NACIONAL COMUM CURRICULAR

Girlene Jesus

1. O que você vai aprender aqui?

- Definição de objetivos globais, educacionais e instrucionais e suas diferenças
- Conceito de competências, habilidades e matrizes de referência
- Taxonomia de Bloom revisada
- Relação entre objetivos educacionais, Base Nacional Comum Curricular (BNCC) e currículo

2. Iniciando a construção do conhecimento

Vamos começar a nossa discussão a partir da leitura do pequeno texto abaixo.

CASO DE UM PROFESSOR

Quando ouvi pela primeira vez sobre a possibilidade de uma base curricular nacional, fiquei intrigado. Pensei que seria bom ter uma ideia clara do que os alunos deveriam saber e poder fazer isso em cada disciplina, em cada série. Mas quando vi os rascunhos das habilidades esperadas... Havia muitas. Havia 85 em língua portuguesa da sexta série (minha área de especialidade); e havia mais de 100 em matemática da sexta série. E as habilidades eram tão vagas. Lembro-me de uma em particular: "Descrever as conexões entre influências históricas e culturais e as seleções de alfabetização". Que conexões? Que influências? Quais seleções? E o que elas querem dizer com descrever? Perguntei-me: "Como essas coisas podem me ajudar a ensinar melhor e a meus alunos a aprender melhor?".

Fonte: ANDERSON et al., 2001. [com adaptações].

O caso desse professor nos faz refletir sobre a importância de elaborar objetivos de forma clara e simples. Uma preocupação que deve nortear desde objetivos globais e educacionais, presentes nos currículos e nas matrizes das avaliações educacionais em larga escala, até os objetivos instrucionais, utilizados por professores em sala de aula.

Mas, antes de definirmos e diferenciarmos cada um desses objetivos, talvez seja importante respondermos a uma pergunta fundamental: como elaborar objetivos com um bom grau de clareza e simplicidade? Saiba que há algumas ferramentas que são indispensáveis para a realização desse trabalho. E o propósito deste Capítulo 1 é justamente conduzir você por uma trilha de aprendizagem composta pelos principais elementos que poderão auxiliá-lo nessa tarefa, que não é nada trivial.

De início, é importante saber que uma ferramenta fundamental para a construção de objetivos globais, educacionais e instrucionais são as taxonomias. Taxonomia é simplesmente uma palavra utilizada para designar uma classificação, cuja função principal é disponibilizar um modelo para que os educadores possam elaborar objetivos, como é o caso da Taxonomia de Bloom.

A Taxonomia de Bloom comprehende três domínios principais de aprendizagem que todos os professores devem conhecer, além de utilizá-los em seu dia a dia. Esses domínios são: (1) cognitivo (pensamento), (2) afetivo (emoção/sentimento) e (3) psicomotor (físico/cinestésico).

Cabe, antes de mais nada, fazer alguns comentários sobre a origem desses três domínios. Comumente, encontramos referências que atribuem todos esses domínios a Benjamin Bloom, o que não é verdade. Bloom participou da descrição dos domínios cognitivo e afetivo, tendo aparecido como primeiro autor na descrição do cognitivo. Como resultado, esse domínio ainda hoje leva o seu nome e, constantemente, se ouve falar, entre os educadores, sobre a Taxonomia de Bloom, mesmo que outros colegas o tenham ajudado nessa tarefa.

Por sua vez, a descrição do domínio afetivo foi publicada em 1964, com David Krathwohl, que acabou sendo nomeado como primeiro autor, embora Bloom também tenha trabalhado no seu desenvolvimento. Já o domínio psicomotor não foi completamente descrito até os anos 1970, período em que surgem três trabalhos fundamentais sobre esse domínio: de Anita Harrow (1972), de E. J. Simpson (1972) e de R. H. Dave (1970).

Em diversos momentos deste capítulo, falaremos da Taxonomia de Bloom. Ao mesmo tempo, consideramos importante destacar que, embora receba esse nome, outros autores importantes participaram da pesquisa que deu origem a ela. Feita essa ressalva, também cabe apontar que, mesmo sendo o domínio cognitivo o nosso foco, devemos considerar que a maioria dos comportamentos se dá a partir de uma combinação do cognitivo, do afetivo e do psicomotor. Além disso, mesmo que seja impossível categorizar comportamentos como sendo pertencentes exclusivamente a apenas um desses três, ao escrever objetivos para um determinado estudante ou grupo de estudantes, o educador deve concentrar-se no objetivo ou na sua intenção principal e organizar a avaliação para refletir uma diferença real no comportamento do aluno (HARROW, 1972). Há casos em que a declaração de um objetivo é cognitiva, mas o estudante precisa realizar um ato psicomotor para demonstrar a sua cognição. O educador, no entanto, avaliará os aspectos cognitivos do comportamento do aluno. Um exemplo disso é o desenvolvimento da caligrafia (HARROW, 1972).

Por fim, outro ponto relevante a ser destacado é o grau de correspondência entre objetivos, processo de ensino-aprendizagem e avaliação. Esses três elementos precisam estar alinhados. Afinal, se o processo de ensino-aprendizagem não estiver alinhado com as avaliações, provavelmente não influenciará no desempenho do estudante na avaliação, mesmo sendo um processo de alta qualidade. Da mesma forma, se a avaliação não estiver alinhada com os objetivos, os seus resultados não refletirão o alcance desses objetivos (ANDERSON *et al.*, 2001).

A partir desse contexto, fica clara a relevância do conteúdo deste capítulo não apenas para a elaboração de objetivos, mas para se pensar em um processo de ensino que esteja alinhado com esses objetivos e em uma avaliação que reflete uma verificação de alcance do que foi proposto.

3. Construindo o conhecimento

3.1 O que são objetivos?

No dia a dia, os objetivos nos ajudam a concentrar nossa atenção e nossos esforços. Eles indicam o que queremos realizar. Na educação, os objetivos indicam o que queremos que os estudantes aprendam, pois são “formulações explícitas das maneiras pelas quais esperamos que os alunos sejam alterados pelo processo educativo” (BLOOM, 1956, p. 26). Tendo em vista que a prática de ensinar é um ato intencional e racional, sempre ensinamos com algum objetivo, principalmente para facilitar o aprendizado do estudante. Por isso, na educação, os objetivos são especialmente importantes.

Nesse sentido, os ambientes, as atividades e as experiências de aprendizagem devem estar alinhados ou consistentes com os objetivos esperados (ANDERSON et al., 2001).

3.1.1 Como elaborar um objetivo?

Marzano e Kendall (2007) propõem uma estrutura específica para elaborar os objetivos educacionais. Primeiro, inicia-se com a frase “*O estudante (ou estudantes) será capaz de...*”, que deverá, depois, ser acrescida de um verbo e um objeto. O verbo deverá indicar o processo cognitivo – isto é, memorizar, compreender, aplicar, analisar etc. – que é empregado no objetivo, enquanto o objeto precisará indicar o tipo de conhecimento que é o foco desse processamento mental. Exemplo: “*Espera-se que, ao final desta unidade, o estudante seja capaz de **distinguir** [verbo que indica o processo cognitivo] entre **fato e opinião** [objeto que indica o conhecimento sobre o qual o processo cognitivo vai operar]*”.

3.1.2 Diferenciando os objetivos

Existem diferentes tipos de objetivos e é de suma importância saber distingui-los. O domínio geral dos objetivos pode ser representado como um *continuum* que varia de muito geral a bem específico. Nesse *continuum*, podemos distinguir três níveis: objetivos globais, objetivos educacionais e objetivos instrucionais (ANDERSON et al., 2001).

O Quadro 1, a seguir, apresenta a relação entre os três níveis de objetivos, fazendo uma comparação entre o escopo e o nível de abrangência de cada tipo.

Quadro 1: Relação entre objetivos globais, educacionais e instrucionais

ESCOPO	NÍVEL DO OBJETIVO		
	GLOBAL	EDUCACIONAL	INSTRUCIONAL
	Amplo	Moderado	Estreito
Tempo necessário para aprendizagem	Um ou mais anos	Semanas ou meses	Horas ou dias
Propósito ou função	Fornecer uma visão global, por meio de competências gerais	Desenvolver currículo e matrizes de referência	Preparar planos de aulas
Exemplo de uso	Planejar um currículo nacional para muitos anos	Planejar unidades de ensino; elaborar habilidades/desritores para matrizes de avaliações	Planejar atividades, experiências e exercícios diários

Fonte: ANDERSON et al., 2001 [com adaptações].

Agora, vamos compreender, de forma mais detalhada, cada um desses tipos de objetivos separadamente.

Objetivos globais

De acordo com Anderson et al. (2001), os objetivos globais são o resultado de um aprendizado complexo e multifacetado, os quais requerem tempo e ensino substanciais para serem alcançados. Eles são amplamente declarados e abrangem um grande número de objetivos que são mais específicos. A BNCC traz diversos exemplos de objetivos que podemos classificar como globais por meio das competências gerais. A seguir, apresentamos alguns exemplos de competências gerais da BNCC que podemos classificar como objetivos globais:

- Utilizar diferentes linguagens – verbal (oral ou visual-motora, como Libras, e escrita), corporal, visual, sonora e digital –, bem como conhecimentos das linguagens artística, matemática e científica, para se expressar e partilhar informações, experiências, ideias e sentimentos em diferentes contextos e produzir sentidos que levem ao entendimento mútuo.

[...]

- Argumentar com base em fatos, dados e informações confiáveis, para formular, negociar e defender ideias, pontos de vista e decisões comuns que respeitem e promovam os direitos humanos, a consciência socioambiental e o consumo responsável em âmbito local, regional e global, com posicionamento ético em relação ao cuidado de si mesmo, dos outros e do planeta. (BRASIL, 2017, p. 9)

Como se pode verificar, esses objetivos gerais da BNCC levam tempo substancial para serem desenvolvidos e dependem do trabalho de anos de escolarização. Para alcançá-los, é necessário que se trabalhe, ao longo do tempo, objetivos um pouco mais específicos, como os educacionais.

Objetivos educacionais

Para que professores usem objetivos globais em seu planejamento do processo de ensino-aprendizagem, os objetivos devem ser divididos de forma mais focada e delimitada. Por exemplo, se o objetivo global estiver relacionado a “estimular a imaginação” do estudante, tal generalidade o torna difícil de ser utilizado para planejar atividades em sala de aula, definir procedimentos de avaliação adequados e avaliar o desempenho dos alunos de maneira significativa. Dessa forma, objetivos mais específicos são necessários para essas tarefas (ANDERSON et al., 2001).

Vale a pena destacar que um dos principais propósitos da taxonomia original de Bloom era concentrar a atenção em objetivos um pouco mais específicos que os objetivos globais, que são chamados de objetivos educacionais. Os objetivos educacionais ocupam a faixa intermediária no *continuum* dos objetivos. Como tal, eles são mais específicos do que os objetivos globais, porém ainda mais gerais do que os objetivos necessários para orientar as instruções diárias da sala de aula fornecidas pelos professores (ANDERSON et al., 2001).

Para ilustrar esse nível de objetivo, utilizaremos como exemplo a matriz de Língua Portuguesa do Saeb de 2019 referente ao 2º ano do ensino fundamental. No Quadro 2, a seguir, destacamos alguns exemplos interessantes, que podemos classificar como objetivos educacionais:

Quadro 2: Trecho da matriz de Língua Portuguesa do 2º ano do ensino fundamental

EIXO DO CONHECIMENTO	HABILIDADE
Apropriação do Sistema de Escrita Alfabética	Relacionar elementos sonoros das palavras com sua representação escrita
	Ler palavras
	Escrever palavras
Leitura	Ler frases
	Localizar informações explícitas em textos
	Inferir o assunto de um texto
	Reconhecer a finalidade de um texto
	Inferir informações em textos verbais
	Inferir informações em textos que articulam linguagem verbal e não verbal
Produção textual	Escrever texto

Fonte: BRASIL, 2019.

Observe, no Quadro 2, que existem duas dimensões: Eixo do Conhecimento e Habilidade. Na Taxonomia de Bloom revisada – da qual trataremos mais à frente neste capítulo –, você também verá a distinção de duas dimensões nos objetivos. Ao prosseguirmos na análise desse exemplo, verificamos que as habilidades elencadas – as quais também podemos chamar de objetivos educacionais – podem levar semanas ou meses para serem desenvolvidas. Além disso, elas são específicas o suficiente para que os professores possam planejar os objetivos instrucionais.

A fim de tornar mais clara a compreensão, veja alguns exemplos de objetivos educacionais, retirados da BNCC, referentes ao 1º ano do ensino fundamental e ao componente curricular de Ciências:

(EF01CI03) Discutir as razões pelas quais os hábitos de higiene do corpo (lavar as mãos antes de comer, escovar os dentes, limpar os olhos, o nariz e as orelhas etc.) são necessários para a manutenção da saúde.

(EF01CI04) Comparar características físicas entre os colegas, reconhecendo a diversidade e a importância da valorização, do acolhimento e do respeito às diferenças. (BRASIL, 2017, p. 329)

Objetivos instrucionais

Os objetivos instrucionais têm uma especificidade substancialmente maior do que os objetivos educacionais. O propósito dos objetivos instrucionais é enfocar o ensino e a avaliação em partes específicas do aprendizado diário, em áreas de conteúdo bastante particulares. Eles são úteis para o planejamento das aulas diárias.

Veja, a seguir, alguns exemplos de objetivos instrucionais (ANDERSON *et al.*, 2001):

- ✓ O aluno aprende a somar dois números de um dígito.
- ✓ O aluno é capaz de citar três causas da 2^a Guerra Mundial.
- ✓ O aluno é capaz de classificar objetivos como globais, educacionais e instrucionais.

Como é possível verificar nos exemplos acima, os objetivos instrucionais podem ser alcançados em horas ou poucos dias, sendo muito úteis para o planejamento diário do trabalho docente.

3.1.3 O que os objetivos não são?

Até aqui, definimos os objetivos e diferenciamos seus níveis. Agora, faz-se importante destacar o que os objetivos não são, a fim de evitarmos alguns lugares-comuns e armadilhas que podem atrapalhar o trabalho.

Alguns educadores tendem a confundir meios e fins. Os objetivos descrevem resultados e mudanças pretendidos. Por isso, atividades instrucionais, como ler um livro, ouvir o professor, conduzir um experimento e fazer uma excursão, são todos meios pelos quais os objetivos são alcançados. Ou seja, as atividades instrucionais, se escolhidas com sabedoria e usadas adequadamente, levam ao alcance dos objetivos declarados. Sendo assim, uma sentença como “os alunos serão capazes de aplicar as regras adequadas para escrever parágrafos coerentes” é uma declaração de um objetivo. Por outro lado, o ato de escrever parágrafos é uma atividade que pode ou não levar ao alcance do objetivo (ANDERSON *et al.*, 2001).

Quando os objetivos não são declarados explicitamente, eles geralmente estão implícitos na atividade instrucional. Por exemplo, uma atividade pode ser definida como “ler o livro *Dom Casmurro*”. Para determinar o objetivo associado a essa atividade, podemos perguntar ao professor ou à professora: “O que você deseja que seus estudantes aprendam lendo *Dom Casmurro*?” A resposta a essa pergunta é o objetivo, que poderia ser “Quero que meus alunos diferenciem a linguagem de Machado de Assis de outros escritores do mesmo período”. Se, no lugar dessa resposta, forem dadas várias outras, é provável que haja vários objetivos (ANDERSON *et al.*, 2001).

Outro ponto importante a ser destacado é que as atividades instrucionais não são objetivos, tampouco testes ou outras formas de avaliação. Por exemplo, a declaração “Os estudantes devem ser capazes de obter uma boa nota no Exame Nacional do Ensino Médio” não é um objetivo educacional. Para determinar o objetivo educacional, devemos buscar o conhecimento e os processos cognitivos que os alunos devem aprender ou possuir para obter uma boa nota no teste (ANDERSON *et al.*, 2001).

Na próxima seção, veremos o papel dos objetivos educacionais como base para a construção de competências e habilidades utilizadas nas matrizes de referência das avaliações educacionais em larga escala.

3.2 Matriz de referência

A matriz de referência é o documento utilizado nas avaliações em larga escala para guiar o processo de elaboração dos itens que compõem os testes. Ela é construída por especialistas da área e consiste em um extrato representativo do conjunto de conhecimentos que fazem parte do currículo de determinada área. A matriz de referência não avalia cada tópico do currículo, mas deve refletir o que se espera como essencial no processo de ensino-aprendizagem para determinado ano ou etapa que esteja sendo avaliado.

Ao elaborar a matriz de referência, os especialistas consideram, essencialmente, dois aspectos: (a) o arcabouço de conhecimentos da área a ser avaliada; e (b) o tipo de teste que será elaborado: objetivo e/ou discursivo, ou prático.

Os passos a serem seguidos no processo de elaboração da matriz de referência são: 1) definição do modelo teórico que será adotado, sendo o mais comum um modelo que tem como foco competências, as quais, para serem alcançadas, dependem de um conjunto de habilidades e objetos de conhecimento; 2) estudo minucioso do currículo do ano ou etapa a ser avaliado, por disciplina; 3) eleição de competências, com

suas respectivas habilidades e objetos de conhecimento, bem como as relações possíveis entre esses componentes; 4) estudos de validação da matriz de referência proposta por meio de revisão de conteúdo por especialistas da área, bem como análise semântica com o objetivo de verificar a compreensão da matriz por uma amostra do seu público-alvo e estudo piloto por meio de oficinas de elaboração de itens com base na matriz.

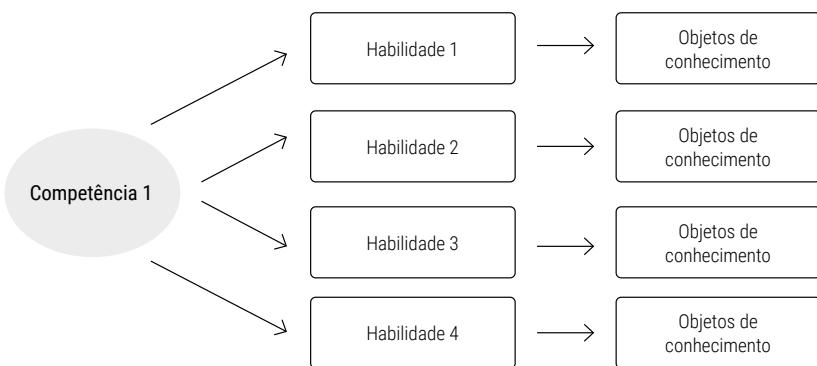
Vamos, agora, detalhar cada um desses quatro passos para a elaboração de uma matriz de referência.

Passo 1 – Definição do modelo teórico que será adotado

No Brasil, em geral, as matrizes de referência são construídas com base no modelo de competências. A matriz do Exame Nacional do Ensino Médio (Enem), por exemplo, é composta por eixos cognitivos comuns a todas as áreas do conhecimento. Para que esses eixos sejam alcançados, são eleitas diversas competências, por área. Por sua vez, para que cada competência de uma área seja alcançada, os especialistas definem um conjunto de habilidades. Consequentemente, para que as habilidades sejam trabalhadas, é necessário fazer uso de objetos de conhecimento, que comumente são conhecidos também como conteúdos programáticos.

Em suma, poderíamos descrever esse modelo da seguinte forma:

Figura 1: Fluxograma de competências, habilidades e objetos de conhecimento



Mas o que seria exatamente uma competência? Para compreendermos esse conceito fundamental para a realização deste primeiro passo, vamos partir da leitura de uma narrativa bem simples.

FÁBULA

Para atrair mais atenção para o seu reino, conquistar fama e trazer mais turistas, o rei da Cabelomania decidiu raspar a barba. Ao mandar chamar o barbeiro do rei, tomou conhecimento de que não havia esse profissional naquele reino. O real comitê informou: – Majestade, há mais de cem anos ninguém faz a barba! Não temos um só barbeiro! Surpreso com a notícia, o rei ordenou que o comitê partisse em busca do melhor barbeiro de todo o país. E foi o que aconteceu. Quando o mais famoso barbeiro foi encontrado, o comitê o submeteu ao exame real. Ele deveria mostrar que estava apto a cortar a barba do rei. Para tanto, o primeiro membro do comitê solicitou: – Fale-nos sobre a história de barbear. E o barbeiro fez.

- Fale-nos sobre a importância de barbear, pediu o segundo membro. E o barbeiro fez.
- Fale-nos sobre os instrumentos que usará para fazer a barba do rei, solicitou o terceiro membro. E o barbeiro fez.

Com a medalha de aprovação no pescoço, o barbeiro foi levado até o rei. Após ter preparado seus instrumentos e desdobrado a roupa do rei, o barbeiro pegou a navalha e, num movimento brusco, accidentalmente tirou um pedaço da orelha do rei.

- Caramba! – gritou o rei. O que vocês fizeram? – indagou ele ao comitê. Eu exijo habilidade e é isso que vocês me trazem?

E para que aquele ato não se repetisse, o rei ordenou que cortassem as cabeças dos membros do comitê.

Moral da estória: para cada tipo de desempenho existe uma forma adequada de buscar sua demonstração. Perguntas erradas podem levar a conclusões equivocadas.

Fonte: SENAI-SP e CESPEUnB, 2009.

A fábula apresentada relaciona os conceitos de competência e avaliação. E essa é uma grande preocupação no mundo educacional nos dias atuais. Como trabalhar competências e avaliá-las? A avaliação educacional em larga escala tem promovido, nos ambientes escolares, a discussão sobre o tema da avaliação de competências. Mas o que é competência e como as matrizes de referência tentam operacionalizá-las para uso nas avaliações educacionais em larga escala?

O mundo atual é complexo e imprevisível, trazendo, muitas vezes, um distanciamento entre a escola e as exigências do mercado de trabalho. Para lidar com essa realidade, surge a demanda por um novo paradigma educativo (SÁ; PAIXÃO, 2013). E esse novo paradigma emerge da discussão sobre competências, de modo que cada situação no dia a dia se torna uma oportunidade para colocar em ação as competências. De acordo com Fleury e Fleury (2001), a noção de competência surge relacionada a verbos como: saber agir, mobilizar recursos, integrar saberes múltiplos e complexos, saber aprender, saber engajar-se e assumir responsabilidades. Le Boterf (2003) e Zarifian (2003) definem competência como a capacidade do sujeito de agir diante de situações-problema inusitadas, a partir da mobilização dos recursos (cognitivos, sociais e afetivos) necessários à sua resolução.

Na BNCC, competência é definida como “a mobilização de conhecimentos (conceitos e procedimentos), habilidades (práticas cognitivas e socioemocionais), atitudes e valores para resolver demandas complexas da vida cotidiana, do pleno exercício da cidadania e do mundo do trabalho” (BRASIL, 2017, p. 8). A competência não reside, portanto, nos recursos a serem mobilizados, mas na sua mobilização e articulação. Conforme nos diz Araujo (2003), ser competente significa, diante de uma situação-problema, *mobilizar recursos subjetivos*, comportamentos e conhecimentos disponíveis e articulá-los aos pontos críticos identificados, para que seja possível tomar decisões e fazer encaminhamentos adequados e úteis ao enfrentamento da situação. Em síntese, a competência pertence à ordem do “saber mobilizar”.

Essa visão traz grandes desafios para o contexto escolar, pois a sala de aula se torna “pequena” para o desenvolvimento de competências. Muitas vezes, é necessário sair. Não basta saber “o quê”; é preciso ser capaz de enfrentar diferentes situações-problema. O raciocínio precisa ser instigado, estimulado, cobrado, pois o conhecimento necessita de um contexto, de modo que não basta apenas aprender uma fórmula – é necessário saber para que ela serve, em que pode ser utilizada e como pode ser útil na vida do estudante.

Se o estudante está nos anos iniciais do ensino fundamental, a própria cantina da escola ou uma loja de balas se tornam situações-problema a serem enfrentadas. Ao chegar a esses ambientes para efetuar compras, a criança precisará mobilizar e articular habilidades e objetos de conhecimento. A capacidade de mobilização e articulação levará o estudante a tomar decisões e fazer encaminhamentos adequados à resolução da situação-problema.

Vamos, agora, analisar alguns exemplos concretos de habilidades extraídas da matriz de Matemática do Saeb referente ao 5º ano do ensino fundamental.

Quadro 3: Exemplos de habilidades da matriz de Matemática do Saeb do 5º ano do ensino fundamental

5M1.1	Reconhecer a unidade de medida ou instrumento mais apropriado para medições de comprimento, área, massa, tempo, capacidade ou temperatura.
5M1.2	Estimar/Inferir medida de comprimento, capacidade ou massa de objetos, utilizando unidades de medida convencionais ou não OU Medir comprimento, capacidade ou massa de objetos.
5M1.3	Medir OU Comparar perímetro ou área de figuras planas desenhadas em malha quadriculada.
5M1.4	Reconhecer volume como grandeza associada a sólidos geométricos OU Medir volumes por meio de empilhamento de cubos.
5M1.5	Identificar horas em relógios analógicos OU Associar horas em relógios analógicos e digitais.
5M1.6	Relacionar valores de moedas e/ou cédulas do sistema monetário brasileiro, com base nas imagens desses objetos.

Fonte: BRASIL, 2019.

Essas habilidades pertencem ao objeto de conhecimento “Grandezas e medidas”, que contempla conhecimentos sobre comprimento (incluindo perímetro), massa, tempo, temperatura, área (de triângulos e retângulos), capacidade e volume (de sólidos formados por blocos retangulares), unidades e instrumentos de medida, incluindo problemas envolvendo essas grandezas. Além disso, contempla conhecimentos sobre moedas e cédulas do sistema monetário brasileiro e problemas que envolvam situações de compra e venda.

As habilidades elencadas para esse objeto de conhecimento integram um rol de outras mais, que foram delineadas para a avaliação de Matemática do Saeb referente ao 5º ano do ensino fundamental. A base para a elaboração dessas e das demais habilidades foram as competências gerais e específicas de Matemática da BNCC. Isso nos leva a uma conclusão tão simples quanto fundamental: avaliando as habilidades é que se avaliam, afinal, as competências previstas em uma determinada avaliação.

Em sala de aula, ao pensar em trabalhar com base em competências, o professor ou a professora deve sempre se questionar: como posso contextualizar esse conteúdo para os meus alunos? Como isso pode fazer sentido para a vida deles? A sala de aula é suficiente para isso? Esse questionamento se mantém no momento da avaliação, em que o professor ou a professora continuará buscando contextos, situações-problema, estudos de caso, experimentos etc. Eis o motivo pelo qual as grandes avaliações nacionais têm feito uso de itens contextualizados com a finalidade de avaliar competências previstas nas matrizes de referência.

Passo 2 – Estudo minucioso do currículo e seleção de competências e habilidades

Como já é possível perceber, a matriz de referência não pode ser confundida com o currículo. Tampouco o currículo pode ser substituído pela matriz de referência. Isso porque o currículo faz parte da política educacional dos sistemas de ensino brasileiros, e é a partir dele que se estabelece a formação mínima comum exigida para os diferentes anos e componentes da educação básica nas unidades da federação. Atualmente, temos uma base para essa discussão do currículo pelos estados e Distrito Federal, que já foi mencionada aqui: a Base Nacional Comum Curricular (BNCC). A BNCC deve embasar os currículos da educação básica das diferentes redes de ensino em todas as unidades da federação. De acordo com o Ministério da Educação (MEC), a BNCC determina quais competências, habilidades e conhecimentos espera-se que os estudantes desenvolvam ao longo da educação básica.

É com base nesse estudo minucioso do currículo que são selecionados competências, habilidades e objetos de conhecimento para compor as matrizes de referência das diferentes áreas. Esse trabalho é feito por especialistas de cada área e deve ser fruto de uma construção coletiva. Quando se trata das matrizes das avaliações nacionais, como o Saeb, é utilizada a BNCC como base.

Passo 3 – Eleição de competências, com suas respectivas habilidades e objetos de conhecimento

Como já vimos, para que as competências sejam desenvolvidas, é necessário trabalhar habilidades. Estas, por sua vez, são compostas por objetos de conhecimento, também chamados de conteúdos. Um exemplo dessa relação entre competência, habilidade e objetos de conhecimento pode ser visto no trecho da matriz do Enem que apresentamos a seguir.

Com base no que já discutimos sobre objetivos educacionais, preste atenção nos verbos que iniciam cada competência e habilidade, lembrando que *verbo*, aqui, significa *processo cognitivo*.

Quadro 4: Fragmento da matriz de Matemática do Enem

Enem – Exemplo da matriz de Matemática e suas tecnologias

Competência de área 1 – Construir significados para os números naturais, inteiros, racionais e reais

Habilidades

- H1. Reconhecer, no contexto social, diferentes significados e representações dos números e operações – naturais, inteiros, racionais ou reais.
- H2. Identificar padrões numéricos ou princípios de contagem.
- H3. Resolver situação-problema envolvendo conhecimentos numéricos.
- H4. Avaliar a razoabilidade de um resultado numérico na construção de argumento sobre afirmações quantitativas.
- H5. Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos.

Objetos de conhecimento

- ✓ Conhecimentos numéricos: operações em conjuntos numéricos (naturais, inteiros, racionais e reais), desigualdades, divisibilidade, fatoração, razões e proporções, porcentagem e juros, relações de dependência entre grandezas, sequências e progressões, princípios de contagem.

Fonte: BRASIL [s.d.]

Como você pôde perceber, nesse exemplo, para o desenvolvimento da Competência de área 1, é necessário trabalhar as Habilidades 1, 2, 3, 4 e 5. Por sua vez, uma única habilidade pode ser composta por diversos objetos de conhecimento.

Outro ponto a ser destacado refere-se ao verbo que inicia a competência e cada uma de suas habilidades, sem deixar de ressaltar que os objetos de conhecimento, por outro lado, não possuem um verbo no início. Isso porque o *verbo* reflete o *processo cognitivo* esperado. Tomando como exemplo as Habilidades 4 e 5, ambas iniciam com o verbo “*avaliar*”, ou seja, o processo cognitivo requerido por essas habilidades é a capacidade do estudante de avaliar. Já o objeto de conhecimento reúne os conteúdos que precisam ser mobilizados para que esses processos cognitivos possam ser realizados.

Vamos, agora, analisar outro exemplo, a matriz de referência de Linguagens para o 9º ano do ensino fundamental do Saeb. Como pode ser observado no quadro a seguir, as matrizes do Saeb estão estruturadas em duas dimensões: na primeira, estão elencados os eixos do conhecimento; na segunda, os eixos cognitivos.

Quadro 5: Exemplos de habilidades da matriz de Linguagens do Saeb do 9º ano do ensino fundamental

EIXOS DO CONHECIMENTO	EIXOS COGNITIVOS		
	RECONHECER	ANALISAR	AVALIAR
Educação Física	1. Identificar as diferentes valências físicas necessárias à realização de práticas corporais (jogos eletrônicos, lutas, práticas corporais de aventura, ginásticas, esportes e dança). 2. Identificar o valor do patrimônio urbano e natural nas vivências das práticas corporais de aventura urbana e na natureza. 3. Identificar as características (códigos, rituais, elementos técnico-táticos, indumentária, materiais, instalações, instituições) das lutas.	1. Analisar as práticas corporais frente à disponibilidade de locais para sua vivência. 2. Diferenciar os esportes com base nos critérios de sua lógica interna. 3. Diferenciar as danças urbanas, seus elementos constitutivos e seu valor cultural nas demais manifestações da dança. 4. Analisar as transformações históricas, o processo de esportivização e a midiatização das práticas corporais, com ênfase nas lutas.	1. Avaliar a multiplicidade de padrões de estética corporal disseminados pela mídia, que geram uma prática excessiva de exercícios e o uso de recursos ergogênicos. 2. Avaliar os problemas presentes nos esportes e abordados pela mídia, tais como doping, violência ou corrupção. 3. Avaliar a relação entre as práticas corporais e a promoção da saúde.

Fonte: BRASIL, 2019.

Observe que no cruzamento dos eixos cognitivos – reconhecer, analisar e avaliar – com o eixo do conhecimento, estão elencadas algumas habilidades. Lembrando, novamente, que o verbo de cada habilidade reflete o *processo cognitivo*, e que, neste caso, pertencem ao eixo do conhecimento da educação física.

Agora, você talvez esteja se perguntando: como escolher o verbo, isto é, o processo cognitivo, para a competência ou habilidade? Neste momento, nos deparamos com a relevância das taxonomias, que nos ajudam a classificar e, consequentemente, escolher os processos cognitivos mais adequados para os propósitos que desejamos. Entre as atuais taxonomias que vêm sendo utilizadas na educação, a mais comum é a Taxonomia de Bloom revisada. Já falamos um pouco aqui sobre ela e a aprofundaremos mais adiante ainda neste capítulo.

Passo 4 – Estudos de validação da matriz de referência

Os estudos de validação da matriz de referência se iniciam com a análise minuciosa do conteúdo por especialistas da área. Esse estudo deve englobar uma pequena amostra do público que utilizará a matriz, tendo como objetivo verificar a compreensão de todo o conteúdo por parte de quem utilizará a matriz de referência para elaborar os itens que irão compor as avaliações. Por fim, recomenda-se também um estudo piloto da matriz por meio de uma oficina de elaboração de itens com base na versão proposta para ser a final.

3.3 Taxonomia de Bloom revisada

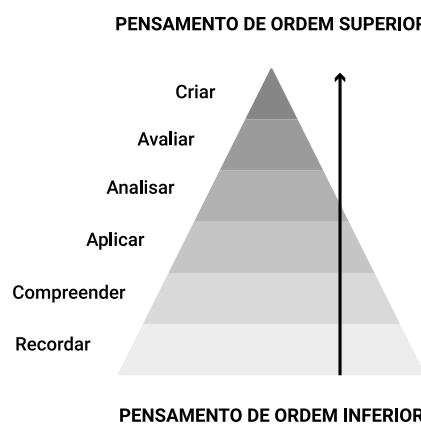
A Taxonomia de Bloom não é apenas um esquema de classificação, mas uma possibilidade de organização hierárquica dos processos cognitivos de acordo com níveis de complexidade e objetivos do desenvolvimento cognitivo desejado e planejado.

Há duas Taxonomias de Bloom, a original e a revisada. A taxonomia original é *hierárquica*, ou seja, o processo cognitivo é considerado hierárquico, por meio do qual um nível somente é alcançado após o anterior ter sido conquistado. Por exemplo, um indivíduo seria capaz de *analisar* somente após *aplicar*. A taxonomia original é também *unidimensional*, isto é, trata apenas da dimensão do processo cognitivo.

Ao analisar a relação direta entre *verbo* e *substantivo*, os pesquisadores chegaram à conclusão de que verbos e substantivos deveriam pertencer a dimensões separadas, nas quais os *substantivos* formariam a base para a dimensão *conhecimento* (*o quê*) e o *verbo* para a dimensão relacionada aos *aspectos cognitivos* (*como*) da taxonomia revisada. Essa separação entre substantivos e verbos, conhecimento e aspectos cognitivos deu um caráter *bidimensional* à taxonomia revisada (ANDERSON et al., 2001).

Na revisão da taxonomia, houve a alteração da terminologia *domínio cognitivo* para *domínio do processo cognitivo*, sendo este, conforme já abordamos aqui, o meio pelo qual o conhecimento é adquirido ou construído e como é usado para resolver problemas. Ao separar, conceitualmente, o conhecimento (conteúdo) do processo cognitivo, a taxonomia revisada estruturou-se da seguinte forma na categoria do domínio do processo cognitivo (ANDERSON et al., 2001):

Figura 2: Progressão do domínio do processo cognitivo da Taxonomia de Bloom revisada



Fonte: MISS GLAU, [s.d.]

- ✓ **1. Lembrar (ou Recordar):** Relacionado a reconhecer e reproduzir ideias e conteúdos. Reconhecer requer distinguir e selecionar uma determinada informação, enquanto reproduzir ou recordar está mais relacionado à busca por uma informação relevante memorizada. Representado pelos seguintes verbos no gerúndio: Reconhecendo e Reproduzindo.
- ✓ **2. Entender:** Relacionado a estabelecer uma conexão entre o novo e o conhecimento previamente adquirido. A informação é entendida quando o aprendiz consegue reproduzi-la com suas “próprias palavras”. Representado pelos seguintes verbos no gerúndio: Interpretando, Exemplificando, Classificando, Resumindo, Inferindo, Comparando e Explicando.
- ✓ **3. Aplicar:** Relacionado a executar ou usar um procedimento numa situação específica, podendo também incluir a aplicação de um conhecimento numa situação nova. Representado pelos seguintes verbos no gerúndio: Executando e Implementando.

- ✓ **4. Analisar:** Relacionado a dividir a informação em partes relevantes e irrelevantes, importantes e menos importantes, bem como entender a inter-relação existente entre as partes. Representado pelos seguintes verbos no gerúndio: Diferenciando, Organizando, Atribuindo e Concluindo.
- ✓ **5. Avaliar:** Relacionado a realizar julgamentos baseados em critérios e padrões qualitativos e quantitativos ou de ciência e eficácia. Representado pelos seguintes verbos no gerúndio: Checando e Criticando.
- ✓ **6. Criar:** Relacionado a colocar elementos com o objetivo de criar uma nova visão, uma nova solução, estrutura ou modelo utilizando conhecimentos e habilidades previamente adquiridos. Envolve o desenvolvimento de ideias novas e originais, produtos e métodos por meio da percepção da interdisciplinaridade e da interdependência de conceitos. Representado pelos seguintes verbos no gerúndio: Generalizando, Planejando e Produzindo.

Na taxonomia revisada, o princípio da progressão da complexidade foi mantido: do simples e concreto (Lembrar) para o complexo e abstrato (Criar). Mas, ao mesmo tempo, foi atribuída mais flexibilidade ao conceito cumulativo e dependente de cada categoria. Embora a nova taxonomia mantenha o *design* hierárquico da original, ela é flexível, pois possibilita considerar a possibilidade de interpolação das categorias do processo cognitivo quando necessário, devido ao fato de que determinados conteúdos podem ser mais fáceis de serem assimilados a partir do estímulo pertencente a uma categoria mais complexa. Por exemplo, pode ser mais fácil entender um assunto após aplicá-lo e só então ser capaz de explicá-lo. A interpolação das categorias não é total, de modo que, especificamente, na categoria conhecimento, a ordem deve ser respeitada (ANDERSON *et al.*, 2001).

Em relação ao domínio do conhecimento, a segunda dimensão da Taxonomia de Bloom revisada, podemos verificar, no Quadro 6, os tipos e subtipos mais relevantes desse domínio.

Quadro 6: Maiores tipos e subtipos da dimensão do conhecimento

Tipos e subtipos	Exemplos
Tipo 1. Conhecimento factual – trata dos elementos básicos que os estudantes devem conhecer para familiarizarem-se com uma disciplina ou resolverem problemas dela	
Subtipos: <ul style="list-style-type: none"> ✓ Conhecimento de terminologia ✓ Conhecimento de detalhes e elementos específicos 	<ul style="list-style-type: none"> > Vocabulário técnico, símbolos musicais > Principais recursos naturais, fontes de informação confiáveis
Tipo 2. Conhecimento conceitual – refere-se às inter-relações entre os elementos básicos dentro de uma estrutura maior que as habilitam a funcionar juntos	
Subtipos: <ul style="list-style-type: none"> ✓ Conhecimento de classificações e categorias ✓ Conhecimento de princípios e generalizações ✓ Conhecimento de teorias, modelos e estruturas 	<ul style="list-style-type: none"> > Períodos geológicos > Teorema de Pitágoras, lei da oferta e procura > Teoria da evolução, estrutura do congresso
Tipo 3. Conhecimento procedural – refere-se a como fazer alguma coisa, aos métodos de investigação e aos critérios para mobilizar habilidades	
Subtipos: <ul style="list-style-type: none"> ✓ Conhecimento de habilidades e algoritmos específicos da disciplina ✓ Conhecimento de técnicas e métodos da disciplina ✓ Conhecimento de critérios para determinar quando usar procedimentos apropriados 	<ul style="list-style-type: none"> > Habilidade usada para pintar com as cores da aquarela, algoritmo de divisão de número inteiro > Técnicas de entrevista, método científico > Critério usado para determinar quando aplicar um procedimento envolvendo a segunda lei de Newton; > Critério usado para julgar a viabilidade de usar um método particular de estimar custos de negócios

Fonte: ANDERSON et al., 2001 [com adaptações].

Como podemos observar nesse quadro, temos três tipos de conhecimento: de fatos, de conceitos e de procedimentos. Cada um deles tem a sua importância dentro das respectivas disciplinas e cumpre função importante no processo de ensino-aprendizagem.

Finalmente, a Taxonomia de Bloom revisada pode ser vista de forma sintetizada, por meio de uma tabela, que tem como objetivo servir como base para o planejamento de objetivos para o ensino e a avaliação. O próximo quadro apresenta a tabela da Taxonomia de Bloom revisada.

Quadro 7: Tabela da Taxonomia de Bloom revisada

Dimensão do conhecimento	Dimensão do processo cognitivo					
	1. Lembrar	2. Entender	3. Aplicar	4. Analisar	5. Avaliar	6. Criar
A. Conhecimento factual						
B. Conhecimento conceitual						
C. Conhecimento procedural						

Fonte: ANDERSON *et al.*, 2001 [com adaptações].

Geralmente, quando se elaboram objetivos/habilidades, usa-se a tabela da taxonomia para planejar os cruzamentos entre as duas dimensões: do processo cognitivo e do conhecimento. O uso da tabela possibilita uma visão mais ampliada do conjunto de objetivos ou habilidades, ajudando a evitar super ou subestimação dos domínios contemplados. A seguir, um exemplo de uso da tabela da Taxonomia de Bloom revisada com habilidades da matriz de referência de Matemática do Saeb para o 2º ano do ensino fundamental.

Quadro 8. Exemplo de uso da tabela da Taxonomia de Bloom revisada com habilidades da matriz de referência de Matemática do 2º ano do EF

Dimensão do conhecimento	Dimensão do processo cognitivo					
	1. Lembrar	2. Entender	3. Aplicar	4. Analisar	5. Avaliar	6. Criar
A. Conhecimento factual	Reconhecer/nomear figuras geométricas planas (círculo, quadrado, retângulo e triângulo)					
B. Conhecimento conceitual			Resolver problemas de adição ou de subtração, envolvendo números naturais de até 3 ordens, com os significados de juntar, acrescentar, separar ou retirar	Comparar dados estatísticos expressos em gráficos (barras simples, colunas simples ou pictóricos).	Analizar argumentações sobre a resolução de problemas de adição, subtração, multiplicação ou divisão envolvendo números naturais.	
C. Conhecimento procedimental		Calcular o resultado de adições ou subtrações, envolvendo números naturais de até 3 ordens				Representar os dados de uma pesquisa estatística ou de um levantamento em listas, tabelas (simples ou de dupla entrada) ou gráficos (barras simples, colunas simples ou pictóricos)

4. O que você aprendeu até aqui?

Neste capítulo, você aprendeu a distinguir objetivos globais, educacionais e instrucionais, bem como comprehendeu os seus usos. Por meio dos tópicos abordados, você percebeu que os objetivos se localizam em um *continuum* que parte do mais geral (objetivos globais), para o mais específico (objetivos instrucionais), estando os objetivos educacionais localizados na parte intermediária desse percurso.

Você também entendeu que, para elaborar objetivos educacionais, deve-se utilizar uma estrutura específica, que se inicia com um verbo no infinitivo – que denota o processo cognitivo – e um objeto ou complemento – que indica o tipo de conhecimento que será o foco do processamento mental.

Neste capítulo, também abordamos o que os objetivos não são, sendo possível assimilar que estes são fins, e não meios. Os objetivos referem-se àquilo que esperamos em termos de mudanças, de desenvolvimento nos estudantes, de modo que atividades instrucionais, como ler um livro específico, não se configuram como objetivos.

Além disso, você aprendeu que a matriz de referência é um documento que embasa a construção dos itens que compõem as avaliações educacionais em larga escala e que ela é elaborada por especialistas, tendo como base o currículo. Você também viu que a matriz de referência é um documento científico, que exige um percurso de diferentes passos para ser elaborada e validada.

Um dos passos importantes no processo de elaboração da matriz de referência é a eleição ou elaboração de competências e habilidades. Nesse contexto, você aprendeu que a competência, para ser desenvolvida, necessita de um conjunto de habilidades. A competência é uma capacidade desenvolvida no indivíduo que lhe permite mobilizar recursos para agir frente a situações-problema com as quais se depara no seu dia a dia.

Por fim, promovemos, neste capítulo, uma ampla abordagem da Taxonomia de Bloom, que possui uma versão revisada que tem sido a mais utilizada atualmente. A Taxonomia de Bloom revisada continua hierárquica, mas não mais unidimensional, pois estrutura-se a partir de dois domínios: do processo cognitivo e do conhecimento. Conceitualmente, o conhecimento foi separado do processo cognitivo, e esta talvez seja a mudança mais marcante da taxonomia revisada em relação à taxonomia original de Bloom.

Sendo assim, você aprendeu que, na classificação do processo cognitivo, proposta pela taxonomia, a trajetória se dá a partir do pensamento de ordem inferior para níveis mais elevados de pensamento de ordem superior. E que, na taxonomia revisada, o processo cognitivo é classificado de forma hierárquica, em seis categorias: Lembrar, Entender, Aplicar, Analisar, Avaliar e Criar. Por sua vez, a dimensão do conhecimento é categorizada em três tipos principais: Factual, Conceitual e Procedimental.

5. Referências

ANDERSON, L. W.; KRATHWOHL, D. R.; AIRASIAN, P. W.; CRUIKSHANK, K. A.; MAYER, R. E.; PINTRICH, P. R.; RATHS, J.; WITTROCK, M. C. (ed.). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. 2. ed. New York: Longman, 2001.

ARAUJO, C. M. M. *Psicologia escolar e o desenvolvimento de competências: uma opção para a capacitação continuada*. Tese (Doutorado) - Instituto de Psicologia, Universidade de Brasília, Brasília, 2003.

BLOOM, B. S.; ENGELHART, M. D.; FURST, E. J.; HILL, W. H.; KRATHWOHL, D. R. *Taxonomy of Educational Objectives – the classification of educational goals. Handbook 1 – Cognitive Domain*. Michigan, US: Longmans, 1956.

BRASIL. Ministério da Educação. *Base Nacional Comum Curricular*. Brasília, 2017.

BRASIL. Ministério da Educação. *Sistema de Avaliação da Educação Básica: Documentos de referência – versão preliminar*. Brasília, 2019.

BRASIL. Ministério da Educação. Matriz de Referência ENEM. Brasília, s.d.

FLEURY, M. T. L.; FLEURY, A. Construindo o conceito de competência. *RAC*, Edição Especial, p. 183-196, 2001.

HARROW, A. *A Taxonomy of Psychomotor Domain: A Guide for Developing Behavioral Objectives*. New York: David McKay, 1972.

KRATHWOHL, D. R.; BLOOM, B. S.; MASIA, B. B. *Taxonomy of Educational Objectives – The Classification of Educational Goals – Handbook II: Affective Domain*. New York: David McKay Co., Inc., 1964.

LE BOTERF, G. *Desenvolvendo a competência dos profissionais*. Porto Alegre: Artmed Editora, 2003.

MARINHO-ARAUJO, C. M.; RABELO, M. L. Avaliação educacional: a abordagem por competências. *Avaliação*, v. 20, n. 2, p. 443-466, 2015.

MARZANO, R. J.; KENDALL, J. S. *Designing & Assessing Educational Objectives: Applying the New Taxonomy*. Thousand Oaks, California: Corwin Press, 2008.

MARZANO, R. J.; KENDALL, J. S. *The new taxonomy of educational objectives*. 2. ed. California (USA): Corwin Press, 2007.

MISS GLAU. Taxonomia de Bloom revisada. [s.d.] 1 ilustração. Disponível em: <http://missglauedu.weebly.com/taxonomia-de-bloom-e-tecnologia.html>. Acesso em: 15 mar. 2022.

RODRIGUES JR., J. F. *A taxonomia de objetivos educacionais*. Brasília: Editora UnB, 2007.

SÁ, P.; PAIXÃO, F. Contributos para a clarificação do conceito de competência numa perspectiva integrada e sistémica. *Revista Portuguesa de Educação*, Braga, v.26, n.1, 2013.

SENAI-SP & CESPEUnB. *Provei 2009 – capacitação de elaboradores*. Brasília, 2009.

SIMPSON, E. J. *The classification of Educational Objectives: Psychomotor Domain*. University of Illinois, Research Project No. OE 5-85-104, 1966.

VIANNA, H. M. *Testes em educação*. 2. ed. São Paulo: IBRASA, 1976.

ZARIFIAN, P. (2003). *O modelo da competência: trajetória histórica, desafios atuais e propostas*. Tradução: Eric R. R. Heneault. São Paulo: Senac, 2003.

6. Testando o que você aprendeu

- 1) Pensando no que os objetivos são e no que eles não são, classifique cada um dos itens a seguir. Para o que for objetivo, insira a letra O; para o que for atividade ou meio para alcançar um objetivo, insira a letra M.

-
- Ler o livro *Romeu e Julieta*.
-
- Diferenciar fato de opinião.
-
- Inferir o sentido de uma palavra ou expressão.
-
- Participar da Feira de Ciências.
-
- Estabelecer relações entre as partes do texto, identificando repetições, substituições e os elementos coesivos que contribuem para a continuidade do texto e sua progressão temática.
-
- Fazer o dever de casa com constância e assertividade.
-
- Estabelecer relações entre as partes do texto, identificando repetições, substituições e os elementos coesivos que contribuem para a continuidade do texto e sua progressão temática.
-
- Relacionar figuras geométricas espaciais (cones, cilindros, esferas e blocos retangulares) a objetos familiares do mundo físico.
-
- Resolver as atividades dos testes propostos pela escola.
-
- Utilizar diferentes procedimentos de cálculo mental e escrito para resolver problemas significativos envolvendo adição e subtração com números naturais.
-

- 2) Considerando o que foi discutido neste capítulo, escolha uma disciplina e um ano escolar para, em seguida, elaborar um objetivo educacional com os seus respectivos objetivos instrucionais.

- 3) Considerando a Taxonomia de Bloom revisada, relate as habilidades da matriz apresentada a seguir com os respectivos processos cognitivos associados. Assinale o(s) cruzamento(s) da habilidade com o progresso cognitivo que ela pretende avaliar.**

Habilidade	Processo cognitivo					
	Lembrar	Entender	Aplicar	Analisar	Avaliar	Criar
Identificar teses/opiniões/ posicionamentos explícitos e argumentos em textos						
Analizar o uso de figuras de linguagem como estratégia argumentativa						
Avaliar diferentes graus de parcialidade em textos jornalísticos						
Identificar elementos constitutivos de gêneros de divulgação científica						
Analizar as variedades linguísticas em textos						
Producir texto em língua portuguesa, de acordo com o gênero textual e o tema demandados						
Utilizar marcas de parcialidade em textos jornalísticos						
Avaliar a adequação das variedades linguísticas em contextos de uso						

ELABORAÇÃO DE ITENS DE RESPOSTA SELECIONADA

Carlos Palacios

Glauker Amorim

Hilda Micarello

Josiane Toledo

Rosângela Veiga

1. O que você vai aprender aqui?

- Conceito de item e sua função em uma avaliação educacional
- Diferentes partes que compõem um item de resposta selecionada
- Habilidades mais recomendadas para serem avaliadas por itens de resposta selecionada de alfabetização, língua portuguesa e matemática
- Características dos itens de resposta selecionada e seus diferentes formatos
- Diretrizes para elaboração e revisão bem-sucedidas de itens de resposta selecionada de alfabetização, língua portuguesa e matemática

2. Iniciando a construção do conhecimento

Em educação, há dois grandes contextos para uso dos itens de resposta selecionada: nas avaliações em larga escala e em sala de aula. Nas avaliações em larga escala, geralmente, a finalidade é somativa, ou seja, são aplicados ao final de um ciclo ou de um bimestre ou ano específico a fim de verificar os resultados do processo de ensino-aprendizagem. Já o uso dos itens de resposta selecionada em sala de aula pode ter diferentes finalidades, desde diagnóstica, até formativa ou somativa.

Os itens de resposta selecionada, conhecidos por professores e estudantes como atividades de múltipla escolha ou questões objetivas, são o tipo de instrumento mais utilizado para compor avaliações e testes em larga escala, dada a facilidade com que os seus resultados podem ser processados. Isso porque a sua correção é bastante simples e objetiva, bastando, geralmente, apenas verificar se o estudante marcou a alternativa correspondente ao gabarito.¹

Sendo assim, podemos afirmar que as principais vantagens dos itens de resposta selecionada consistem em:

- › pouca ou nenhuma subjetividade ao corrigir;
- › possibilidade de testar uma grande variedade de conteúdos em pouco tempo;
- › facilidade para serem corrigidos usando o computador.

Por outro lado, há algumas desvantagens que devem ser destacadas. São elas:

- › impedimento da elaboração de respostas criativas pelos estudantes;
- › dificuldade para construir distratores (alternativas incorretas) de boa qualidade;
- › impossibilidade de aferir certas habilidades mais complexas, como produção textual e desenho geométrico com régua e compasso;
- › suscetibilidade ao chute.

A praticidade da correção do item de resposta selecionada também vale para o trabalho do professor ou da professora em sala de aula. Embora seja indispensável o uso de outros modelos de instrumentos de avaliação – como os que abordaremos no Capítulo 3 –, os itens de resposta selecionada permitem que o profissional elabore, aplique e corrija, em um curto espaço de tempo, testes que verificam o desenvolvimento de habilidades importantes e previstas no currículo de um determinado ano escolar e área do conhecimento.

1. Thomas P. Hogan ressalta que há outros modos de correção de itens de resposta selecionada que vão além do certo e do errado: “Uma variação envolve a concessão de créditos parciais à escolha de uma resposta que não é a melhor possível mas que, não obstante, não está completamente errada. Outra variação envolve a atribuição de pesos extras a itens de especial importância” (HOGAN, 2006, p. 166). Conforme o próprio autor pondera, não há consenso entre autores sobre a eficácia desses outros métodos, mas não há como discordar de que são significativamente mais complexos, exigindo tanto dos elaboradores de itens quanto dos corretores um trabalho maior. Neste capítulo, abordamos alguns modelos de itens de resposta selecionada que permitem, com maior facilidade, a atribuição de pontuações variadas.

Por outro lado, essa praticidade só conduzirá a uma melhora da aprendizagem se os itens forem elaborados de acordo com critérios de qualidade bem estabelecidos. Afinal, um item que contenha falhas provavelmente não nos fornecerá a informação de que tanto precisamos, isto é, se o estudante desenvolveu ou não determinada habilidade. Na pior das hipóteses, o resultado conduzirá a um diagnóstico em desacordo com a realidade dos estudantes, de modo que as ações que virão a ser desenvolvidas com base nesses resultados poderão não surtir os efeitos positivos esperados.

Por isso é fundamental o domínio da técnica de elaboração de itens para o uso de avaliações educacionais e a gestão do currículo escolar. Como sabemos, sem itens bem construídos, os testes não nos fornecerão as informações necessárias para que acompanhemos a aplicação do currículo. Por outro lado, mesmo se a avaliação não tiver sido elaborada pelo professor ou pela professora, o conhecimento sobre o instrumento utilizado permitirá que o profissional se aproprie melhor dos resultados gerados e faça um uso mais direcionado. Afinal, para sermos capazes de interpretar bem os dados de desempenho de nossos estudantes, precisamos, antes de tudo, compreender de que forma eles foram gerados.

Sendo assim, neste capítulo, faremos um percurso partindo da definição e do conceito de item para chegarmos, de forma mais específica, aos itens de resposta selecionada. A partir daí, abordaremos diferentes modelos e exemplos desse tipo de instrumento, dando atenção a cada uma das partes que o compõe, de modo a compreendermos suas potencialidades e limitações em avaliações na etapa da alfabetização e nos componentes de língua portuguesa e matemática nos ensinos fundamental e médio. Por fim, apresentaremos um conjunto de diretrizes fundamentais para a elaboração de qualquer item de resposta selecionada, muitas das quais também podem ser aproveitadas para instrumentos de outro formato.

Ao final desse percurso formativo, você terá uma compreensão mais clara dos testes compostos por itens de resposta selecionada que venham a ser aplicados aos seus estudantes. Tanto aqueles externos à escola, como no caso das avaliações em larga escala, quanto os que você mesmo poderá elaborar com base nos critérios e métodos aqui indicados.

3. Construindo o conhecimento

3.1 O que é um item de resposta selecionada?

De acordo com Thomas Haladyna (2004, p. 3), o item é a unidade básica de observação em um teste, por meio da qual o estudante recebe uma instrução que o induz a uma resposta, cuja pontuação pode seguir métodos variados e se inserir em diferentes escalas. Por sua vez, o teste é um instrumento de avaliação composto por itens e destinado a descrever, numericamente, o grau ou a quantidade de aprendizado em condições uniformes e padronizadas. Uniformidade e padronização são termos importantes que precisam ser levados em conta em uma avaliação educacional, tendo em vista que os itens de um teste precisam ser resolvidos pelos estudantes sob as mesmas condições e com o mínimo de interferência externa possível.

Esse é um ponto que já nos leva, de início, a diferenciarmos o item da questão ou da atividade que é comumente aplicada de maneira informal em sala de aula. Enquanto, no primeiro caso, precisa haver uma padronização para que o processo não sofra interferência e as informações produzidas não estejam distorcidas, no segundo, o professor ou a professora pode estabelecer critérios de acordo com o contexto, como fornecer apoio diferenciado para estudantes em níveis de desenvolvimento distintos. Isso se deve ao fato de que a atividade comum de sala de aula costuma ter como objetivo o próprio desenvolvimento ou aperfeiçoamento do estudante em relação à habilidade, à competência ou ao conhecimento em questão, enquanto o item busca somente a sua aferição.

No modelo de item de resposta selecionada, o estudante realiza um percurso cognitivo a partir de algum estímulo que o leva a um conjunto de alternativas possíveis. Caso escolha a alternativa correta, ele terá acertado a questão e demonstrado dominar a habilidade; caso contrário, será verificado que ainda não a desenvolveu. Nas avaliações em larga escala, para que a correção dos testes possa ser feita automaticamente por um computador, o item apresenta um conjunto de alternativas das quais apenas uma corresponde ao gabarito, sem abrir margem para que seja possível obter diferentes pontuações em uma mesma tarefa.

Esse é o formato de item de resposta selecionada que aprofundaremos aqui. Com vistas a aproximar você da temática central do capítulo, antes de explorarmos a estrutura de um item dessa modalidade, bem como suas potencialidades e seus limites, vamos dar uma olhada em outros modelos que, certamente, são familiares a qualquer estudante ou professor.

3.2 Diferentes modelos de itens de resposta selecionada

Com base nos modelos descritos por Haladyna (2004, p. 67-96), destacamos cinco formatos de itens de resposta selecionada bastante conhecidos, desde alguns mais limitados e pouco recomendáveis, até outros capazes de exigir tarefas mais complexas dos estudantes, que podem ser recomendáveis a depender do tipo de habilidade ou competência que se deseja avaliar. Para cada um, trouxemos exemplos de diferentes áreas do conhecimento.

Trechos suprimidos

Este é um formato de item extremamente simples e que exige pouco do estudante, tendendo mais a avaliar memorizações. No exemplo abaixo, o avaliando pode até excluir alternativas caso domine apenas parcialmente o conteúdo trabalhado. Além disso, o espaço em branco deixado no enunciado costuma confundir o estudante, que pode ter dúvidas sobre como interagir com o item e inserir a sua resposta.

- Castro Alves é um exemplo de poeta _____, enquanto Oswald Andrade é _____.
- A) modernista; romântico.
 - B) parnasiano; simbolista.
 - C) romântico; modernista.**
 - D) simbolista; parnasiano.

Caso utilize esse modelo, é importante garantir que o estudante tenha compreendido corretamente a proposta do item, pois, caso contrário, não será possível afirmar se ele não chegou ao gabarito por ainda não ter desenvolvido a habilidade trabalhada ou porque não compreendeu a presença das lacunas no item e o seu papel para resolver a tarefa proposta.

Preenchimento

Este formato ainda apresenta limitações, principalmente em relação a habilidades e competências mais complexas, porém, é mais interessante que o modelo anterior, caso o objetivo ainda seja colocar o estudante diante de duas possibilidades de res-

posta. Recomenda-se, neste caso, evitar tarefas de memorização, como nos trechos suprimidos, e focar em um conteúdo mais amplo, como no caso abaixo, que avalia habilidades básicas de escrita de forma simples. Exemplo:

Para cada par de opções numeradas, escolha a letra que indica a alternativa associada à grafia correta da palavra e depois preencha sua folha de respostas com a letra escolhida de acordo com o número do item.

(1. A. *Há* ou B. *À*) ____várias formas de investir seu dinheiro. Você pode depositá-lo em uma (2. A. *poupança* ou B. *polpança*)_____, ou você pode comprar (3. A. *assões* ou B. *ações*) _____de uma empresa. Além disso, é possível também (4. A. *investir* ou B. *envestir*)_____ em seu próprio negócio. Para cada caso, é (4. A. *preciso* ou B. *precizo*) _____ficar atento às oportunidades e aos riscos.

Gabarito: Há (1A) - poupança (2A) - ações (3B) - investir (4A) - preciso (5A)

A seguir, apresentamos outro exemplo referente à etapa de alfabetização.

Complete o texto abaixo usando somente as palavras que têm a letra T na sílaba.

PENA

Bartolomeu Campos de Queirós

É UMA PENA

A GALINHA TER TANTAS PENAS.

ELA TEM PENA DO PATO

TEM PENA DO GATO

TEM PENA DO RATO.

O _____ (BADU ou PATO) TEM BICO DE PENA.

O _____ (GADO ou GATO) TEM PENA DO RATO.

O _____ (RATO ou SAPO) TEM MEDO DE GATO.

SÓ O _____ (DADO ou RATO) NÃO TEM PENA.

Gabarito: PATO - GATO - RATO - RATO

QUEIRÓS, Bartolomeu Campos de. *O guarda-chuva do guarda*. São Paulo: Moderna, 2019. p. 10.

Aqui, as possibilidades de resposta apresentam pares de palavras que se aproximam foneticamente – T/D – e possuem o mesmo centro silábico, o que faz com que seja um item interessante – e provavelmente difícil – para estudantes que se encontram em processo de alfabetização. Conforme já ressaltado, por ser um modelo que avalia habilidades menos complexas, pode ser bastante útil para os anos iniciais do ensino fundamental. Deve-se, porém, verificar a necessidade de o professor ou a professora ler para o estudante o enunciado do item e explicar bem o que está sendo proposto para que ele realize o percurso cognitivo esperado.

Modelo complexo de múltipla escolha

Aqui, é apresentado um comando e um conjunto de opções que serão reagrupadas em alternativas, das quais uma deve ser o gabarito, enquanto as demais, os distratores. O professor ou a professora pode, também, contextualizar o item por meio de um texto ou de uma situação-problema.

Observe os nomes dos poetas abaixo.

1. Álvares de Azevedo.
2. Castro Alves.
3. Gregório de Matos.

Quais desses poetas fizeram parte do romantismo brasileiro?

A) 1 e 2.

B) 1 e 3.

C) 1, 2 e 3.

D) 2 e 3.

Esse modelo carrega uma dificuldade na elaboração, pois, muitas vezes, permite que o estudante o acerte por meio de exclusão, sem necessariamente dominar o conteúdo como um todo. No exemplo acima, saber apenas que Gregório de Matos não pertencia ao romantismo já conduz à resposta correta.

O item a seguir, correspondente à etapa da alfabetização, avalia adequadamente a habilidade de reconhecer a unidade de uma palavra, porém, por trazer muitas informações, torna mais complexa a realização da tarefa.

Leia as frases abaixo.

1. A menina brinca com o irmão.
2. Os professores passaram tarefas.
3. O menino joga bola no salão.
4. As crianças chegaram animadas.

Quais frases são formadas por 4 palavras.

- A) 1 e 2
- B) 1 e 3**
- C) 2 e 4
- D) 2 e 3

Nesse caso, o elaborador deverá refletir se o modelo é adequado aos anos iniciais do ensino fundamental e se o mais indicado não seria avaliar a mesma habilidade a partir de uma única frase, em vez de quatro.

Associações

Outro formato possível e mais complexo é o de múltiplas associações. Neste caso, o item inicia-se com um conjunto de alternativas no topo, seguido por um conjunto de opções correspondentes abaixo. As instruções que precedem as alternativas e as opções informam ao aluno sobre como responder e onde marcar as respostas. Exemplo:

A seguir, são apresentados nomes de escritores e características que definem o estilo de cada um deles. Relacione os nomes dos escritores às suas características.

A) Álvares de Azevedo.

B) Castro Alves.

C) Gonçalves Dias.

D) José de Alencar.

() Em seus romances urbanos, abordou a hipocrisia da vida na corte.

() Em uma de suas obras famosas, denunciou os horrores da escravidão.

() O fatalismo e o estilo gótico são representativos de sua obra.

() O ufanismo e o indianismo são marcantes em seus versos.

Gabarito: A, B, D, C.

No exemplo apresentado, o gabarito não se dá por meio de uma alternativa específica, mas pela ordem correta das alternativas, de modo que o professor ou a professora tem a possibilidade de estabelecer uma graduação de acerto, caso o avaliando acerte apenas parte da sequência. Isso pode fazer com que o estudante que domina parcialmente o conteúdo avaliado sinta-se mais motivado a aprender mais.

Importante ressaltar que esse é um modelo de item mais conteudista, por meio do qual você consegue trabalhar conceitos e eventos históricos, relacionando-os a definições e características. No caso de cálculos e interpretações textuais, esse tipo de item possui limitações maiores, como você pode perceber no item de língua portuguesa a seguir.

Leia o texto abaixo.

WATTERSON, Bill. *Calvin e Haroldo: e foi assim que tudo começou*. São Paulo: Conrad Editora do Brasil, 2007. p. 35.

Relacione as personagens desse texto à ação realizada por elas na história.

- A) Calvin.
- B) Mãe.
- C) Tigre.

- Demonstra impaciência com a falta de resposta.
- Desmascara o plano do outro.
- Disfarça-se com as roupas de outro.
- Irrita-se com a falta de sucesso do plano.
- Tenta faltar à escola para ficar em casa.

Gabarito: B, B, C, A, A.

Percebe-se, aqui, que o modelo de associação acaba tornando mais complexa a avaliação de uma habilidade que poderia ser aferida de forma mais simples por meio de um item tradicional de múltipla escolha. Dificilmente, um estudante capaz de fazer as duas primeiras associações de forma correta não conseguirá realizar as demais, pois já terá demonstrado uma compreensão global da tirinha. Sendo assim, nesse caso, sugere-se um item que questione o estudante a respeito do sentido global do texto, representando em uma única alternativa de múltipla escolha, evitando trabalho excessivo para a elaboração do instrumento e possíveis incompreensões por parte dos estudantes em relação ao que é exigido.

Verdadeiro ou falso

Esse modelo é bastante comum para testes em sala de aula, mas raramente utilizado em avaliações em larga escala. Possui algumas vantagens, como ser de fácil elaboração e poder avaliar diversos conteúdos de uma só vez e que podem ser respondidos em pouco tempo.

Marque V, para verdadeiro, e F, para falso.

- () Em *Canção do Exílio*, Gonçalves Dias exalta a natureza brasileira.
- () Em *Lucíola*, José de Alencar exibe uma visão idealizada da sociedade carioca.
- () Em *Navio Negreiro*, Castro Alves denuncia os horrores da escravidão.
- () Em *Noite na Taverna*, Álvares de Azevedo cria uma imagem ingênuua do homem.

Gabarito: V, F, V, F.

Como no caso anterior, não há apenas um gabarito, e sim um conjunto de alternativas corretas e incorretas. Sendo assim, o gabarito corresponderá à sequência correta de verdadeiros e falsos, de tal forma que, como no modelo anterior, o professor ou a professora poderá estabelecer uma pontuação para cada alternativa marcada corretamente, possibilitando que sejam alcançadas diferentes notas em um mesmo item.

Por outro lado, itens desse tipo tendem a avaliar habilidades e competências pouco complexas, levando a operações de memorização a partir de uma atividade mais concreta. Uma forma de atenuar essa limitação é recorrer a um texto suporte ou a uma situação-problema, de modo a tornar o item um pouco mais sofisticado. Veja, a seguir, este outro exemplo de matemática, que possui uma tabela como suporte:

Em uma academia de ginástica, foram oferecidas duas novas modalidades de aula: dança e ioga, em três turnos. Nessa academia, os alunos interessados em frequentar essas aulas fizeram uma inscrição em que deveriam optar pela modalidade e o turno de sua preferência. A mensalidade cobrada pelas aulas de dança é R\$ 50,00. Já as aulas de ioga custam R\$ 60,00 por mês. A tabela a seguir apresenta o número de alunos inscritos nas duas modalidades, por turno, no primeiro mês em que foram ofertadas essas duas novas modalidades.

Turnos	Quantitativo de alunos	
	Dança	Ioga
Manhã	10	20
Tarde	15	8
Noite	22	12

Marque V, para verdadeiro, ou F, para falso.

Em relação ao primeiro mês em que foram ofertadas essas duas modalidades,

- () a academia de ginástica teve uma receita maior com as aulas de dança do que com as aulas de ioga.
- () o turno da noite foi o que gerou maior receita com as aulas de dança e de ioga.
- () o turno da tarde teve a menor receita em relação aos demais.
- () houve mais alunos nas aulas de dança do que nas aulas de ioga.

Gabarito: F, V, V, V.

Uma variação possível desse modelo é exigir que o estudante apenas marque as alternativas verdadeiras ou corretas, deixando as falsas ou incorretas em branco. Essa proposta é interessante, pois se entende que a pressão por indicar se uma afirmação é falsa, após uma sequência de verdadeiras, pode influenciar o estudante a marcar como falsa uma alternativa que ele ainda acredita estar correta. Por isso, nessa outra proposta, o estudante poderá marcar apenas alternativas sobre as quais tiver maior certeza e evitará arriscar um “chute”.

3.3 Modelos de itens de resposta selecionada mais utilizados em avaliações em larga escala

Até aqui, você foi apresentado a diferentes modelos de itens de resposta selecionada. Embora sejam bastante utilizados em sala de aula, costumam apresentar limitações para avaliar habilidades e competências mais complexas, visto que não se apoiam em situações-problema ou suportes textuais que precisam ser levados em consideração para responder à questão, além de serem pouco práticos para serem aplicados em avaliações em larga escala.

Por isso, os sistemas de avaliação costumam focar em outros instrumentos de resposta selecionada. Antes de entrarmos em exemplos concretos, vamos compreender, em linhas gerais, a arquitetura desse modelo.

Embora cada item de resposta selecionada tenha características próprias, com níveis de dificuldade diferentes, todos irão possuir uma mesma estrutura, que se divide entre **enunciado** e **alternativas**.

O **enunciado** é o estímulo que faz com que o estudante mobilize recursos cognitivos, a fim de solucionar o problema apresentado, podendo conter um texto, uma imagem ou outros recursos que recebem o nome de **suporte**, ou apresentar uma situação-problema que funcione como um questionamento ou uma contextualização. Com ou sem suporte, o enunciado deve apresentar todos os dados e informações necessários à solução do item. Itens sem suporte ou situação-problema serão denominados *itens de múltipla escolha convencional*; quando possuem um texto como suporte, cuja leitura seja necessária para a realização da tarefa, ou uma situação-problema, serão definidos como *itens de múltipla escolha contextualizados*.²

Como você poderá observar, nos testes de matemática, alguns itens não apresentam suporte, enquanto nos de língua portuguesa, salvo a etapa de alfabetização, a presença do suporte costuma ser obrigatória. Há diferentes tipos de suporte, que podem ser criados pelo professor ou pela professora – no caso de gráficos e tabelas – e retirado de livros, periódicos e meios digitais – no caso de contos, poemas, fragmentos de romance, artigos, notícias, verbetes, tirinhas de jornal, entre diversos outros –, os quais devem ter os direitos autorais respeitados e referenciados.

2. Essas nomenclaturas foram elaboradas com base em Haladyna (2004): *Conventional Multiple-Choice* e *Context-Dependent Item Set*.

Além do suporte, outro elemento que faz parte do enunciado é o **comando**, que tem como função apresentar o problema que irá direcionar o estudante a uma resposta correspondente, podendo ser elaborado em formato de pergunta ou de sentença a ser completada pelas alternativas. De antemão, já é possível perceber que o comando deve ser sempre claro e objetivo, de modo que o estudante que domina a habilidade avaliada não tenha dúvidas daquilo que lhe está sendo solicitado.

Depois do comando, chegamos às **alternativas**, que se dividem entre **distratores** e **gabarito**. Os distratores são as alternativas incorretas, de modo que qualquer alternativa que não seja um gabarito é um distrator. O **distrator** é uma das partes mais importantes e mais difíceis na hora da elaboração do item, pois deve ser plausível – para quem ainda não desenvolveu a habilidade avaliada –, de tal maneira que o estudante não o descarte por conta de algum elemento absurdo, ao mesmo tempo que seu conteúdo precisa, sem sombra de dúvida, ser incorreto, considerando o que foi solicitado no comando. Por último, o **gabarito** é a alternativa correta do item, ou seja, é a resposta para o que está problematizado no comando. Tal qual os distratores, um gabarito não pode dar margem para que a sua correção seja questionada.

Para compreendermos melhor cada uma dessas partes, precisamos antes entender as particularidades de itens de alfabetização, língua portuguesa e matemática, bem como observar alguns exemplos concretos, com suas respectivas descrições e análises pedagógicas.

3.3.1 Itens de resposta selecionada na etapa de alfabetização

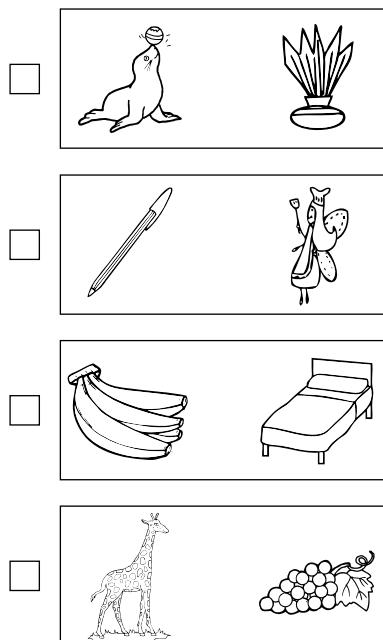
Para avaliar estudantes que se encontram na etapa da alfabetização, podem ser utilizados itens de resposta selecionada para verificar o desenvolvimento tanto de habilidades relacionadas ao processo de apropriação das regras que organizam o sistema de escrita alfabetica, quanto da leitura e compreensão de textos de gêneros variados. Entretanto, há algumas especificidades que devem ser observadas na elaboração dos itens e também na sua aplicação.

Itens de resposta selecionada que avaliam habilidades relacionadas à apropriação do sistema de escrita alfabetico

No caso dos instrumentos de resposta selecionada, que avaliam habilidades relativas à apropriação das regras que organizam o sistema de escrita alfabetica, os itens devem ser lidos, parcialmente, pela pessoa responsável pela aplicação do teste. Isso porque eles avaliam a apropriação de habilidades que antecedem à leitura propriamente dita,

com autonomia e compreensão. Destinam-se, portanto, a estudantes que podem não ter, ainda, desenvolvido habilidades que lhes permitam ler e compreender o próprio comando do item. Desse modo, a leitura pelo aplicador é uma condição fundamental para que possam ser avaliadas habilidades que dizem respeito ao processo de alfabetização ainda em curso. Vejamos três exemplos de itens de resposta selecionada desse tipo.

 **Faça um X no quadradinho onde os nomes das figuras terminam com a mesma sílaba (pedaço).**



No caso desse item, que avalia a habilidade de identificar sílabas de uma palavra, não é utilizado nenhum suporte, ou seja, nenhum texto ou imagem como estímulo à resolução do item.

O item é construído a partir de um enunciado que apresenta a situação-problema que o estudante deve resolver. Esse enunciado deve ser objetivo, claro e mobilizar a habilidade que o item se propõe a avaliar. No caso desse item, o enunciado é “Faça um X no quadradinho onde os nomes das figuras terminam com a mesma sílaba (pedaço)”. Conforme já destacamos, por se tratar de um item que avalia uma habilidade relativa à apropriação das regras que organizam o sistema de escrita alfabética, o enunciado é lido pelo aplicador.

Em seguida, são apresentadas as alternativas de resposta, das quais, como já destacamos, uma será o gabarito, enquanto as demais, os distratores. No caso desse item, a primeira alternativa apresenta as imagens do par “FOCA/PETECA”, que termina com a mesma sílaba, “CA”, e se trata, portanto, do gabarito do item. Ao escolher essas duas figuras para compor o gabarito, o elaborador do item teve o cuidado de escolher imagens pertencentes a campos semânticos diferentes – um animal e um brinquedo –, de modo que o estudante que venha a escolher essa alternativa utilize como critério a semelhança fonética da sílaba final, e não um critério semântico, por exemplo.

As demais alternativas, que são os distratores, apresentam os pares “CANETA/FADA”; “BANANA/CAMA”; e “GIRAFÁ/UVA”, respectivamente. Em cada uma, os nomes de ambas as figuras terminam pela vogal “A”, porém, as consoantes que compõem a sílaba final das palavras são diferentes. A escolha desses pares de palavras não foi, portanto, aleatória. O elaborador buscou palavras que tivessem alguma semelhança na sílaba final – nesse caso, o som vocálico –, mas cujas consoantes fossem diferentes. Sendo assim, os estudantes que venham a optar por qualquer uma dessas alternativas demonstram não ter observado essa diferença entre as sílabas finais.

Observemos mais um exemplo:

 **Veja a letra abaixo.**

Q

Faça um X no quadradinho onde a letra que você viu está escrita de outra forma.

- g
- a
- p
- q

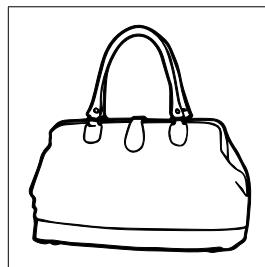
Nesse item, que avalia a habilidade de reconhecer as diferentes formas de grafar uma mesma letra ou palavra, já temos a presença do suporte, ainda bastante simples, porém indispensável para a resolução do item. Como é possível verificar, o suporte escondido foi uma letra do nosso alfabeto: a letra “Q” maiúscula.

O enunciado, por sua vez, foi construído de forma clara e objetiva, evidenciando a habilidade que se pretende avaliar: "Faça um X no quadradinho onde a letra que você viu está escrita de outra forma". Assim como no caso do item exemplificado anteriormente, por avaliar uma habilidade relativa à apropriação das regras que organizam o sistema de escrita alfabética, o enunciado é lido pelo aplicador, contendo, portanto, a presença do ícone do megafone.

Quando observamos as alternativas, verificamos que o gabarito corresponde à última alternativa, pois nela é apresentada a letra "q" minúscula. As demais alternativas de resposta, que são os distratores, foram construídas pelo elaborador do item de modo a apresentar letras que tivessem alguma semelhança, gráfica ou sonora, com a letra "Q". Desse modo, as escolhas do elaborador, mesmo no caso das alternativas que não são gabarito do item, não foram aleatórias. Elas procuraram contemplar raciocínios possíveis de serem desenvolvidos pelos estudantes, mesmo aqueles que não acertaram o item.

Vamos analisar um terceiro exemplo de item de alfabetização. No item a seguir, é avaliada a habilidade de contar as sílabas de uma palavra. Para tal, foi utilizada uma figura como suporte: a imagem de uma bolsa.

 **Veja a figura abaixo.**



 **Faça um X no número de sílabas (pedaços) do nome da figura que você viu.**

- 1
- 2
- 3
- 5

Novamente, o enunciado foi construído de forma clara e objetiva, de modo a ficar bem claro qual é a habilidade que se pretende avaliar: “Faça um X no número de sílabas (pedaços) do nome da figura que você viu”. E, assim como nos outros dois exemplos de itens que discutimos, por estar em jogo uma habilidade relativa à apropriação das regras que organizam o sistema de escrita alfabético, o enunciado é lido pelo aplicador e contém o ícone do megafone.

Verificamos, de imediato, que a segunda alternativa representa o gabarito do item, pois nela é apresentado o número 2, que corresponde à quantidade de sílabas da palavra “bolsa”. As demais alternativas de resposta, que são os distratores, apresentam raciocínios possíveis do estudante ao tentar resolver a situação-problema proposta pelo item. A primeira alternativa, que indica somente uma sílaba, seria marcada por estudantes que considerassem a palavra como um todo, e não a sua divisão em sílabas; já a terceira, que indica 3 sílabas, seria considerada correta por estudantes que tivessem separado as sílabas da palavra de forma inadequada – por exemplo, BO – U – SA. Finalmente, a última alternativa, que indica 5 sílabas, poderia ser escolhida por estudantes que tivessem contado todas as letras, e não as sílabas da palavra. Como se percebe, a escolha dos distratores do item seguiu uma lógica de raciocínios possíveis aos estudantes, não consistindo em opções de resposta aleatórias.

Itens de resposta selecionada que avaliam habilidades relacionadas à leitura com compreensão de textos de diferentes gêneros

No caso dos itens de resposta selecionada que avaliam habilidades de leitura e compreensão de textos de diferentes gêneros, o enunciado pode ou não ser lido pelo aplicador, pois se destinam a estudantes que, espera-se, já tenham desenvolvido habilidades que lhes permitam ler e compreender o comando dos itens. Nos exemplos que apresentamos a seguir, o enunciado não é lido pelo aplicador.

Leia o texto abaixo.

Cor especial, veneno mortal

A cobra coral azul da Malásia tem um estilo bem diferente. A combinação da cor azul do corpo com o vermelho da cabeça deixa o bicho impressionante. A espécie vive na Ásia, em países como Indonésia e Tailândia, e, apesar de ter apenas 140 centímetros, possui veneno que impede a respiração e pode ser letal. Ela come lagartos, rãs, pássaros e, acredite, engole até outras cobras!

Disponível em: <<http://www.recreio.com.br/licao-de-casa/as-cobras-mais-diferentes-do-mundo>>. Acesso em: 15 mar. 2013. (P010122E4_SUP)

Qual é o assunto desse texto?

- A cobra coral azul da Malásia.
- A respiração dos animais.
- Os alimentos dos animais.
- Os países da Ásia.

No caso desse item, que avalia a habilidade de reconhecer o assunto de um texto lido, o suporte utilizado é um texto de curiosidades, de extensão curta, com sintaxe simples (períodos breves, sem muitas intercalações, construídos na ordem direta) e que aborda um tema acessível aos estudantes aos quais o item será aplicado. É, portanto, um texto com nível de complexidade adequado ao público que será avaliado. Percebe-se também que o texto escolhido pertence a um gênero adequado à avaliação da habilidade em questão. Por se tratar de um texto de curiosidades, há vários tópicos que contribuem para a abordagem do assunto tratado, o que permite, ao elaborador do item, a construção do gabarito e das alternativas de resposta.

Percebemos também que o enunciado do item foi construído na forma de uma pergunta – “Qual é o assunto desse texto?” –, remetendo, de forma clara e objetiva, à habilidade que está sendo avaliada. Após a sua leitura, quando nos direcionamos à primeira alternativa, verificamos que se trata do gabarito, isto é, a alternativa correta, que apresenta o assunto do texto. Não se trata de um item difícil, pois a primeira frase do texto já apresenta o seu assunto, exatamente do modo como aparece no gabarito. As demais alternativas, que representam os distratores, não foram escolhidas ao acaso, pois apresentam tópicos secundários do texto, sendo, portanto, plausíveis.

A seguir, podemos analisar mais um item de alfabetização que contém um texto como suporte – neste caso, pertence ao gênero poema. A habilidade avaliada, por sua vez, é a de localizar informações explícitas em textos.

Leia o texto abaixo.

Faz de conta
<p>Eu vi um coelhinho nadando dois patinhos dançando três boizinhos cantando!</p> <p>Eu vi um gatinho pintando dois peixinhos voando três burrinhos falando!</p> <p>E você o que viu?</p>

DIAS, Iêda. *Reflexo*. Belo Horizonte: Ática, 1995. (P020172C2_SUP)

De acordo com esse texto, quais animais estavam dançando?

- Boizinhos.
- Burrinhos.
- Patinhos.
- Peixinhos.

O poema é pouco extenso, composto por versos curtos, numa linguagem acessível ao público a que se destina o item. Essa adequação do nível de dificuldade do texto aos estudantes aos quais o item será aplicado é necessária para que o teste seja adequado ao público que o realizará. Além disso, o texto está de acordo com a avaliação da habilidade de localizar informações explícitas, pois o poema cita ações de diversos animais, o que permite ao elaborador do item construir o gabarito com distratores plausíveis, que façam referência a elementos presentes no texto.

O comando, construído na forma de uma pergunta – “De acordo com esse texto, quais animais estavam dançando?” –, remete, de forma clara e objetiva, à habilidade que está sendo avaliada, ao solicitar que o aluno identifique, dentre os animais citados, quais estão dançando. Essa é uma informação que se encontra explícita no texto – mais especificamente, no terceiro verso.

Encontramos o gabarito na terceira alternativa, que apresenta a informação solicitada pelo item. As demais alternativas, que são os distratores, apresentam animais também citados nos versos do poema, mas que não são aqueles que estavam dançando. Sendo assim, todas as alternativas são plausíveis, mas apenas uma corresponde à informação solicitada pelo item.

Vamos, agora, analisar um último item da etapa de alfabetização, que avalia a habilidade de reconhecer o gênero de um texto. Temos, neste caso, um bilhete como suporte, de extensão curta e redigido em linguagem coloquial, com sintaxe simples e de circulação na esfera da vida cotidiana. Trata-se, portanto, de um texto acessível aos estudantes aos quais se destina a avaliação.

Leia o texto abaixo.

*Lu,
Precisei sair e não pude esperar
que acordasse. Fui levar seu avô
ao médico. Tome seu café e
arrume seu quarto. Volto logo.
Beijos
Mamãe*

Disponível em: <http://www.azinteligencia.com.br/site/index.php?option=com_content&view=article&id=214:bilhete&catid=56:4o-ano&Itemid=109>.

Acesso em: 18 mar. 2012. (P010124E4_SUP)

Esse texto é

- um bilhete.
- um convite.
- uma lista.
- uma notícia.

No enunciado, o comando é construído na forma de uma frase incompleta – “Esse texto é” –, cujo gabarito deverá completá-la de forma correta. Trata-se de um comando curto, claro e objetivo, que remete à habilidade que está sendo avaliada.

Na primeira alternativa, encontramos o gabarito: “um bilhete”. As demais, que representam os distratores, apresentam gêneros que poderiam ser confundidos com o que dá suporte ao item, por possuírem estrutura ou função social semelhante àquela do bilhete.

Como foi possível observar, a elaboração de itens de resposta selecionada e para a etapa da alfabetização requer que o elaborador disponha de conhecimentos sobre os requisitos técnicos desse tipo de item; de conhecimentos sobre a língua, de modo a construir de forma intencional o gabarito e as demais alternativas de resposta; e, ainda, de conhecimentos pedagógicos sobre as características dos estudantes aos quais os itens se destinam.

3.3.2 Itens de resposta selecionada de língua portuguesa

A partir do 4º ano do ensino fundamental, espera-se que o estudante já tenha alcançado um nível de desenvolvimento em leitura que lhe permita ler e compreender, por conta própria, o enunciado dos itens, de modo que não encontramos mais o recurso da leitura do item por parte da pessoa responsável pela aplicação. Além disso, para esses anos de escolaridade, os itens de resposta selecionada de língua portuguesa virão sempre acompanhados de um texto como suporte, que também deverá ser lido pelo estudante, sozinho, para poder resolver a tarefa proposta e chegar ao gabarito.

Portanto, itens de resposta selecionada de língua portuguesa de qualidade, que avaliam habilidades e competências importantes para o desenvolvimento do estudante, serão sempre contextualizados. Quando não possuem um texto como suporte, esses itens tendem a ser excessivamente conteudistas, focando mais em memorização e conteúdo gramatical no nível da frase, o que se distancia do propósito de uma avaliação em consonância com a BNCC e as competências do século XXI.

Sendo assim, para entendermos melhor os itens de resposta selecionada de língua portuguesa, selecionamos quatro exemplos – do 5º e do 9º do ano ensino fundamental, com suas respectivas análises estruturais e pedagógicas, tal qual fizemos para os itens da etapa de alfabetização.

Vejamos, a seguir, um item que avalia a capacidade do estudante de reconhecer diferentes formas de tratar uma informação na comparação de textos que abordam o mesmo tema, elaborado para o 5º ano do ensino fundamental.

Leia os textos abaixo.

Texto 1
Segredos das rochas
<p>Para saber mais sobre as criaturas do passado, os cientistas estudam algumas pistas, chamadas fósseis. No caso dos dinossauros, são ossos, dentes e pedaços de pele que ficaram conservados entre camadas da terra.</p> <p>Com o passar de milhões de anos, essas camadas de terra se transformam em rochas, que ficam com marcas de folhas, pedacinhos de ossos e dentes ou até esqueletos inteiros.</p> <p>Outro sinal importante são as pegadas. Com esse tipo de fóssil é possível descobrir se uma espécie de dino era veloz ou pesada, se andava em bandos e como se movia.</p> <p>Os especialistas que estudam fósseis de animais e vegetais são os paleontólogos.</p>
Recreio. Detetives do passado. São Paulo: Abril, ano 1, n. 15, 22 jun. 2000. p.15.
Texto 2
O local das descobertas Santa Maria, Rio Grande do Sul Staurikosaurus Pricei <p>É um dos quatro dinossauros mais antigos já encontrados. Viveu entre 220 e 221 milhões de anos atrás. Sua ossada foi achada em 1936 por paleontólogos da Universidade Harvard. Era carnívoro. Podia atingir até 2 metros de comprimento e 1 metro de altura.</p>
OLTRAMARI, Alexandre. O mundo perdido. Veja. São Paulo: Abril, n.12, 1998. (P050091EX_SUP)

(P050091EX) O que há em comum entre esses textos?

- A) Descrevem os tipos de rochas.
- B) Informam sobre a vida de um paleontólogo.
- C) Relatam a caçada aos dinossauros.
- D) Tratam da descoberta de fósseis de dinossauros.

Para resolverem a tarefa proposta nesse item, os estudantes precisam ter consolidado outras habilidades, como a de identificar o assunto de um texto, de forma que possam proceder a comparações entre os textos apresentados como suporte. Nesse caso, o suporte oferecido é constituído por dois textos que apresentam, de modo geral, períodos curtos, linguagem clara e objetiva, o que facilita a leitura por parte dos estudantes. O comando do item, construído em forma de pergunta – “O que há em comum entre esses textos?” –, solicita que os estudantes identifiquem o assunto comum aos dois textos. O assunto comum já se anuncia no parágrafo de abertura do primeiro texto e logo no início do segundo.

Os estudantes que marcaram a alternativa D – gabarito – demonstraram ter desenvolvido a habilidade avaliada pelo item. Aqueles que optaram pela alternativa A – distrator – podem ter focado no título do primeiro texto e associado o termo “ossada”, no segundo, à ideia de pedra, dadas as semelhanças físicas. Já aqueles que marcaram a alternativa B – também distrator – podem ter realizado uma leitura superficial e atentado apenas à referência aos paleontólogos. Por fim, os estudantes que selecionaram a alternativa C – outro distrator – podem ter direcionado a atenção somente ao termo “dinossauro”, sem identificar a abordagem que o texto realiza. Tudo isso mostra, mais uma vez, como os distratores foram cuidadosamente pensados para que se tornem plausíveis, de modo a não serem descartados por conta de algum conteúdo absurdo ou muito distante do que é abordado no texto.

Percebemos, também, como as alternativas possuem paralelismo, tendo extensão semelhante – A e C são mais longas, enquanto B e D, mais curtas. Além disso, todas começam com um verbo e possuem uma estrutura similar, de tal maneira que nenhuma chamaria excessivamente a atenção do estudante.

Agora, vejamos um item que avalia a interpretação de um texto com auxílio de material gráfico, também voltado ao 5º ano do ensino fundamental.

Leia o texto abaixo.



Recreio. São Paulo: Abril, ano 10, n. 479, p. 24, 14 maio 2009. (P050619A9_SUP)

(P050620A9) A intenção da mãe ao mandar Antoninho pular na água era

- A) afastá-lo do predador.
- B) escondê-lo do macaco.
- C) brincar com o filho.
- D) dar um banho no filho.

Nesse exemplo, temos um suporte de fácil leitura, do gênero tirinha, adequado à etapa de escolarização à qual o item é destinado. Para resolvê-lo, o estudante precisa ler o texto e compreender que, na narrativa construída pela sequência de quadros, a mãe conta com a ajuda de outro personagem para enganar o filho e fazer com que ele tome banho. Isso implica entender que não há nenhum predador na água, de modo que o que ocorre no primeiro quadro é apenas uma simulação, revelada no seguinte, quando observamos a mãe dando banho no filho. A situação toda é desvendada no terceiro e último quadro, quando a mãe se dirige ao amigo dizendo, explicitamente, que aquela era apenas uma estratégia para dar banho no filho.

Antes de escolher uma das alternativas, o estudante precisa realizar a leitura do comando – “A intenção da mãe ao mandar Antoninho pular na água era” –, que, ao contrário do exemplo anteriormente apresentado, não se estrutura em forma de pergunta, mas de sentença a ser completada. Nada disso, porém, torna o item mais complexo ou difícil, pois ainda temos um comando bastante simples e objetivo, cuja leitura não deve exigir grandes esforços de uma criança que se encontra no final da primeira etapa do ensino fundamental. Sendo assim, os estudantes que conseguem acompanhar a

progressão da narrativa quadro a quadro e articular as linguagens verbal e não verbal chegam facilmente à alternativa D – gabarito.

Quanto aos distratores, percebemos que cada um foi pensado com base em elementos presentes no texto, de modo a garantirem sua plausibilidade. Aqueles que não consolidaram ainda a habilidade de ler textos com auxílio de materiais gráficos, dentro desse nível de dificuldade, podem interpretar que, de fato, existe um predador na água, marcando a alternativa A – distrator; outros estudantes podem, por outro lado, apenas focar no último quadro e entender que o macaco estaria atrás do filho, optando pela alternativa B – também distrator. Por fim, em relação à alternativa C, os estudantes que chegaram a esse distrator podem até ter compreendido que não há nenhum predador, porém, direcionaram o raciocínio apenas ao caráter divertido e humorístico da tirinha, focando na ideia de brincadeira.

Vamos, agora, analisar um item referente ao 9º ano do ensino fundamental, o qual avalia a habilidade de estabelecer relações lógico-discursivas presentes em um texto, marcadas por conjunções, advérbios etc.

Leia o texto abaixo.

		Jovens trocam livros por “leitura digital”
5		<p>No bolso do <i>jeans</i>, um celular. Na escrivaninha do quarto, um <i>laptop</i>. [...] Tudo ao redor dos jovens de hoje oferece conexão 24 horas por dia nas mais diversas redes sociais. Como deixar de lado todas as infinitas possibilidades que o mundo digital oferece e se dedicar à leitura de um livro, com suas centenas de páginas, cheias de palavras [...] exigindo concentração para serem decifradas?</p>
10		<p>Dados do Programa Internacional de Avaliação de Estudantes (Pisa) divulgados nesta semana afirmam que a leitura não está entre as prioridades dos jovens de 15 anos. [...] 46% dos estudantes afirmam que leem apenas para obter as informações de que precisam; [...]. Apenas um terço disse que a leitura é um dos <i>hobbies</i> favoritos.</p>
15		<p>Apesar dos dados do Pisa, especialistas em educação e tecnologia discordam da ideia de que o jovem de hoje lê menos. Muito pelo contrário: afirmam que os adolescentes nunca leram tanto. A diferença é que, agora, não são só os livros que são “ídios”, mas vídeos, <i>sites</i>, <i>SMS</i>, <i>e-mails</i> e uma gama imensa de informações. “O adolescente lê e escreve muito, comunica-se muito mais por escrito. As gerações anteriores liam só os livros da escola. Os jovens de hoje não: estão sempre se informando dentro dessa vida social digitalizada”, diz Rosa Maria Farah, [...] da PUC-SP. [...]</p>
20		<p>Para os educadores, a falta de interesse pela leitura formal pode levar à perda da habilidade de se concentrar quando necessário. “O jovem não consegue mais ler um texto inteiro. [...],” explica Teresa Ferreira, psicopedagoga da Unifesp. [...]</p> <p>Ainda é cedo para afirmar o quanto isso pode ser prejudicial no futuro. Mas os especialistas alertam: ler apenas o essencial e aquilo que interessa pode levar à perda da aptidão para analisar situações com mais profundidade. “O jovem sabe de tudo o que acontece, mas não aprofunda o conhecimento dos fatos”, destaca a psicóloga Dora Sampaio Góes [...].</p>

MANDELLI, Mariana. Disponível em: <<http://www.estadao.com.br/noticias/geral/jovens-trocaram-livros-por-leitura-digital-imp-652713>>.

Acesso em: 5 mar. 2015. *Adaptado para fins didáticos. Fragmento. (P090288H6_SUP)

(P090291H6) No trecho “... mas não aprofunda o conhecimento dos fatos.” (l. 22-23), a palavra destacada expressa ideia de

- A) conclusão.
- B) consequência.
- C) finalidade.
- D) oposição.

Nesse terceiro exemplo de item de resposta selecionada de língua portuguesa, temos um texto mais longo, pertencente ao gênero reportagem, cujo conteúdo é mais sofisticado, pois agora tratamos de itens direcionados a estudantes do último ano do ensino fundamental. Ao fazermos sua leitura, encontramos dados estatísticos, oriundos de uma pesquisa, e opiniões de especialistas, que contribuem para reforçar e dar densidade ao texto e enfatizar a relevância do tema para a sociedade. Ao mesmo tempo, por se tratar de uma reportagem, cujo objetivo é, ao mesmo tempo, informar e levar o leitor a formar uma opinião, o texto faz uso de uma linguagem clara, objetiva e sem termos complicados. Além disso, aborda um tema caro ao universo educacional e conectado a questões contemporâneas.

Ao leremos o comando – “No trecho “...**mas** não aprofunda o conhecimento dos fatos” (l. 22-23), a palavra destacada expressa ideia de” –, verificamos que a tarefa exigida pelo item é a de identificar o valor semântico da conjunção coordenada adversativa “mas”, de tal maneira que o estudante precisa entender que entre “saber tudo o que acontece” e “não aprofundar o conhecimento” há uma relação de oposição, o que está indicado na alternativa D, o gabarito. As demais alternativas, os distratores, apresentam outras relações semânticas que seriam expressas por outras conjunções, com exceção daquela destacada no comando. O estudante que escolher qualquer um dos distratores revela que ainda tem dificuldades de identificar e compreender as relações estabelecidas pelos elementos coesivos, o que é fundamental para a construção de sentidos em um texto.

Outro ponto fundamental deve ser destacado em relação a esse item. Por possuir como suporte um texto com mais de vinte linhas, é fundamental que ele não se limite a trazer consigo apenas um único item. Textos como esse são ricos e podem dar origem a diversos outros itens, permitindo que o estudante se concentre em diferentes partes e elementos que o compõem, evitando-se, assim, que ele apenas venha a ler o trecho mencionado no comando. Por exemplo, poderiam ser avaliadas as habilidades de reconhecer o assunto ou a informação principal do texto; localizar informações explícitas; inferir informações implícitas; reconhecer efeito de sentido de sinais de pontuação, como as aspas em “lidos”, na linha 12, e o uso do itálico para termos estrangeiros; dentre outras.

Para finalizarmos, vamos analisar outro item referente ao 9º ano do ensino fundamental, que traz como suporte um texto mais simples e acessível, o que não faz necessariamente com que a tarefa exigida seja fácil.

Leia o texto abaixo.

Joãozinho e os pronomes

Na escola:

- Joãozinho!
- Sim, professora!
- Por favor, diga-me dois pronomes.
- Quem, eu?
- Muito bem, garoto!

Disponível em: <<http://recantodacronica.blogspot.com.br/2011/11/joaozinho-e-os-pronomes-historias.html>>. Acesso em: 3 dez. 2015. (P091384H6_SUP)

(P091384H6) O humor desse texto está

- A) na forma como o Joãozinho atende a professora.
- B) na maneira como a professora faz o pedido ao Joãozinho.
- C) no fato de Joãozinho responder corretamente sem intenção.
- D) no jeito como a professora faz um elogio ao Joãozinho.

Esse item avalia a habilidade de identificar efeitos de ironia ou humor em textos variados. Nesse caso, utilizamos como suporte uma anedota, texto bastante comum para avaliar essa habilidade. Embora seja bastante curto, passível de ser lido por estudantes de anos anteriores, para ser capaz de entendê-lo realmente é preciso que o estudante tenha desenvolvido compreensão de humor e ironia, o que não é algo simples.

Nesse exemplo, para resolver a tarefa exigida, o estudante precisa reconhecer que o personagem Joãozinho, no momento em que reage de forma desorientada e pergunta “Quem, eu?”, fornece accidentalmente a resposta correta para o que é solicitado pela professora – dois pronomes. Portanto, para avaliar a habilidade indicada, o comando – “O humor desse texto está” – apresenta uma frase bastante objetiva que deverá ser completada pelas alternativas, com foco no elemento do humor, sem dar margem a outras interpretações.

O estudante que consegue acompanhar a progressão do diálogo, articulando as informações presentes nas falas das personagens, chegará facilmente ao gabarito, que é a alternativa C. Ao mesmo tempo, os distratores apresentam alternativas plausíveis de serem selecionadas por alguém que ainda não compreendeu a estratégia utilizada no texto para gerar o efeito de humor. Os estudantes que marcam a alternativa A podem ter entendido que Joãozinho foi irônico ao atender prontamente à professora e depois não saber o que deveria responder; por outro lado, quem optou pela alternativa B, outro distrator, pode ter focado no uso da exclamação e entender que ela, por si só, já provoca o humor; finalmente, os estudantes que escolheram a alternativa D podem ter interpretado que Joãozinho não atendeu ao que foi solicitado pela professora e, por conta disso, ela estaria sendo irônica no final.

Além de fornecer um gabarito incontestável e distratores plausíveis, as alternativas possuem paralelismo em termos de estrutura e sentido. As duas primeiras começam com “na” (contração da preposição “em” com o artigo definido feminino “a”); as duas seguintes, por sua vez, trazem o mesmo recurso, porém no masculino – “no”. Ao mesmo tempo, nenhuma se destaca em termos de tamanho; ademais, todas fazem referência ao personagem Joãozinho e não fogem dos acontecimentos do texto.

3.3.3 Itens de resposta selecionada de matemática

A BNCC de Matemática do ensino fundamental se estrutura em cinco unidades temáticas: números, álgebra, grandezas e medidas, geometria e probabilidade e estatística, em torno das quais se organizam suas habilidades específicas. Dessa forma, em geral, os objetos de conhecimento ficam evidenciados no próprio descritor da habilidade, os quais, aliados à ação cognitiva demandada pela habilidade, direcionam a elaboração do item.

Dependendo da habilidade a ser avaliada, itens de resposta selecionada de matemática poderão ser contextualizados ou não. Por exemplo, há habilidades focadas em domínio de processos, conceitos ou algoritmos, em relação às quais o item deverá avaliar o domínio que o estudante tem sobre uma determinada técnica, conhecimento ou conceito, podendo, eventualmente, ser elaborado sem contexto algum. Por outro lado, há habilidades sobre a capacidade de se aplicar ferramentas matemáticas na resolução de problemas em contextos diversos, as quais, obviamente, demandam itens de resposta selecionada construídos com o intuito de mobilizar contextos variados, mas sempre compatíveis com o grau de escolaridade e com a realidade do estudante.

Sendo assim, para entendermos melhor os itens de resposta selecionada de matemática, trouxemos alguns exemplos – do 5º e do 9º ano do ensino fundamental –, com suas respectivas análises estruturais e pedagógicas.

O exemplo a seguir apresenta um item proposto a estudantes do 5º ano do ensino fundamental.

Pedro foi à papelaria e comprou uma caixa de lápis de cor por R\$ 12,50, um estojo por R\$ 8,25 e um caderno por R\$ 25,25.

Quanto Pedro gastou nessa papelaria?

- A) R\$ 20,75
- B) R\$ 33,50
- C) R\$ 37,75
- D) R\$ 46,00

Esse item avalia a habilidade de resolver problemas que envolvam representações decimais de valores convencionados de cédulas e moedas do sistema monetário brasileiro.

Para isso, o item aborda uma situação em que Pedro realizou a compra de uma caixa de lápis de cor, um estojo e um caderno, cujos valores foram, respectivamente, R\$ 12,50, R\$ 8,25 e R\$ 25,25. O comando solicita a quantia que Pedro gastou nessa papelaria.

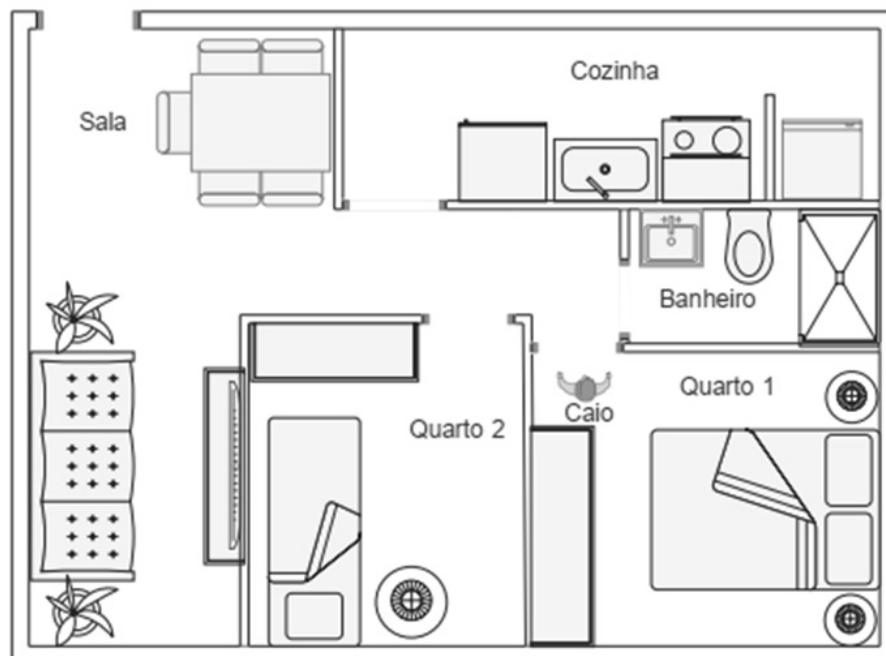
O estudante deve, inicialmente, perceber que as representações decimais de valores convencionados de cédulas e moedas do sistema monetário brasileiro estão relacionadas às representações decimais de números racionais, com décimos e centésimos. Deve-se, portanto, compreender que, para resolver esse item, é preciso somar os valores desses três objetos como adição de números racionais, utilizando estratégias pessoais ou algoritmo operatório convencional com reagrupamento, obtendo R\$ 46,00, conforme apresentado abaixo.

$$\begin{array}{r}
 \text{R\$} \quad 12,50 \\
 \text{R\$} \quad 8,25 \\
 + \quad \text{R\$} \quad 25,25 \\
 \hline
 \text{R\$} \quad 46,00
 \end{array}$$

Por fim, é importante notar que as alternativas apresentam valores não muito distantes um do outro, garantindo paralelismo sintático. E, para o estudante que não tenha consolidado essa habilidade, os distratores são plausíveis, de modo que nenhuma das alternativas se destaca excessivamente.

No exemplo a seguir, apresentamos um item aplicado a estudantes do 6º ano do ensino fundamental. Esse item faz uso de um suporte, que é a planta baixa de um apartamento.

Caio estava em seu apartamento na posição representada no desenho abaixo.



Caio saiu do Quarto 1, virou à esquerda e, em seguida, virou a primeira à direita.

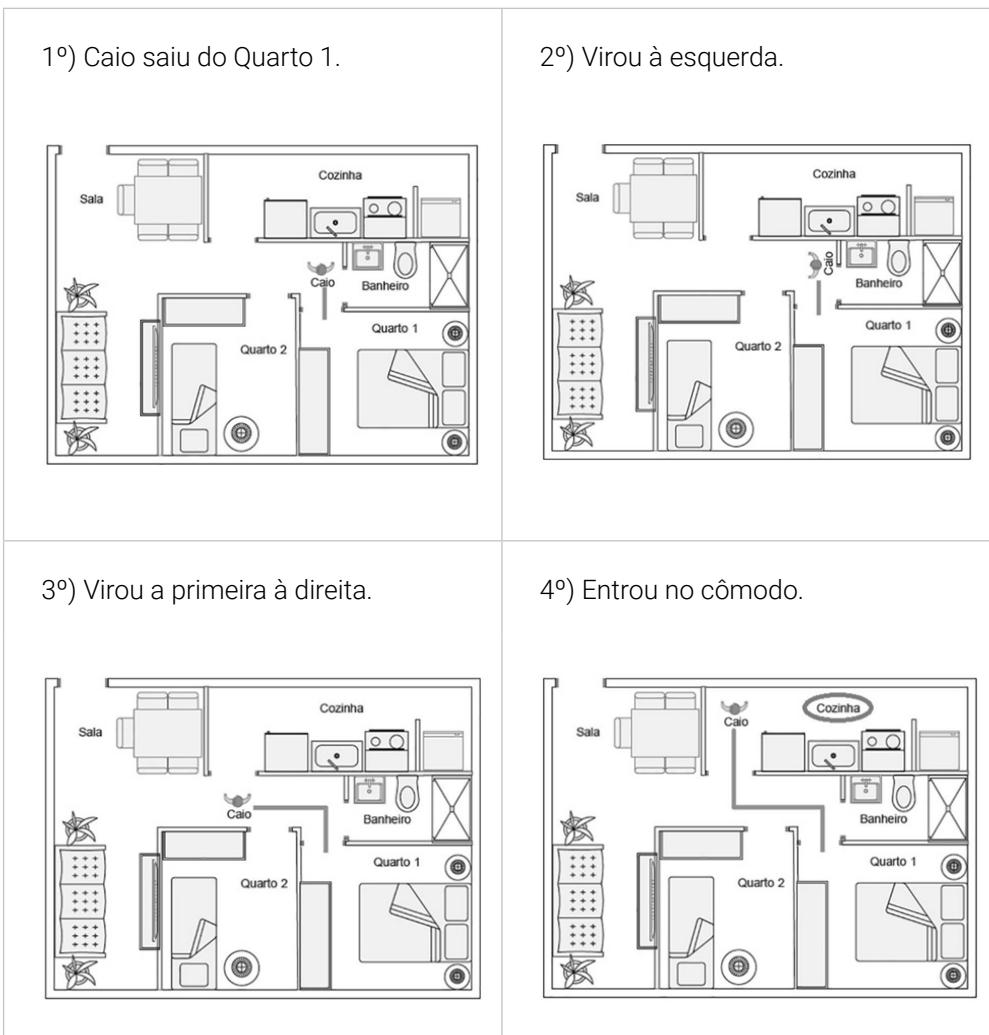
Em qual cômodo Caio entrou?

- A) Banheiro.
- B) Cozinha.
- C) Quarto 2.
- D) Sala.

O item avalia a habilidade de o estudante compreender deslocamentos de pessoas no espaço, em relação a um referencial, por meio de representações esquemáticas no plano, empregando termos “à direita” e “à esquerda” para mudanças de direção e de sentido.

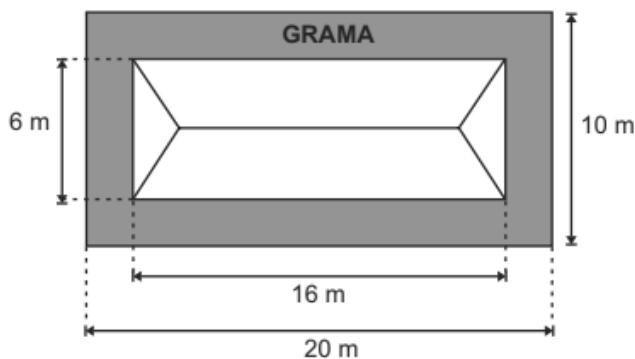
Para isso, o item apresenta a descrição de um deslocamento de Caio em seu apartamento e, como suporte, uma representação da planta baixa do local, com Caio posicionado no Quarto 1. O item solicita, em seguida, que seja identificado o cômodo em que ele entrou após ter feito esse deslocamento.

O estudante deve observar que, para determinar corretamente em qual cômodo Caio entrou, é necessário compreender a descrição desse deslocamento com o auxílio do suporte, considerando e acompanhando cada mudança de direção e de sentido pelo emprego dos termos “à direita” e “à esquerda”, e tomando como referencial o próprio Caio, representado na imagem. Na planta baixa, o estudante pode identificar a posição inicial de Caio no Quarto 1 e seguir as orientações sobre sua movimentação descritas no texto do item, conforme ilustrado na sequência de imagens abaixo.



O último exemplo, voltado ao 9º ano do ensino fundamental, ressalta a importância de produzir distratores associados a resoluções provenientes de raciocínios possíveis de serem elaborados pelo estudante, sendo, portanto, plausíveis e não excludentes entre si.

Carlos comprou um terreno retangular cujas medidas estão representadas no desenho abaixo e, no centro dele, construiu uma casa de base também retangular, medindo 6 metros de largura por 16 metros de comprimento. Ao redor da casa, ele plantou grama de forma a cobrir todo espaço que sobrou do terreno.



Quantos metros quadrados de grama Carlos plantou nesse terreno?

- A) 16
- B) 96
- C) 104
- D) 200

O item avalia a habilidade de o estudante resolver problemas que envolvam o cálculo de medida de área de figuras geométricas planas.

Para isso, o item apresenta um suporte com a imagem do terreno retangular que Carlos comprou, destacando a casa de base retangular que ele construiu, a região ao redor dessa casa em que ele plantou grama e as respectivas medidas. Em seguida, o item solicita a quantidade de metros quadrados de grama que Carlos plantou nesse terreno.

O estudante deve perceber que a medida da área da região ao redor dessa casa em que Carlos plantou grama é determinada pela diferença entre as medidas das áreas desse terreno e da base dessa casa. Como as dimensões do terreno retangular e da base retangular da casa são, respectivamente, 20m por 10m e 16m por 6m, o estudante pode concluir que as medidas respectivas das áreas dessas regiões são $20m \times 10m = 200m^2$ e $16m \times 6m = 96m$. A partir dessa informação, o aluno descobre que a área ao redor da casa mede $200m^2 - 96m^2 = 104m^2$, sendo essa a quantidade de metros quadrados de grama que Carlos plantou nesse terreno.

Observe que a alternativa A pode ser alcançada pelos estudantes que consideram as diferenças entre os comprimentos e entre as larguras do terreno e da base da casa, obtendo $20m - 16m = 4m$ e $10m - 6m = 4m$ e efetuam o produto entre essas diferenças, obtendo: $4m \times 4m = 16m^2$. Já os estudantes que assinalam a alternativa B, provavelmente, calculam somente a medida da área da base da casa, enquanto os que escolhem a alternativa D, possivelmente, efetuam o cálculo da medida da área de todo o terreno.

Com isso, verificamos como a escolha de cada distrator na elaboração do item foi pautada em raciocínios possíveis de serem adotados por estudantes que ainda não desenvolveram totalmente a habilidade avaliada, de modo que nenhum deles aparenta ser, por si, absurdo.

3.4 Habilidades que podem ser avaliadas por itens de resposta selecionada

Até aqui, conhecemos diferentes formatos de itens de resposta selecionada, cada um contendo suas potencialidades e limitações, e direcionamos nossas análises para o tipo de item que é o foco de nosso estudo: o de múltipla escolha, que pode tanto ser convencional – sem apresentar suporte ou situação-problema – ou contextualizado – apresentando suporte ou situação-problema. Ressaltamos que a escolha do formato do item está diretamente relacionada àquilo que se deseja avaliar do estudante. Não existe, portanto, formato *melhor* ou *pior* de item, mas sim item *adequado* ou *inadequado* ao seu propósito avaliativo.

O quadro a seguir apresenta, de forma sintética, o tipo de operação mais indicada para cada um dos formatos de item que apresentamos até aqui.

Formato de item de resposta selecionada	Conhecimento	Habilidade cognitiva	Competência
Trechos suprimidos	X	X	
Preenchimento	X	X	
Modelo complexo de múltipla escolha	X	X	
Verdadeiro ou falso simples	X	X	
Verdadeiro ou falso contextualizado	X	X	
Múltipla escolha convencional	X	X	
Múltipla escolha contextualizado (suporte ou situação-problema)		X	X

Fonte: HALADYNA, 2004, p. 96 [com adaptações].

Percebemos, de imediato, que a competência requer instrumentos mais sofisticados de avaliação, enquanto o conhecimento e a habilidade cognitiva podem ser avaliados por diferentes formatos de itens. Mas o que cada um significa exatamente?

A habilidade cognitiva pode ser compreendida como atos complexos que requerem conhecimento para serem desenvolvidos. Tanto o conhecimento quanto a habilidade cognitiva são facilmente ensinados e aprendidos em um curto período de tempo, enquanto a competência correlata emerge mais tarde e é mais complexa, pois é baseada em conhecimento e habilidade cognitiva. Por esse motivo, a competência costuma ser uma aprendizagem desenvolvida em um período de tempo maior (HALADYNA, 1997).

Costuma-se afirmar que o desenvolvimento da competência é o objetivo mais importante da escolarização. Sendo assim, alguns comportamentos cognitivos dos estudantes que são qualificados como competências são: 1) pensamento crítico; 2) resolução de problemas; e 3) criatividade, incluindo pensamento criativo e artístico e produção científica criativa (HALADYNA, 1997).

Para além da nomenclatura, é fundamental termos uma dimensão clara do que pretendemos avaliar. Desde a decisão sobre o formato de instrumento a ser utilizado, passando pela escolha do suporte e a redação do comando e das alternativas, todos os elementos que compõem o item devem estar de acordo com aquilo que será avaliado. Nos testes de avaliações externas, os conhecimentos, as habilidades cognitivas e

as competências contempladas, geralmente, correspondem ao que está apresentado nas matrizes de referência das avaliações nacionais, estaduais ou municipais, denominados apenas como habilidades, que se apresentam a partir dos descritores, sobre os quais você já aprendeu no Capítulo 1. Nessas matrizes, a maioria dos descritores pode ser avaliada por itens de resposta selecionada; enquanto na BNCC nem tudo será facilmente avaliado por um item desse tipo, principalmente quando estão em jogo competências mais sofisticadas, que podem estar relacionadas ao raciocínio crítico, como a capacidade de formular opiniões.

Itens de resposta selecionada ocupam, portanto, um papel específico no processo de avaliação dos estudantes, que não pode ser entendido como o único, tampouco como melhor. No caso da alfabetização, por exemplo, temos um processo que envolve diferentes dimensões da experiência dos estudantes com o conhecimento: linguística, emocional, social, psicológica, dentre outras. Dessa forma, o professor alfabetizador precisa estar atento às diferentes dimensões do processo, para que construa instrumentos voltados ao acompanhamento do desenvolvimento dos estudantes em todas as dimensões envolvidas na aprendizagem da língua escrita. Dada a especificidade de cada uma delas, são necessários instrumentos diversos para registrar os avanços dos estudantes na apropriação da língua escrita, em sua formação como leitores e escritores competentes.

Sendo assim, itens de resposta selecionada são instrumentos que permitem registrar o desempenho dos estudantes na dimensão linguística do processo de alfabetização. Como vimos, por meio deles, é possível obter informações sobre o processo de apropriação da base alfabética do sistema de escrita, assim como sobre o desenvolvimento de habilidades que contribuem para a leitura com compreensão. Na etapa da alfabetização, esse tipo de instrumento oferece um diagnóstico que pode ser apropriado tanto pelos sistemas de ensino – com o intuito de compreender como se encontra um amplo contingente de estudantes em relação à alfabetização –, quanto pelas escolas e professores – para obter o diagnóstico de uma turma ou mesmo dos estudantes individualmente. Entretanto, quando se trata do trabalho do professor ou da professora em sala de aula, itens de múltipla escolha podem não ser os instrumentos mais adequados para proceder à avaliação dos estudantes. Quando usados, eles precisam ser elaborados respeitando os requisitos técnicos e pedagógicos que já apresentamos aqui. Além disso, o diagnóstico produzido por meio da aplicação desses itens precisa ser cotejado àquele produzido por meio da aplicação de outros instrumentos de avaliação, elaborados pelo professor para acompanhar e registrar aspectos da dimensão linguística da alfabetização, assim como de outras dimensões desse processo.

É oportuno lembrar, ainda, que itens de resposta selecionada são instrumentos de avaliação e não atividades de ensino. Essa distinção é sempre importante, pois a aplicação de instrumentos desse tipo aos estudantes não pode substituir as atividades de ensino, cujo foco é a promoção da aprendizagem. Desse modo, a avaliação permite produzir um diagnóstico que informa ao professor quais as necessidades de aprendizagem dos estudantes e nisso consiste a sua importância para o processo de ensino.

Em relação aos anos iniciais do ensino fundamental, os itens de resposta selecionada, no caso do componente curricular língua portuguesa, podem ser utilizados para avaliar as habilidades relacionadas aos eixos leitura e análise linguística do currículo. Também podem ser elaborados itens cujo foco são habilidades de compreensão de textos, como localizar uma informação explícita, inferir uma informação em um texto ou reconhecer sua finalidade; habilidades que demandam uma reflexão sobre a estrutura dos textos, como a de reconhecer os elementos de uma narrativa; ou, ainda, aquelas que demandam uma reflexão sobre a língua, como a habilidade de reconhecer relações entre partes de um texto marcadas por conjunções, advérbios ou locuções adverbiais. Como essas habilidades não são todas de uma mesma natureza, é importante que o professor ou a professora tenha clareza da natureza da habilidade que cada item avalia, para que possa se beneficiar do resultado da avaliação e compreender em quais aspectos é necessário maior investimento nas práticas de ensino.

Outras habilidades igualmente fundamentais à formação de leitores e escritores competentes não podem ser avaliadas por itens de resposta selecionada, como aquelas relacionadas à argumentação sobre determinado tema, à manifestação de posicionamentos e opiniões em relação ao que leem, ao interesse e envolvimento em práticas de leitura autônoma, à produção de sentidos pessoais para a leitura de textos literários.

Nos anos finais do ensino fundamental e no ensino médio, é essencial que o professor ou a professora de língua portuguesa conheça não apenas a progressão curricular das habilidades e os objetos de conhecimentos relativos a seu componente curricular, mas também que entenda os processos de aprendizagem que caracterizam os estudantes do ano e da etapa de escolaridade para os quais as atividades compostas por itens de resposta selecionada serão destinadas. As aprendizagens vão avançando em complexidade, o que implica, consequentemente, textos mais complexos, tanto no que diz respeito à linguagem e à sintaxe utilizadas, quanto às temáticas abordadas, que se tornam mais especializadas, intensificando-se a abordagem de textos literários. Tudo isso exige que o professor ou a professora seja capaz de dimensionar quais habilidades e conhecimentos são passíveis de serem avaliados por meio de itens de resposta selecionada.

De modo semelhante ao que ocorre na alfabetização e nos anos iniciais, quando são seguidas orientações técnicas e pedagógicas para sua elaboração, itens de resposta selecionada podem ser excelentes instrumentos para realização de um diagnóstico relativo a algumas habilidades de procedimento de leitura e aspectos relativos à análise linguística e semiótica.

Isso aponta tanto para uma vantagem quanto para uma limitação desse tipo de instrumento. A vantagem diz respeito a poder fornecer, de modo rápido e preciso, um diagnóstico acerca de algumas habilidades e conhecimentos, permitindo ao professor conhecer os pontos fortes e as fragilidades nas aprendizagens de seus estudantes. Por sua vez, a limitação do emprego de itens de resposta selecionada se dá no sentido de que há uma série de habilidades importantes relacionadas ao desenvolvimento da oralidade e da produção de textos que não são passíveis de serem mensuradas por esse tipo de atividade avaliativa.

Nesse sentido, é importante que, em sua prática, o professor ou a professora mescle atividades avaliativas, de modo a poder contemplar um conjunto mais amplo de conhecimentos e habilidades e a ter uma visão igualmente ampla das aprendizagens desenvolvidas por seus estudantes.

Por fim, no caso da matemática, há habilidades que demandam ações cognitivas, como identificar, localizar, corresponder, utilizar, interpretar, compreender, para as quais itens de resposta selecionada se mostram bastante eficientes como recurso avaliativo. Entretanto, também há limitações. Por exemplo, instrumentos de resposta selecionada informam se o estudante é capaz de interpretar dados apresentados em tabelas ou gráficos ou corresponder dados apresentados em tabela a seu respectivo gráfico; por outro lado, esse formato de item não nos permite avaliar se o estudante é capaz de construir um gráfico para representar dados fornecidos em uma tabela. Da mesma forma, em geometria, as habilidades relativas às construções geométricas, com régua e compasso, também ficam fora do alcance desse tipo de item.

Itens de resposta selecionada também permitem saber se o estudante comprehende um algoritmo que executa uma determinada tarefa, mas nada informam sobre a capacidade de criar um algoritmo que execute essa tarefa. Nesse sentido, embora seja bastante eficiente, esse tipo de instrumento não é suficiente para explorar as possibilidades pedagógicas demandadas no cumprimento de currículos de matemática na educação básica.

3.5 Etapas e critérios para elaborar itens de resposta selecionada

O ponto de partida para elaborar qualquer tipo de item é escolher a habilidade que será avaliada, o que pode ser feito a partir da BNCC, da matriz de referência do sistema de avaliação ou diretamente do plano de ensino do professor ou da professora.

Após selecionada a habilidade, você deverá refletir se o item será contextualizado, o que pode se dar por meio de um suporte. Um bom critério para essa escolha é estabelecer um nível de complexidade desejado, ou verificar se o estudante precisa de alguma informação adicional para responder ao item. Tenha em mente, porém, que um suporte desnecessário pode confundir o aluno e prejudicar a qualidade do instrumento.

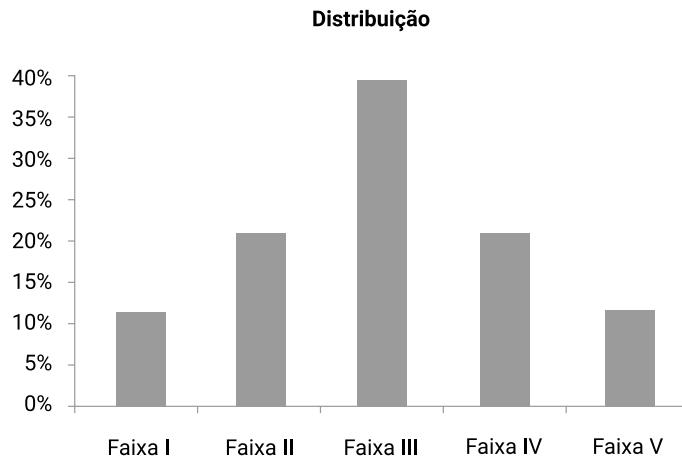
A partir daí, deve-se prosseguir para a elaboração de um comando que induzirá o estudante a realizar um percurso cognitivo em torno da habilidade avaliada; por fim, o percurso esperado deverá coincidir com uma das alternativas apresentadas, que é o gabarito.

Ao construir os distratores, procure pensar em caminhos possíveis e plausíveis que estudantes que ainda não desenvolveram a habilidade poderão percorrer, e fuja sempre das “pegadinhas”, que em nada ajudam no processo avaliativo. Quando redigir o gabarito, seja claro e evite ambiguidades.

Durante a construção de um teste com diversos itens, procure diversificar, principalmente no que diz respeito aos suportes, tanto em termos de formato quanto de nível de complexidade, de tal maneira que depois seja possível identificar quais textos precisam ser mais trabalhados em sala de aula. Isso também vale para as habilidades: tente abranger uma quantidade razoável delas, que possibilite depois, a partir dos resultados, a criação de um mapa mais completo da aprendizagem de seus estudantes.

É importante também que um teste contenha itens com diferentes níveis de dificuldade. Pasquali (2003) sugere uma distribuição do nível de dificuldade dos itens em cinco faixas, como pode ser visto no gráfico a seguir.

Figura 1: Distribuição percentual dos itens em 5 faixas de dificuldade



Fonte: PASQUALI, 2003.

Na figura apresentada, podemos ver os itens distribuídos percentualmente em cinco faixas de dificuldade, que podem ser nomeadas da seguinte forma: 10% na faixa *Muito Fácil*; 20% na faixa *Fácil*; 40% na faixa *Mediana*; 20% na faixa *Difícil*; e 10% na faixa *Muito Difícil*. Essa distribuição dos itens contribui tanto para alcançar estudantes que possuem diferentes níveis de proficiência, quanto para deixar o teste qualitativamente melhor.

Afinal, um teste muito difícil pode desestimular o estudante e prejudicar o processo avaliativo, o que também vale para um teste muito fácil, que poderá focar somente naquilo que já aprenderam. O tamanho também merece atenção, pois, um teste muito extenso pode fazer com que o estudante fique cansado e não consiga concentrar-se plenamente na resolução dos itens, ou, até mesmo, opte pelo “chute”.

Outro ponto relevante é compreender as graduações de dificuldade de uma habilidade. Assim como existem habilidades simples e habilidades complexas, cada uma possui também diferentes níveis de dificuldade. Um suporte menos acessível torna uma habilidade, automaticamente, mais difícil: por exemplo, em língua portuguesa, um poema do século XIX pode tornar o item mais difícil do que seria se tivesse sido utilizada uma crônica jornalística contemporânea, mesmo que ambos contemplam a mesma habilidade.

Isso vale para qualquer outra área do conhecimento. Por exemplo, uma habilidade de matemática que avalie leitura de gráficos se tornará mais difícil se for utilizado, como suporte, um gráfico com uma complexidade maior de informações. Em ciências hu-

manas, um item que contenha um texto de um teórico do século XVIII tem chances de ser mais difícil que outro que avalie a mesma habilidade, mas possua um suporte simples e familiar. O mesmo vale para um suporte de ciências da natureza que consista na imagem de uma cadeia alimentar com seres vivos que não fazem parte do contexto do aluno.

Tudo isso nos leva a uma regra fundamental que deve sempre nortear a elaboração de qualquer instrumento avaliativo: **todos os elementos de um item devem ser construídos tendo em vista o nível de dificuldade desejado, além de estarem de acordo com a etapa de ensino do estudante.**

Não se deve, portanto, aplicar um item fácil ou difícil com o intuito apenas de ajudar ou punir o aluno. A razão deve estar associada a uma informação que será importante para o trabalho pedagógico do professor ou da professora em sala de aula, que é verificar se o estudante alcançou determinado nível de aprendizagem.

Apresentamos, a seguir, uma sequência de 29 diretrizes para elaboração e revisão de itens de resposta selecionada, com base nas instruções de Haladyna (2004, p. 97-126).

Diretrizes relacionadas ao conteúdo do item

- 1) *Todo item deve estar relacionado a um conteúdo específico e a um único processo cognitivo específico.* Para que seja possível determinar especificamente qual conhecimento, habilidade ou competência o aluno domina ou não em um teste, é necessário que cada item não envolva mais de um processo cognitivo.
- 2) *Cada item deve se basear em algo importante para se aprender. Evite conteúdo trivial.* Deve-se dar prioridade a conteúdos enriquecedores, interessantes e de valor humanístico, e evitar aqueles que possam ser considerados pouco atraentes e de viés questionável. Exemplo:

Trivial: Qual o nome da principal substância encontrada na maioria dos cigarros?

Importante: Quais doenças são comumente associadas ao uso de cigarros?

- 3) *Explore maneiras mais criativas de avaliar um conteúdo.* Evite abordagens excessivamente conteudistas e que exijam termos e conceitos decorados. Em vez disso, busque focar na identificação de exemplos que ilustrem um conceito, princípio ou procedimento, ou na compreensão de paráfrases. Exemplo:

Método literal: Qual é a definição de metáfora?

- A) Apresentação de uma ideia pelo uso de palavras de sentidos opostos.
- B) Descrição exagerada por meio do contraste entre duas coisas diferentes.
- C) Produção de sentidos a partir da equivalência entre dois termos distintos.

Método criativo: Qual dos trechos a seguir possui uma metáfora?

- A) Ela é uma rosa, cheia de perfume e espinhos.
 - B) Hoje no café da manhã eu comi como um leão.
 - C) Os pelos do meu gato são fofos como algodão.
- 4) *Mantenha o conteúdo de um item independente do conteúdo de outros itens do teste.* É comum em testes que um item ajude o estudante a resolver outro. Por isso, ao elaborar uma sequência de itens, principalmente quando se baseiam em um mesmo suporte ou conteúdo, é importante verificar se o gabarito ou comando de algum dá pistas para a resolução dos demais, pois, caso isso ocorra, o estudante poderá chegar à resposta correta sem dominar aquilo que está sendo efetivamente avaliado. Exemplo:

De acordo com o texto, quem era o melhor amigo de João?

- A) Fernando.
- B) Marcos.
- C) Rodrigo.
- D) Rubens.

Qual das personagens era inimiga de João?

- A) Fernando.
- B) Marcos.
- C) Rodrigo.
- D) Rubens.

- 5) Evite *conhecimentos muito específicos ou muito amplos*. Itens devem situar-se entre esses dois extremos. Quando muito específicos, correm o risco de avaliar conteúdos triviais, como a memorização de um fato. Quando muito amplos, dão margem a exceções e ambiguidades. Exemplo:

Item específico: Quem é o autor de *Luzia-Homem*?

- A) Aluísio Azevedo.
- B) Bernardo Guimarães.
- C) Domingos Olímpio.
- D) Rodolfo Teófilo.

Item amplo: Qual é o escritor mais importante do naturalismo no Brasil?

- A) Aluísio Azevedo.
- B) Bernardo Guimarães.
- C) Domingos Olímpio.
- D) Rodolfo Teófilo.

- 6) Evite *itens baseados em opiniões*. Itens devem basear-se em fatos, conceitos, princípios e procedimentos cujo fundamento é de conhecimento público. Avaliar um estudante acerca de uma opinião sobre qualquer conteúdo será injusto, a menos que a opinião seja qualificada por alguma análise lógica ou evidência exposta durante o processo de ensino e aprendizagem. O exemplo a seguir mostra um item baseado em opinião, que, por conta disso, não possui gabarito.

Item baseado em opinião: Qual é o melhor romance da literatura brasileira?

- A) *Dom Casmurro*.
- B) *Gabriela, cravo e canela*.
- C) *Grande sertão: veredas*.
- D) *Vidas secas*.

- 7) Evite *itens com “pegadinhas”*. Esse tipo de item pode forçar o estudante a optar por um distrator pelos motivos errados. Nesse caso, o item pode ter um distrator que se aproxima muito do gabarito, um comando que apresenta mais de um sentido,

ou excesso de informações que terminam por confundir o estudante. Tudo isso deve ser evitado. O exemplo a seguir é de um item ambíguo em relação ao termo “romântico”, já que não faz referência explícita ao período da literatura que recebe esse nome, o que pode fazer com que o aluno considere o sentido genérico do termo e enxergue vários gabaritos.

Item com pegadinha: Qual destes escritores é romântico?

- A) Gonçalves Dias.
- B) Manuel Bandeira.
- C) Mario Quintana.
- D) Olavo Bilac.

Item adequado: Qual destes escritores pertence à corrente literária do romantismo?

- A) Gonçalves Dias.
- B) Manuel Bandeira.
- C) Mario Quintana.
- D) Olavo Bilac.

- 8) *Prefira itens verticais em vez de horizontais.* Embora itens horizontais ocupem menos espaço no teste, essa diagramação pode parecer mais desorganizada ao estudante e gerar confusão. Geralmente, itens verticais deixam o teste visualmente mais limpo. Exemplo:

Horizontal: Em um baralho de 52 cartas, quais as chances de se tirar uma carta de número primo?

- A) 16/52
- C) 20/52
- B) 32/52
- C) 36/52

Vertical: Em um baralho de 52 cartas, quais as chances de se tirar uma carta de número primo?

- A) 16/52
- C) 20/52
- B) 32/52
- C) 36/52

- 9) *Itens devem passar por revisões que os deixem mais claros.* É importante que todo item passe por mais de um olhar para garantir que as informações estão claras ao avaliando. Contudo, deve-se tomar cuidado para não abusar de revisões. Quando um item já demonstra ter um bom desempenho, ainda mais quando a habilidade avaliada é mais simples e objetiva, pode não haver mais necessidade de se buscar aperfeiçoamentos.
- 10) *Itens devem passar por revisões que levem em conta gramática, pontuação, formatação textual e ortografia.* Outro tipo de revisão indispensável diz respeito à norma culta da língua e aos formatos pré-estabelecidos. Por exemplo, ao perceber que uma alternativa contém erro de português, o estudante pode descartá-la de imediato, por considerá-la pouco atraente como gabarito. Além disso, a própria avaliação perderá parte de sua autoridade e de seu caráter profissional se não estiver de acordo com a norma-padrão da língua.
- 11) *Procure utilizar um vocabulário simples.* É absolutamente normal um item utilizar um texto científico como suporte, pedir o significado de alguma palavra menos conhecida, ou abordar variações linguísticas e termos técnicos de alguma área do conhecimento. Contudo, a linguagem utilizada nos comandos e nas alternativas deve buscar sempre ser a mais clara e objetiva possível, de modo que a habilidade exigida pelo item não se misture com a habilidade de compreender textos mais complexos.
- 12) *Evite excesso de textos em um único item.* Essa diretriz vale tanto para comandos quanto para suportes e situações-problema de um item. Recomenda-se, inclusive, para o suporte, estabelecer um limite de linhas de acordo com o ano escolar e a área avaliada. Ao mesmo tempo, as informações que contextualizam o item e que apresentam um problema a ser resolvido, bem como o comando que o direciona à resolução da questão, devem ser curtos e objetivos tanto quanto for possível.
- 13) *Evite o uso de humor caso se trate de uma avaliação formal ou em larga escala.* Estudos apontam que utilizar humor em itens não trará prejuízos em uma avaliação de sala de aula, na qual o professor ou a professora conheça os estudantes. Porém, no caso de avaliações maiores ou cujos resultados trarão impactos na trajetória dos alunos, recomenda-se que o recurso seja evitado, pois pode distraí-los do que é proposto. Essa diretriz, porém, não se aplica quando o objetivo é justamente verificar a capacidade do estudante de reconhecer efeitos de sentido decorrentes de humor e ironia, comum em matrizes de língua portuguesa.

Diretrizes relacionadas ao comando e ao suporte do item

- 14) *Deixe o mais claro possível aquilo que o item exige.* O que difere essa orientação de outras anteriores – como 11 e 12 – é o foco no procedimento que o item solicita ao estudante. Trata-se, portanto, de inserir todas as informações necessárias para que o estudante compreenda a operação a ser efetuada. Nesse caso, a falta de conteúdo costuma ser um problema mais comum do que o excesso. Evite, portanto, comandos vagos e ambíguos, que gerem dúvidas sobre o que é solicitado. Exemplo:

Comando vago: Um vaso com uma planta virou de lado. O que acontecerá?

Comando claro: Um vaso com uma planta em crescimento foi virado de lado. Uma semana depois, o que se espera que aconteça?

- 15) *Deixe o comando o mais breve possível.* Se o comando precisa estar claro e completo, deve-se também ter o cuidado para que não dê muitas voltas em torno do conteúdo. Às vezes, contextualizações excessivas podem prejudicar o entendimento do item. Exemplo:

Comando excessivo: Nascido na Europa e importado pelo Brasil, o romantismo trouxe para o mundo da arte formas de expressar ideias e sentimentos diferentes da maneira que predominava em períodos anteriores. Na literatura, uma das suas primeiras e principais obras é o romance *Os sofrimentos do jovem Werther*, de Johann Wolfgang von Goethe, que narra a paixão profunda e fatalista do jovem protagonista que dá o título à obra. Goethe é uma das maiores personalidades da Alemanha e da literatura no século XIX, sendo responsável pela produção de romances, peças de teatro, poemas, escritos autobiográficos, reflexões teóricas nas áreas de arte, literatura e ciências naturais e até mesmo correspondência epistolar com pensadores e personalidades da época, que é hoje fonte de pesquisa e análise. No Brasil, especificamente naquela que chamamos de primeira geração do romantismo, uma das novidades foi

- A) a dominância de um sentimento melancólico.
- B) a manifestação exacerbada do ideal nacionalista.
- C) o envolvimento com causas de cunho político.
- D) o pensamento excessivamente racionalista.

Comando conciso: Nascido na Europa, o romantismo trouxe para a literatura novas formas de expressar ideias e sentimentos. No Brasil, naquela que é considerada a primeira geração do romantismo, uma das principais características da poesia era o

- A) engajamento político.
- B) nacionalismo.
- C) naturalismo.
- D) sentimento fatalista.

16) Concentre a ideia principal do item no comando, não nas alternativas. Ao mesmo tempo que um comando deve ser conciso e objetivo, deve-se tomar cuidado para que não sejam esquecidas informações importantes. Recomenda-se, na maioria dos casos, que o estudante já identifique a operação cognitiva que lhe é exigida antes de entrar nas alternativas. Exemplo:

Comando escasso: O trecho “Essa tese, porém, é falsa” é

- A) destacado por um termo que repete uma informação.
- B) dividido por um termo que estabelece relação de oposição.
- C) finalizado por um termo que contradiz a afirmação inicial.
- D) introduzido por um termo que indica uma relação de lugar.

Comando completo: No trecho, “Essa tese, **porém**, é falsa”, o termo em destaque estabelece uma relação de

- A) ambientação.
- B) explicação.
- C) oposição.
- D) repetição.

17) Evite informações irrelevantes. Essa diretriz reforça a 12, a 14 e a 15 em torno do fato de que itens devem conter somente informações indispensáveis. Procure ter em mente que um item não é uma fonte de conhecimento para o estudante, de modo que uma informação não pode ser inserida pelo simples fato de ser interessante. Suportes devem ser de fato instigantes e atrair a atenção do aluno, mas, acima

de tudo, a razão de sua escolha deve levar em conta aquilo que será exigido do estudante. O mesmo pode ser dito do comando e das alternativas. Quanto mais rápido o estudante entender o item e resolvê-lo, melhor para ele e para a avaliação como um todo. Ao mesmo tempo, se a habilidade que está sendo avaliada exigir textos longos e muitas informações – como no caso de distinguir as informações principais das secundárias em um texto –, elas não serão consideradas irrelevantes e, portanto, não prejudicarão o item.

- 18) *Evite termos negativos em um comando (não, nunca, jamais...).* Diversos especialistas em avaliação concordam que termos negativos em um comando têm um efeito ruim, pois podem dificultar o entendimento da operação que está sendo exigida. Palavras como “não” e “nunca” podem confundir o estudante durante a sua leitura, de modo que ele tenha que retornar ao início e reler o item com a concepção de que todo o conteúdo, em vez de afirmado, está sendo negado. Essas orientações valem, principalmente, para itens que pedem para o estudante identificar a alternativa incorreta, que, além de serem confusos, invertem o processo, ao exigir que ele saiba não o procedimento correto, mas justamente o errado. Exemplo:

Comando com negação: Qual tendência estética a seguir não influenciou a Semana de Arte Moderna de 1922 em São Paulo?

- A) Concretismo.
- B) Cubismo.
- C) Expressionismo.
- D) Futurismo.

Comando adequado: Na Semana de Arte Moderna de 1922, o escritor Oswald Andrade trouxe da Europa as influências do

- A) Concretismo.
- B) Futurismo.
- C) Parnasianismo.
- D) Realismo.

Diretrizes relacionadas aos distratores e gabaritos

- 19) *Fique atento à quantidade necessária de alternativas.* Dependendo do ano escolar para o qual o item estiver indicado, o número de alternativas provavelmente já estará pré-estabelecido. Mas caso o elaborador tenha liberdade para essa questão, a quantidade de alternativas deve levar em conta a qualidade delas. A partir do momento em que se torna difícil produzir mais um distrator plausível, deve-se refletir se a quantidade de alternativas já não é a ideal e o item talvez já não deva ser concluído.
- 20) *Estabeleça uma lógica geral para ordenar as alternativas dos itens.* Para que uma alternativa não pareça muito atraente ao estudante, estando em destaque em relação às demais, procure organizá-las de acordo com uma lógica clara. Caso sejam números, coloque-os em ordem crescente ou decrescente; no caso de frases ou palavras, siga uma ordem alfabética. O importante é ficar evidente que a posição do gabarito segue critérios externos e não foi pensada apenas pelo elaborador. Exemplo:

Quanto custa um produto cujo valor inicial era R\$ 9,90 e recebeu um desconto de 20%?

- | <i>(Inadequado)</i> | <i>(Adequado)</i> |
|---------------------|-------------------|
| A) R\$ 7,92 | A) R\$ 1,98 |
| B) R\$ 8,91 | B) R\$ 4,95 |
| C) R\$ 1,98 | C) R\$ 7,92 |
| D) R\$ 4,95 | D) R\$ 8,91 |

- 21) *Procure variar a posição do gabarito em caso de sequências iguais ou muito ordenadas.* Na hora da montagem do teste, ou na elaboração de vários itens de um mesmo suporte, é comum surgir uma sequência de gabaritos com a mesma letra (A – A – A), ou seguindo uma ordem perfeita (A – B – C – D). Estudantes tendem a desconfiar de respostas assim e podem vir a alterar alguma das que haviam marcado para quebrar esse tipo de sequência. Nesses casos, para que ainda seja mantido o critério de ordenamento, sugere-se inverter a sequência das alternativas de algum dos itens – por exemplo, de ordem numérica crescente para decrescente.
- 22) *Evite distratores que possam englobar parte do gabarito.* Esse problema é comum em itens cujas alternativas são números, geralmente intervalos, em que um distrator poderia estar parcialmente correto. Por exemplo, se o gabarito é “12 a 18 anos”, qualquer distrator que passe por esse intervalo poderá ser contestado futuramente

como passível de ser gabarito. Outro exemplo é o caso abaixo, no qual o gabarito apresenta mais de um advérbio, de modo que, na versão inadequada, a alternativa D) estaria parcialmente correta.

No trecho “**Ontem**, conversei com ela como **nunca** havia feito antes...”, os termos em destaque estabelecem relação de

(Inadequado)

- A) lugar e negação.
- B) lugar.
- C) tempo e negação.
- D) tempo.

(Adequado)

- A) lugar e negação.
- B) lugar e oposição.
- C) tempo e negação.
- D) tempo e oposição.

- 23) *Mantenha as alternativas homogêneas em conteúdo e forma.* Alternativas muito diferentes das demais podem ser uma pista para o estudante, tanto para gabarito quanto para distratores. Sugere-se, portanto, que todas as alternativas apresentem paralelismo em relação: a) ao tamanho e à quantidade de palavras; b) aos termos que iniciam (tempo, modo, gênero, flexão...); e c) ao tipo de conteúdo. Caso não seja possível adequar todos os distratores, pode-se dividi-los em duplas: duas alternativas que seguem um tipo de paralelismo, enquanto outras duas ou três seguem outro. Exemplo:

Na Semana de Arte Moderna de 1922, o escritor Oswald Andrade trouxe da Europa as influências

(Inadequado)

- A) da arte concretista.
- B) da estética futurista.
- C) da poesia parnasiana.
- D) do realismo.

(Adequado)

- A) da arte concretista.
- B) da poesia parnasiana.
- C) do futurismo.
- D) do realismo.

- 24) *Evite a alternativa “Nenhuma das anteriores”.* Esse tipo de alternativa pode dar a impressão de que o elaborador ficou sem criatividade ou que não havia mais possibilidade de distrator a ser construído, de modo que o estudante raramente irá marcá-la. É uma alternativa pouco atraente, além de que, de um ponto de vista formal, espera-se que um item apresente um gabarito – caso contrário, o aluno só demonstrará saber o que *não* é certo ou verdadeiro.

- 25) Evite a alternativa “Todas as anteriores”. Assim como na orientação 24, esse tipo de alternativa é pouco atraente e costuma ser descartada pelo estudante. Além do mais, caso seja um distrator, é fácil de ser excluído, pois basta que o aluno saiba que pelo menos uma das demais alternativas não é correta.
- 26) Evite termos negativos, como “não”, “nunca”, “jamais” e “exceto”. Assim como na orientação 18, as alternativas também devem evitar o uso de termos negativos, pois podem confundir sobre o que de fato está sendo afirmado.
- 27) Evite termos que dão destaque ao gabarito. Da mesma forma que termos negativos podem tornar uma alternativa pouco atraente, palavras que dão ênfase podem ter um efeito oposto. “Totalmente”, “absolutamente” e “sempre”, por exemplo, tornam a afirmação forte e podem atrair mais o estudante. Ou, dependendo do tipo de conteúdo – principalmente histórico –, termos muito abrangentes como “sempre” e “totalmente” podem soar implausíveis.

Outro problema possível é um termo relevante que esteja presente no comando constar somente em uma alternativa. Isso pode dar a entender que aquela alternativa possui uma relação mais forte com o conteúdo abordado, tornando-a mais atraente. Nesse caso, deve-se ou retirar esse termo da alternativa ou inseri-lo nas demais e manter o paralelismo. Exemplo:

Em comparação ao romance *O Sertanejo*, de José de Alencar, uma das inovações de *Grande sertão: veredas*, de Guimarães Rosa, é

(Inadequado)

- A) conceder elementos universais ao **sertão** brasileiro.
- B) construir uma trama com personagens inventadas.
- C) criticar valores nacionais considerados atrasados.
- D) expor uma visão otimista sobre a população pobre.

(Adequado)

- A) conceder elementos universais ao **sertão** brasileiro.
- B) construir uma trama com personagens inventadas.
- C) criticar valores nacionais considerados atrasados.
- D) expor uma visão otimista sobre a política do **sertão**.

- 28) *Mantenha todos os distratores plausíveis.* Em um item, o gabarito deve estar incontestavelmente certo e os distratores totalmente errados. Mas a chave para o desenvolvimento de bons distratores é a plausibilidade, de tal forma que soem como uma resposta correta para aqueles que possuem pouco ou nenhum domínio do que é avaliado. Uma maneira de atestar a plausibilidade é o pré-teste do item, a partir do qual são atribuídas estatísticas às alternativas: por exemplo, uma alternativa marcada por apenas 1% de uma amostra grande e representativa de estudantes pode ser considerada pouco plausível, o que sugeriria reformulação.
- 29) *Explore erros comuns de estudantes ao elaborar distratores.* Distratores devem ser respostas às quais os estudantes poderiam chegar, independentemente de o item apresentá-las como alternativas. Itens com bons distratores são, portanto, aqueles que apresentam alternativas que os alunos marcariam caso realizassem um procedimento errado para resolvê-los. O exemplo a seguir apresenta um gabarito e um distrator bastante plausível para quem ainda não desenvolveu totalmente a habilidade de realizar adições entre dezenas:

Qual o resultado da operação $77 + 34$?

- A) 101
B) 111

4. O que você aprendeu até aqui?

Neste capítulo, você teve contato com os diferentes formatos de itens de resposta selecionada e com a estrutura de cada um. Foi possível verificar que, em sua maioria, esse tipo de instrumento de avaliação se divide entre enunciado, que pode conter um suporte ou uma situação-problema e possui um comando que direciona à resolução da tarefa; e alternativas, que se dividem entre gabarito e distratores.

Além disso, você pôde compreender que cada um dos formatos possíveis de itens de resposta selecionada é mais ou menos adequado para avaliar determinadas operações. Ao mesmo tempo, independentemente do formato adotado, foi destacado que itens de resposta selecionada são geralmente recomendáveis para se medir conhecimentos, habilidades cognitivas e competências menos complexas, cujo percurso para resolver a tarefa requerida não dê margem para múltiplos caminhos e diferentes conclusões.

Por isso, quando estão em jogo competências mais sofisticadas, que exigem posicionamento e raciocínio crítico, bem como habilidades motoras, outros instrumentos podem ser mobilizados, como itens de resposta construída e atividades em sala de aula com mediação do professor ou da professora. Isto é, tarefas cuja correção demanda análise mais cuidadosa e subjetiva, no lugar de apenas verificar se o gabarito foi marcado ou inserido. Neste capítulo, foi possível aprofundar-se nesse tópico por meio diferentes exemplos de itens, com suas respectivas análises pedagógicas, para a etapa de alfabetização e para os componentes curriculares de língua portuguesa e matemática.

Por sua vez, a objetividade dos instrumentos de resposta selecionada, principalmente do item tradicional de múltipla escolha, contribui para a facilidade e a conveniência de sua aplicação e correção, principalmente em avaliações em larga escala, nas quais são medidos os conhecimentos e as habilidades de uma grande população de estudantes. A correção automática, por meio de computador, bem como o simples lançamento no sistema da resposta marcada pelo estudante, permitem que esse processo seja realizado com agilidade e os resultados, que servirão para embasar as ações pedagógicas, sejam divulgados em tempo hábil.

Mas, para que isso seja possível, os instrumentos de resposta selecionada precisam ser elaborados e revisados com base em critérios rigorosos. Para isso, foi apresentado a você um percurso com 29 diretrizes que devem ser seguidas na construção de itens de resposta selecionada de qualidade, que dizem respeito ao conteúdo, ao suporte, ao comando, ao gabarito e aos distratores. Essas diretrizes devem ser estudadas com atenção e consultadas sempre que você for elaborar ou revisar algum item de resposta selecionada.

5. Referências

- BRASIL. Ministério da Educação. *Base Nacional Comum Curricular*. Brasília, 2018.
- CAED/UFJF. *Guia de Elaboração de itens: Língua Portuguesa*. Juiz de Fora, 2009.
- CAED/UFJF. *Guia de Elaboração de itens: Matemática*. Juiz de Fora, 2009.
- HALADYNA, Thomas M. *Developing and validating multiple-choice test items*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2004.
- HALADYNA, Thomas M. *Writing Test Items to Evaluate Higher Order Thinking*. Boston (MA): Allyn & Bacon, 1997.
- HOGAN, Thomas P. *Introdução à prática de testes psicológicos*. Tradução de Luiz Antonio Fajardo. Rio de Janeiro: LTC, 2006.
- PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. Petrópolis (RJ): Vozes, 2003.

6. Testando o que você aprendeu

Atividades de conhecimento geral sobre elaboração de itens

1) Sobre os itens de múltipla escolha, são feitas as seguintes afirmativas:

- I. Possibilitam mensurar a criatividade dos estudantes avaliados.
- II. Possibilitam precisão e padronização no processo de correção.
- III. São ideais para avaliações em larga escala, que têm o objetivo de medir o desempenho escolar.
- IV. São recomendados para mensurar habilidades mais complexas, como a produção textual.

São corretas apenas as afirmativas

- A) I e II.
- B) I e IV.
- C) II e III.
- D) III e IV.

2) Sobre as alternativas de resposta de um item de múltipla escolha, uma característica desejável é

- A) induzirem ao erro quem tem baixa proficiência.
- B) permitirem identificar o gabarito por exclusão.
- C) serem homogêneas quanto ao conteúdo e à forma.
- D) tratarem de processos cognitivos diferentes.

3) Os gabaritos dos itens de resposta selecionada que compõem um teste de proficiência devem

- I) apresentar a informação de forma objetiva.
- II) atrair a atenção dos estudantes.
- III) provocar debates entre especialistas da área.
- IV) ter semelhanças com os distratores.

Classifique cada uma dessas afirmações em (V) Verdadeira e (F) Falsa e escolha a alternativa correta.

- A) VVVF.
- B) VFFV.
- C) FVFF.
- D) FFFV.

4) Na avaliação em larga escala, os itens devem, preferencialmente,

- A) fazer uso de perguntas sobre tópicos específicos, para se testar a capacidade de memorização dos examinados.
- B) privilegiar perguntas que dependam da opinião dos examinados, para descobrir suas impressões sobre o tópico abordado.
- C) recorrer a recursos de humor, a fim de que os participantes não fiquem tensos durante a realização do teste.
- D) ser independentes entre si, de modo que a resposta dada a um item não influence as respostas dadas aos demais.

Atividades sobre itens da etapa de alfabetização

5) Leia o item abaixo para responder às questões 5.1 e 5.2.

VEJA A FIGURA ABAIXO.



QUAL É O NOME DESSA FIGURA?

- A) FARINHA
- B) FAXINA
- C) VACINA
- D) VARINHA

5.1) Esse item avalia qual habilidade?

- A) Compreender texto que articula linguagem verbal e não verbal.
- B) Distinguir letras do alfabeto de outros sinais gráficos.
- C) Ler palavras formadas por sílabas canônicas.
- D) Localizar informação explícita em textos de gêneros diversos.

- 5.2) Escreva um texto explicando a habilidade avaliada e o percurso cognitivo esperado para que os estudantes cheguem ao gabarito desse item.

6) Leia o item abaixo para responder à questão a seguir.

LEIA A FRASE ABAIXO.

A ENFERMEIRA CUIDA DO PACIENTE NO HOSPITAL.

Faça um X na imagem que representa o que você leu nessa frase.

-
- A)
- 
-
- B)
- 
-
- C)
- 
-
- D)
- 

Os distratores desse item são plausíveis, pois apresentam imagens que

- A) mostram mulheres em destaque.
- B) ocupam o mesmo espaço na página.
- C) possuem traços e estilos semelhantes.
- D) remetem à ideia de cuidado e de saúde.
- E) são acessíveis a alunos da alfabetização.

7) Leia o item abaixo para responder às questões 7.1 a 7.4.

Leia o texto abaixo.

CUIDADO DIÁRIO

ESCOVAR SEU CÃO TODOS OS DIAS PODE AJUDAR. NESSE MOMENTO, OS PELOS VÃO CAIR PERTO DE VOCÊ E FICARÁ MAIS FÁCIL MANTER A CASA LIMPA.

ESCOVE SEMPRE NA DIREÇÃO DOS PELOS. A ESCOVAÇÃO REMOVE AQUELES FIOS QUE JÁ ESTÃO SOLTOS E IMPEDE QUE ELES CAIAM NOS TAPETES, ROPAS E MÓVEIS.

YAZBEK, Letícia. Quanto pelo! *Recreio*. São Paulo, p. 11, jun. 2021. Fragmento. Adaptado.

Qual é o assunto desse texto?

- A) A limpeza das casas.
- B) A maneira de escovar os cães.
- C) A saúde dos animais.
- D) A sujeira que cai nos tapetes.

7.1) Qual é o gabarito desse item?

7.2) Qual habilidade esse item avalia?

7.3) O texto escolhido como suporte nesse item está adequado para estudantes do 1º ou 2º ano? Justifique sua resposta.

- 7.4) Escreva uma resolução justificada para o gabarito e os distratores que compõem esse item, detalhando o caminho cognitivo necessário para chegar à alternativa correta e os possíveis erros cometidos que podem levar o estudante a marcar cada uma das incorretas.

8) Leia as orientações abaixo para responder às questões 8.1 a 8.3.

Para testar sua aprendizagem, você precisará, neste exercício, concluir a elaboração do item, considerando as orientações a seguir:

Habilidade: Localizar informações explícitas.

Gênero escolhido como suporte: Receita culinária.

Etapa de escolarização: Estudantes em processo de alfabetização, do 1º ao 3º ano do ensino fundamental. Indique para qual ano você elaborou.

Suporte indicado:

POLENTA

INGREDIENTES:

400 GRAMAS DE FUBÁ

1 COLHER DE SOPA DE SAL

2 COLHERES DE SOPA DE MANTEIGA

2 LITROS DE ÁGUA

MODO DE FAZER:

COLOQUE A ÁGUA, O SAL E A MANTEIGA NA PANELA E LEVE AO FOGO.

DEPOIS QUE A ÁGUA ESTIVER QUENTE, COLOQUE O FUBÁ E MISTURE SEM PARAR.

QUANDO A MASSA ESTIVER GROSSA, ABAIXE O FOGO, TAMPE A PANELA E DEIXE COZINHAR POR 30 MINUTOS.

COLOQUE EM UMA VASILHA, CUBRA COM O MOLHO PREFERIDO E QUEIJO RALADO POR CIMA.

Disponível em: <https://comidas-tipicas.info/comidas-do-sul/receita-de-polenta.html>. Acesso em: 6 ago. 2021. Fragmento adaptado.

Comando sugerido: DE ACORDO COM ESSE TEXTO, PARA FAZER A POLENTA, O FUBÁ DEVE SER COLOCADO.

Quantidade de alternativas: 4

- 8.1) De acordo com as orientações acima, elabore o gabarito e os distratores para o item.
- 8.2) Escreva a resolução justificada de cada alternativa, apoiando-se nos conhecimentos sobre o processo de alfabetização de estudantes do 1º ao 3º ano. Indique qual ano você vai escolher.
- 8.3) Explique o porquê de ser importante avaliar essa habilidade, elencando elementos que a caracterizam.

9) Leia as orientações abaixo para responder às questões 9.1 e 9.2.

Habilidade: Ler palavras formadas por sílabas não canônicas.

Etapa de escolarização: Estudantes do 1º ano do ensino fundamental.

Suporte indicado:



Disponível em: <https://br.pinterest.com/pin/365706432221475205/>. Acesso em: 10 ago. 2022.

- 9.1) Elabore um item completo, seguindo as orientações acima.
- 9.2) Escreva um texto explicando em que consiste a habilidade e o percurso cognitivo esperado para que os estudantes cheguem ao gabarito do item que você elaborou.

Atividades sobre itens de língua portuguesa

10) Leia o item abaixo para responder às questões 10.1 e 10.2.

Leia os textos abaixo.

Texto 1**Motoristas enfrentam lentidão na volta do feriado prolongado no Rio**

A Concessionária Rio-Teresópolis (CRT) informou que os motoristas enfrentavam retenção com chuva em vários trechos da pista, sendo o maior na Serra, que vai do Km 89 ao Km 104, às 16h50. [...]

Por volta das 17h, a Via Lagos apresentou trânsito intenso e lento na extensão que vai do Km 1 até o Km 32. Isso representa mais da metade da via expressa, que tem 57 km.

A ponte Rio-Niterói apresentava trânsito lento na extensão que vai da Ilha de Mocanguê até o acesso à Avenida Brasil, devido ao grande fluxo de veículos. A previsão de tempo de travessia era de 20 a 25 minutos, segundo a CCR Ponte, às 17h05. [...]

Disponível em: <<http://g1.globo.com/rio-de-janeiro/transito/noticia/2013/11/motoristas-enfrentam-lentidao-na-volta-do-feriado-prolongado-no-rio.html>>. Acesso em: 8 jan. 2014. Fragmento.

Texto 2

Disponível em: <<http://www1.folha.uol.com.br/ilustrada/cartum/cartunsdiarios/#3/1/2014>>. Acesso em: 8 jan. 2014.

(P080205F5_SUP)

(P080205F5) Qual é a informação em comum nesses textos?

- A) A extensão da Ponte Rio-Niterói.
- B) A lentidão no trânsito no período de feriado.
- C) O estresse provocado pelos engarrafamentos.
- D) O trânsito intenso da Via Lagos.

10.1) Esse item avalia qual habilidade?

- A) Compreender o sentido global de textos que articulam linguagem verbal e não verbal.
- B) Diferenciar informações principais das secundárias em textos de gêneros variados.
- C) Identificar posições distintas entre duas ou mais opiniões relativas ao mesmo fato ou ao mesmo tema.
- D) Reconhecer diferentes formas de tratar uma informação na comparação de textos que tratam do mesmo tema.

- 10.2) Escreva um breve texto explicando a habilidade avaliada e o percurso cognitivo esperado para que os estudantes cheguem ao gabarito desse item.

11) Leia o item abaixo para responder à questão.

Leia o texto abaixo.

	<p>Uma vaca da raça Girolando que participa do torneio leiteiro da 41ª Exposição Agropecuária de Jataí, no sudoeste de Goiás, é recordista mundial em produção de leite. Em apenas um só dia, Dengosa, como foi batizada, já forneceu 107 kg da bebida.</p> <p>Por causa da marca, a vaca mineira recebe um tratamento de estrela. Passa a maior parte do tempo em uma cama de palha de arroz e com direito a climatizador para deixar o ar mais agradável e um ventilador para se refrescar do calor.</p> <p>Com apenas 6 anos e pesando 800 kg, Dengosa já acumula várias conquistas. [...] O veterinário responsável por Dengosa, Minoro Júnior, diz que cuidar de um animal tão especial é difícil, mas compensador. [...]</p>
5	<p>Com apenas 6 anos e pesando 800 kg, Dengosa já acumula várias conquistas. [...] O veterinário responsável por Dengosa, Minoro Júnior, diz que cuidar de um animal tão especial é difícil, mas compensador. [...]</p>

10 Um dos segredos para os bons resultados na produção de leite pode estar na alimentação balanceada que o bovino recebe.

Com a genética apurada e todas as conquistas obtidas, os responsáveis pelo animal já trabalham no intuito de lucrar ainda mais através dele. "A produção de embriões está a todo vapor, porque a procura pelo material genético dela está muito grande", diz Minoro Júnior.

Disponível em: <<http://migre.me/ow87u>>. Acesso em: 6 fev. 2015. Fragmento. (P050191H6_SUP)

(P050192H6) De acordo com esse texto, um dos segredos para a boa produção de leite pode ser

- A) o uso de uma cama de palha.
- B) o acúmulo de conquistas.
- C) a participação nos torneios.
- D) a alimentação balanceada.

Os distratores desse item são plausíveis, pois

- A) apresentam frases de curta extensão.
- B) contêm linguagem condizente com a etapa avaliada.
- C) possuem estrutura sintática semelhante.
- D) referem-se a informações presentes no suporte.
- E) relacionam-se a situações de conhecimento geral.

12) Leia o item abaixo para responder às questões 12.1 a 12.4.**Leia o texto abaixo.**

		Domingão
5		<p>Domingo, eu passei o dia todo de bode. Mas, no começo da noite, melhorei e resolvi bater um fio para o Zeca.</p> <p>– E aí, cara? Vamos no cinema?</p> <p>– Sei lá, Marcos. Estou meio pra baixo...</p> <p>– Eu também tava, cara. Mas já estou melhor.</p>
10		<p>E lá fomos nós. O ônibus atrasou, e nós pagamos o maior mico, porque, quando chegamos, o filme já tinha começado. [...]</p> <p>Saímos de lá, comentando:</p> <p>– Que filme massa!</p> <p>– Maneiro mesmo!</p> <p>Mas já era tarde, e nem deu para contar os últimos babados pro Zeca. Afinal, segunda-feira é dia de trampo e eu detesto queimar o filme com o patrão. Não vejo a hora de chegar o final de semana de novo para eu agitar um pouco mais.</p>

CAVÉQUIA, Márcia Paganini. Disponível em: <<http://migre.me/rP9xe>>. Acesso em: 16 out. 2015. Fragmento. (P121011H6_SUP)

(P121012H6) Nesse texto, a história tem início quando

- A) Marcos convida Zeca para ir ao cinema.
- B) o filme começa.
- C) o ônibus atrasa.
- D) Zeca aceita o convite feito por Marcos.
- E) Zeca e Marcos chegam ao cinema.

12.1) Qual é o gabarito desse item?**12.2) Qual habilidade esse item avalia?****12.3) O suporte é adequado para avaliar estudantes a partir de qual ano escolar?****Justifique sua resposta.****12.4) Escreva uma resolução justificada para o gabarito e os distratores que compõem esse item, detalhando o caminho cognitivo necessário para chegar à alternativa correta e os possíveis erros cometidos que podem levar o estudante a marcar cada uma das incorretas.**

13) Leia as orientações abaixo para responder às questões 13.1 a 13.3.

Para testar sua aprendizagem, você precisará, neste exercício, concluir a elaboração do item, considerando as orientações a seguir:

Habilidade: Identificar o objetivo comunicativo de um texto.

Gênero escolhido como suporte: Notícia.

Etapa de escolarização: Estudantes da segunda etapa do ensino fundamental, do 6º ao 9º ano. Indique qual ano você escolheu.

Suporte indicado:

Novo endereço da Feira Hippie divide opiniões

A possibilidade de mudança da feira de Arte, Artesanato e Produtores de Variedades da avenida Afonso Pena para a avenida Augusto de Lima, entre as ruas Barbacena e Araguari, no Barro Preto, divide opiniões de moradores e trabalhadores da região.

Os que são a favor argumentam que o bairro tem mais comércio que moradores, e o movimento traria mais lucro e lazer para o local. Aqueles que se posicionam contra a novidade reclamam do aumento do trânsito e do barulho, principalmente no início da manhã dos domingos.

A jornalista Silvânia Ariel mora na região há mais de 20 anos. Ela teme a desvalorização dos imóveis residenciais.

O empresário do ramo imobiliário Nilton dos Reis mora na Augusto de Lima há 60 anos, desde que nasceu. Ele vê com bons olhos a mudança.

Super Notícia, p. 6, 15 nov. 2009.

Comando sugerido: Esse texto tem como finalidade

Quantidade de alternativas: 4

- 13.1) Elabore o gabarito e os distratores para o item, de acordo com as orientações acima.
- 13.2) Escreva a resolução justificada de cada alternativa, apoiando-se na habilidade avaliada e observando o ano que você escolheu.
- 13.3) Explique por que avaliar essa habilidade é importante, elencando elementos que a caracterizam.

14) Leia as orientações abaixo para responder às questões 14.1 e 14.2.

Habilidade: Estabelecer relação causa/consequência entre partes e elementos do texto.

Etapa de escolarização: Estudantes do 9º ano do ensino fundamental.

Suporte indicado:

	De onde vieram os tomates?
5	A história do tomate é cheia de rumores, boatos e especulações, mas uma coisa é certa: essa fruta vermelha favorita de muita gente (sim, o tomate é uma fruta) não tem sua origem na Itália. Apesar do fato de ser um ingrediente essencial para massas, pizzas e saladas, o tomate é originário do México e da América Central.
10	O tomate em sua forma original, no entanto, não tinha nada a ver com esse globo vermelho que nós conhecemos e adoramos hoje em dia. Tratava-se de uma pequena fruta perfumada (imagine algo como o tomate cereja) que os grupos nativos americanos combinavam com "ahi", um tipo de pimenta para fazer um molho bem temperado. Embora os nativos americanos o tenham consumido por séculos, os tomates rapidamente ganharam uma má reputação nas Américas. Os colonizadores acreditavam que o tomate era venenoso e nenhum descendente europeu se atreveu a comer a fruta até o início do século 19 – com medo de morrer.
15	
20	Na verdade, credita-se à Fundação Americana Padre Thomas Jefferson o início do cultivo de tomate para consumo nos Estados Unidos. Os registros de Jefferson contam que ele plantava a fruta todos os anos em seu "Garden Kalendar" que manteve de 1809 a 1824. Talvez essa seja a primeira referência escrita do cultivo de tomate pelos colonizadores do Novo Mundo [...]. Seus registros meticulosos indicavam que ele frequentemente vendia seus tomates em mercados de Washington, além de apresentar diferentes usos para o mesmo em sua coleção pessoal de receitas.

Disponível em: <http://lazer.hsw.uol.com.br/origem-tomates.htm>. Acesso em: 13 jan. 2018. Fragmento.

- 14.1) Elabore um item completo, seguindo as orientações acima.
- 14.2) Escreva um breve texto explicando a habilidade e o percurso cognitivo esperado para que os estudantes cheguem ao gabarito do item que você elaborou.

Atividades sobre itens de matemática

15) Leia o item abaixo para responder às questões 15.1 a 15.3.

Ana e Paula colecionam figurinhas. Ana tem 255 figurinhas e Paula, 98. Quantas figurinhas Ana tem a mais que Paula?

- A) 353
- B) 267
- C) 157
- D) 143

- 15.1) Esse item avalia qual habilidade? Qual o contexto que o item atribuiu à operação que é empregada na resolução da tarefa proposta?
- 15.2) Qual é o gabarito desse item?
- 15.3) Apresente uma resolução detalhada para esse item, descrevendo o percurso cognitivo esperado para que os estudantes cheguem ao gabarito.

16) Leia as orientações abaixo para responder às questões 16.1 a 16.4.

Utilize o texto abaixo como base para elaboração de um item de resposta selecionada, com quatro alternativas, que avalie a seguinte habilidade: “Resolver problemas que envolvam a operação de subtração entre números inteiros”.

Um centro de pesquisas possui uma câmara fria, na qual são armazenados alguns experimentos. Em um dia, a temperatura no interior dessa câmara fria atingiu -18°C , enquanto a temperatura ambiente do centro de pesquisa, nesse instante, era de 21°C .

- 16.1) Qual comando você proporia para esse item?
- 16.2) Qual é o gabarito para esse item?
- 16.3) Quais distratores você proporia para esse item?
- 16.4) Escreva as justificativas tanto para o gabarito quanto para os distratores.

17) Leia as orientações abaixo para responder às questões 17.1 a 17.3.

No quadro abaixo, encontram-se o texto-base, com o suporte de um gráfico, e o comando para um item incompleto.

O gráfico abaixo apresenta o desempenho do Brasil em relação à quantidade de medalhas olímpicas conquistadas entre os anos de 1996 a 2012.



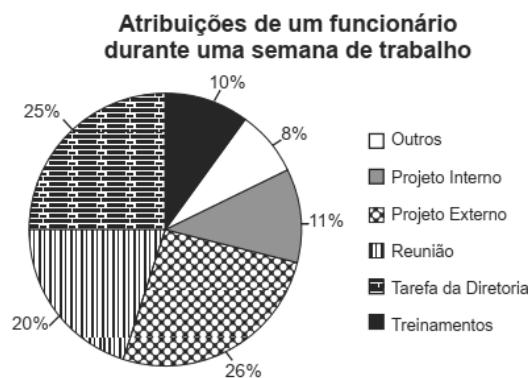
Disponível em: <<http://www.cob.org.br/pt/time-brasil/brasil-nos-jogos/medalhas-do-time-brasil/>>. Acesso em: 28 jul. 2015. *Adaptado para fins didáticos.

De acordo com esse gráfico, a quantidade total de medalhas obtidas pelo Brasil nas Olimpíadas de 2000 a 2008 é

- 17.1) Qual habilidade esse item avalia?
- 17.2) Complete o item, formulando quatro alternativas de resposta, com gabarito claro e distratores plausíveis, mantendo o paralelismo sintático e semântico entre as alternativas.
- 17.3) Escreva a resolução justificada de cada alternativa, apoiando-se na habilidade avaliada.

18) Leia o item abaixo para responder às questões 18.1 e 18.2.

O gráfico a seguir apresenta as atribuições de um funcionário durante uma semana de trabalho.



- 18.1) Elabore um comando e quatro alternativas para finalizar o item, de modo que ele avalie a habilidade “Corresponder representação gráfica de dados de uma pesquisa à sua respectiva representação em uma tabela”. Em seguida, indique o gabarito.
- 18.2) Elabore um comando e indique um gabarito para esse mesmo enunciado, de modo que o item agora avalie a habilidade “Interpretar dados representados graficamente”.

ELABORAÇÃO DE ITENS DE RESPOSTA CONSTRUÍDA

Glauker Amorim

Hilda Micarello

Josiane Toledo

Rosângela Veiga

1. O que você vai aprender aqui?

- Conceito de item de resposta construída e seu papel em uma avaliação educacional
- Características dos itens de resposta construída e seus diferentes formatos
- Habilidades mais recomendadas para serem avaliadas por itens de resposta construída de alfabetização, língua portuguesa e matemática
- Diretrizes para elaboração, revisão e correção bem-sucedidas de itens de resposta construída de alfabetização, língua portuguesa e matemática

2. Iniciando a construção do conhecimento

Como visto no Capítulo 2, as avaliações em larga escala e as atividades avaliativas em sala de aula são os instrumentos por meio do qual os variados tipos de itens podem e devem ser utilizados para se observar o desempenho dos estudantes. Se nas avaliações externas, principalmente no Brasil, são usados, essencialmente, itens de resposta selecionada, sendo em sua maioria os de múltipla escolha, no caso das atividades escolares o mais comum é o emprego de itens de resposta construída, que costumam ser chamados de questões abertas.

Esse modelo de item, se comparado aos itens de resposta selecionada, permite que o professor ou a professora visualize melhor as habilidade e os conhecimentos desenvolvidos pelos alunos por meio das respostas formuladas por eles. Embora a sua elaboração, a princípio, seja mais simples, exige também alguns critérios básicos para a sua construção e posterior correção.

Nesse sentido, podemos afirmar que os itens de resposta construída apresentam as seguintes vantagens:

- > permitem avaliar habilidades mais complexas que exijam, por exemplo, a aplicação de um conhecimento e o desenvolvimento de um raciocínio;
- > permitem avaliar habilidades mais complexas que exijam, por exemplo, desenvolvimento de argumentação, análise, raciocínio crítico e propostas de intervenção na realidade, bem como a construção de objetos;
- > possibilitam, de acordo com a habilidade e com o conhecimento avaliados, observar a capacidade de elaborar respostas criativas;
- > abrem pouco espaço para o acerto ao acaso, o chute;
- > viabilizam identificar os níveis de desenvolvimento de uma dada habilidade ou conhecimento – por exemplo, na alfabetização, por meio da escrita de uma palavra, em que se pode observar em que estágio a criança se encontra na aprendizagem da escrita e na apropriação das regras do sistema de escrita alfabetica (SEA).

Mesmo diante de tais vantagens, os itens de resposta construída também têm suas desvantagens, as quais elencamos a seguir:

- > demandam mais tempo para correção;
- > exigem mais tempo para serem respondidos, sendo importante saber dimensionar o tempo disponível para a realização da avaliação e o número e a complexidade das tarefas a serem demandadas nos itens;
- > podem ser suscetíveis à subjetividade de quem vai realizar a correção.

A elaboração dos itens de resposta construída é, a princípio, mais simples, se comparada aos de resposta selecionada, embora exijam, também, alguns princípios universais para sua construção e, posteriormente, uma maior atenção para a correção. Por isso, é importante a definição de critérios de elaboração e de correção bem definidos, de modo a se evitar o efeito da subjetividade no momento de se avaliar as respostas dos estudantes.

Portanto, para que um item de resposta construída possa realmente fornecer informações sobre habilidades e conhecimentos desenvolvidos pelos estudantes, é fundamental que tenham qualidade técnica e pedagógica. Isso significa que devem ser seguidas orientações técnicas e pedagógicas para a elaboração tanto dos itens quanto das chaves de correção, também chamadas de guias de correção. Além disso, seguir orientações para garantir a qualidade de um item de resposta construída possibilita que outros professores, mesmo que não tenham elaborado os itens, possam utilizá-los, corrigi-los, interpretar seus resultados e fazer uso deles, direcionando as ações pedagógicas a partir das informações coletadas por meio das respostas dos estudantes.

Visando instrumentalizá-lo quanto às características e aos usos dos itens de resposta construída, bem como às habilidades e aos conhecimentos passíveis de serem avaliados por meio deles, este capítulo apresentará um percurso em que retomaremos o conceito de item, já visto no capítulo anterior, porém sob a óptica da especificidade de tarefas de resposta construída. Em seguida, passaremos aos diferentes tipos de itens de resposta construída, com seus respectivos exemplos, de modo a entendermos sua aplicabilidade em atividades avaliativas da etapa de alfabetização e dos componentes de língua portuguesa e matemática.

3. Construindo o conhecimento

3.1 O que é um item de resposta construída?

A presença do item de resposta construída em avaliações externas, antes rara, vem se tornando cada vez mais comum, dada a necessidade de se buscar informações mais amplas e fidedignas sobre o desempenho e as aprendizagens dos estudantes, de modo que o uso de diversos tipos de itens tem se revelado essencial. No que diz respeito ao contexto da avaliação interna, realizada comumente no cotidiano escolar, mesclar instrumentos avaliativos permite uma maior abrangência de habilidades, competências e conhecimentos que compõem o currículo escolar.

Os itens de resposta construída são assim denominados pois exigem que, a partir de um dado estímulo, que pode ser um comando, uma instrução, ou uma situação-problema, o estudante construa a sua resposta. Portanto, para que as respostas sejam corrigidas, é necessária também, como já mencionado, a elaboração de chaves ou guias de correção, especificando as expectativas de resposta de referência, bem como a pontuação a ser atribuída a tais respostas.

São muitos os tipos de itens e atividades que podem ser compreendidos como itens de resposta construída, conforme apontam Osterlind (2002) e Haladyna e Downing (2006), os quais também podem ser chamados, em alguns casos, de itens de *performance*. Por isso, neste capítulo, iremos nos dedicar aos tipos e formatos de itens de resposta construída mais comumente utilizados tanto na avaliação educacional em larga escala, quanto na avaliação realizada no cotidiano escolar, podendo envolver respostas curtas ou respostas mais extensas.

3.2 Diferentes modelos de itens de resposta construída

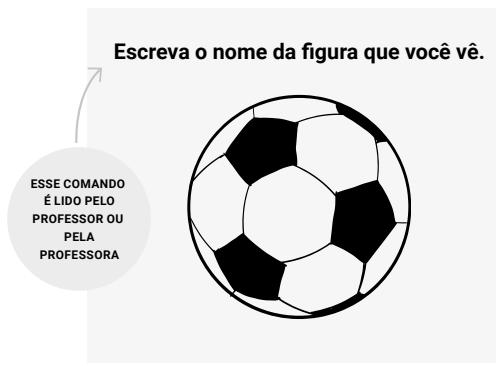
Haladyna e Rodriguez (2013) categorizam os itens de resposta construída em dois grupos: com respostas objetivas e com respostas subjetivas. A seguir, detalhamos cada um deles.

3.2.1 Itens de resposta construída com respostas objetivas

Esse tipo de item é, frequentemente, corrigido a partir de critérios de certo ou errado, principalmente quando demandam respostas curtas do estudante. Ainda assim, é possível estabelecer gradações de acerto para as respostas, como no caso da matemática, em que o aluno pode errar o resultado final, mas realizar corretamente a maioria dos cálculos; ou na alfabetização e em língua portuguesa, em que o estudante pode seguir o caminho correto, mas ainda cometer alguns erros ao escrever uma palavra ou frase. Nesses casos, para corrigir as respostas, é necessária uma chave de correção, que apresentaremos mais adiante.

Nesta seção, vamos analisar por quanto alguns exemplos que exigem respostas curtas, com pouca margem para desvios.

Resposta curta a partir de uma imagem: esse é um formato de item muito comum nos anos iniciais do ensino fundamental, durante o processo de alfabetização.



Resposta: BOLA

Resposta curta com situação-problema: comumente utilizado em situações de ensino e avaliação de habilidades e conhecimentos da matemática, em todas as etapas da educação básica.

Se um chocolate custa 1 real e Ana comprar 5 chocolates, quanto ela vai pagar no total?

Resposta: 5 reais

Resposta curta que exige o conhecimento de fatos: em alguns casos, pode haver mais de uma maneira de se dar a resposta relativa ao fato. A seguir, trazemos um exemplo de História.

A respeito da Primeira Guerra Mundial, responda à seguinte questão:

A) Qual foi o principal fato responsável pelo início da Primeira Guerra Mundial?

Resposta:

O assassinato do arquiduque Francisco Ferdinando.

Ou

O assassinato do herdeiro do trono austro-húngaro.

Ou

O assassinato de Francisco Ferdinando, arquiduque do Império Austro-Húngaro.

Preenchimento de lacunas: costuma envolver uma única frase com um ou dois espaços em branco. Normalmente, esse modelo de item avalia a recordação de um conhecimento, tratando-se, pois, de um processo com baixa demanda cognitiva. Para esse tipo de abordagem, as questões de múltipla escolha são mais recomendadas.

O arcadismo brasileiro sofreu forte influência de um movimento intelectual e cultural que nasceu na Europa, a partir do século XVII, denominado

_____.
Resposta: Iluminismo.

Cloze: nesse tipo de atividade, o aluno intervém no corpo do texto, que apresenta alguns termos suprimidos. Trata-se de um formato bastante parecido com o item de resposta selecionada de preenchimento, apresentado no Capítulo 2, com o diferencial de que, aqui, o estudante não escolhe uma palavra dentre duas alternativas a cada lacuna, preenchendo por conta própria o espaço em branco.

Esse tipo de item pode ser utilizado para avaliar compreensão e fluência de leitura, bem como o domínio de vocabulário, pois exige que o estudante reconheça o sentido do contexto e acesse seu léxico mental de modo a preencher a lacuna.

No exemplo a seguir, foram apagadas todas as sétimas palavras em sequência no texto. As respostas corretas encontram-se entre parênteses e podem sofrer variações, permitindo-se o uso de sinônimos ou termos com significados próximos que mantenham a coerência do texto.

Desenhos são uma fonte de informação _____ (e) uma maneira especial de obter revelações _____ (sobre) a mente. Diferentes disciplinas, como a _____ (psicologia), a educação e as artes, buscaram _____ (investigar) essa atividade. Muitos pesquisadores concordam que _____ (os) desenhos são uma importante forma de _____ (expressão) utilizada pelas crianças para comunicar naturalmente _____ (os) seus pensamentos, suas emoções e a _____ (maneira) de ver o mundo ao seu _____ (redor).

Fonte: STUDART, 2008.

Ditado: a aplicação de um ditado em sala de aula exige uma correção objetiva, isto é, que aponte se o estudante escreveu ou não a palavra corretamente. Ao mesmo tempo, é possível levar em conta gradações de acerto, o que permite verificar diferentes níveis de desempenho dos estudantes, tornando esse tipo de item um pouco mais complexo do que os anteriores de resposta objetiva.

Ditados são instrumentos que podem ser aplicados como avaliação interna principalmente no início do ano escolar, de modo a se verificar o desenvolvimento das habilidades de escrita dos estudantes. As etapas que correspondem à sua elaboração, aplicação e correção exigem um detalhamento maior.

O ponto de partida é a escolha de um tema, que pode nascer de uma história que a professora ou o professor tenha contado, de um fato ocorrido na comunidade, de um assunto que esteja sendo estudado em outro componente curricular, dentre outras possibilidades. O segundo momento é a definição de critérios para a escolha das palavras que comporão o ditado. Tal escolha deve ser orientada pela clareza de quais aspectos devem ser objeto de avaliação. No exemplo que apresentaremos, a ser aplicado no início do ano letivo do 3º ano do ensino fundamental, a atividade tem o objetivo de produzir um diagnóstico que permita identificar desde estudantes que ainda não desenvolveram a hipótese alfabética até aqueles que já desenvolveram, inclusive, alguns aspectos da escrita ortográfica, em um contexto no qual ainda se sabe pouco sobre cada um. A escolha da etapa de ensino tomou como base habilidades relacionadas à escrita de palavras previstas para serem desenvolvidas até o 2º ano do ensino fundamental.

Diante desse objetivo, os critérios para a escolha de palavras foram os seguintes: (i) palavras compostas exclusivamente por sílabas no padrão consoante/vocal e palavras compostas por sílabas em outros padrões silábicos; (ii) palavras pouco extensas e palavras mais extensas; (iii) palavras que apresentem marcas de nasalização.

Com base nesses critérios, as palavras escolhidas para compor o ditado foram organizadas em cinco grupos, apresentados no Quadro 1.

Quadro 1: Grupos de palavras segundo critérios de complexidade

GRUPOS DE PALAVRAS PARA COMPOR O DITADO		
GRUPOS	CARACTERÍSTICAS	PALAVRAS
1	Palavras dissílabas, compostas exclusivamente por sílabas canônicas	BOTE - REMO
2	Palavras trissílabas, compostas exclusivamente por sílabas canônicas	PIRATA - SACOLA
3	Palavras dissílabas, compostas por diferentes padrões silábicos	PRAIA - BARCO
4	Palavras trissílabas, compostas por diferentes padrões silábicos	NAVIO - MARINHA
5	Palavras com marcas de nasalização	SUNGA - ÂNCORA

Assim como nos itens apresentados anteriormente, é necessário estabelecer critérios de correção. Portanto, uma rubrica para correção do ditado faz-se necessária. Esse instrumento consiste em uma tabela na qual são registrados os possíveis desempenhos dos estudantes em relação ao que está sendo avaliado. Uma possibilidade de rubrica para a correção desse ditado é a que apresentamos no Quadro 2. Neste quadro, incluímos algumas possibilidades de escrita que se adequariam a cada um dos desempenhos possíveis dos estudantes³.

Quadro 2: Rubrica para correção do ditado de palavras

Aluno: _____		A palavra foi escrita de forma convencional	A palavra foi escrita de forma alfabética	A palavra foi escrita de forma silábico-alfabética	A palavra foi escrita de forma silábica	A palavra foi escrita de forma pré-silábica	A palavra foi representada por uma garatuja ou desenho
Data: _____							
Grupo 1	BOTE	BOTE	POTE – POTI BOTI	BOT BTE BTI	BT	GSTV	
	REMO	REMO	REMU	RMO REM	RM	AOED	
Grupo 2	PIRATA	PIRATA	BIRATA	PIRTA	PRT – IRT - PRA	UERTDS	0-0-0-0-0-0
	BALEIA	BALEIA	PALEA	PALA	BLA - PLI	AEIATR	
Grupo 3	PRAIA	PRAIA	PAIA	PAA	AI	SURTE	oo
	BARCO	BARCO	BACO	BAO	BO	DFRVY	
Grupo 4	NAVIO	NAVIO	NAVO	AVIO	AIO	SLTRFV-GTE	
	MARI-NHA	MARINHA	MARINA	MAIA	MRA	GHVBR-TFC	
Grupo 5	SUNGA	SUNGA	SUGA	SUA	UA	AIO	
	ÂNCO-RA	ANCORA	ACORA	ACOA	AOA	VQYR	

3. Para um maior conhecimento sobre as hipóteses que os estudantes formulam sobre a escrita e como identificá-las, bem como outras possibilidades, consultar Ferreiro e Teberosky (1999).

Como é possível observar, a rubrica de correção deve ser preenchida individualmente, para cada estudante. Com base na escrita apresentada para cada palavra, o professor ou a professora marcará um X na coluna correspondente, de modo a identificar não apenas se o estudante errou ou acertou a escrita da palavra, mas a qualidade de seu erro. Esses resultados deverão, por sua vez, embasar o planejamento pedagógico e orientar as ações a serem realizadas em sala de aula

3.2.2 Itens de resposta construída com respostas subjetivas

Itens de resposta construída que exigem respostas subjetivas, por sua vez, adéquam-se à avaliação de demandas cognitivas mais complexas, refletindo a habilidade do estudante e a sua capacidade de desenvolver um raciocínio crítico. Nesse tipo de atividade, o aluno pode demonstrar a capacidade de articular informações, posicionar-se e expressar-se por escrito ou até oralmente.

Esse tipo de item pode solicitar desde uma resposta elaborada em poucas sentenças ou frases, até a produção de uma redação, o que é muito comum na área de linguagens e de ciências humanas⁴. Para serem corrigidos, uma chave de correção se faz necessária.

Por demandarem um raciocínio mais complexo e subjetivo, o que dá margem a múltiplos caminhos por parte do estudante, essas atividades precisam trazer orientações bem definidas, de modo que o comando para a tarefa deve:

- › mostrar com clareza o processo a ser seguido por meio de verbos adequados: narrar, descrever, argumentar, analisar, persuadir, informar ou justificar, por exemplo;
- › indicar o tempo disponível para execução da tarefa;
- › sugerir ou deixar espaço que indique a extensão da resposta, podendo, quando necessário, limitar o número de linhas, como ocorre, por exemplo, nas redações do Enem.

4. Em relação à matemática, os itens de resposta construída com respostas subjetivas não são, em geral, empregados durante a educação básica.

Exemplo de item de escrita de frase a partir de uma imagem:

Veja a cena abaixo.



Escreva uma frase contando o que acontece nessa cena.

Fonte: CAEd/UFJF.

Esse item, bastante comum em avaliações da alfabetização, solicita ao estudante que escreva uma frase que seja plausível em relação à imagem apresentada, ou seja, que mantenha uma relação de sentido com o que o estudante vê. Além disso, esse tipo de item avalia também o quanto o estudante já assimilou dos conhecimentos e habilidades relacionados à língua escrita, normalmente no que diz respeito à apropriação do sistema de escrita alfabética, isto é, ao domínio da ortografia.

Exemplo de item de resposta construída em que se solicita a redação de uma narrativa:

Imagine que, ao apertar o botão da televisão da sua casa, uma luz muito forte apareceu e puxou você para dentro do aparelho. Escreva um texto contando o que você viveu dentro da televisão, lembrando-se de falar de coisas que você viu e fez, e também de quem participou. Dê um título à sua história.

Fonte: CAEd/UFJF.

Nesse exemplo, por se tratar de um enunciado para a elaboração de uma narrativa, pode-se solicitar a produção de um texto ficcional ou de uma história verdadeira, como um relato. Como se pode ver, o enunciado costuma trazer os seguintes tipos de verbos: contar, descrever ou escrever (“escreva uma história”).

Exemplo de item de resposta construída em que se solicita a realização de uma inferência com a aplicação de um conhecimento específico:

Leia o texto abaixo.

Helena

Capítulo III

[...] Era uma moça de dezesseis a dezessete anos, delgada sem magreza, estatura um pouco acima de mediana, talhe elegante e atitudes modestas. A face, de um moreno-pêssego, tinha a mesma imperceptível penu-gem da fruta de que tirava a cor; naquela ocasião tingiam-na uns longes cor-de-rosa, a princípio mais rubros, natural efeito do abalo. As linhas pu-ras e severas do rosto parecia que as traçara a arte religiosa. Se os ca-belos, castanhos como os olhos, em vez de dispostos em duas grossas tranças lhe caíssem espalhadamente sobre os ombros, e se os próprios olhos alçassem as pupilas ao céu, disséreis um daqueles anjos ado-les-centes que traziam a Israel as mensagens do Senhor. Não exigiria a arte maior correção e harmonia de feições, e a sociedade bem podia con-tar-se com a polidez de maneiras e a gravidade do aspecto. Uma só coisa pareceu menos aprazível ao irmão: eram os olhos, ou antes o olhar, cuja expressão de curiosidade sonsa e suspeitosa reserva foi o único senão que lhe achou, e não era pequeno. [...]

ASSIS, Machado de. Disponível em: <http://www.dominiopublico.gov.br/download/texto/bn000079.pdf>.

Acesso em: 10 ago. 2020.

Aponte um recurso muito utilizado no romantismo, predominante nesse frag-mento, e, em seguida, explique o efeito provocado por seu emprego. (5 linhas)

Aqui, para responder ao que foi solicitado, o estudante precisa demonstrar conhecimento de características do romantismo (“Aponte um recurso muito utilizado no romantismo”), o que, no caso desse texto, se manifesta pelo emprego recorrente da descrição da figura feminina. Em seguida, solicita-se que se explique qual o efeito gerado por esse recurso. Para concluir com sucesso a tarefa, o estudante precisa colocar em ação mais conhecimentos acerca do romantismo e da obra tomada para estímulo, de modo a ser capaz de dizer que o uso sistemático da descrição, que gera um efeito de exagero nas características da personagem, cria uma imagem idealizada da figura feminina.

Exemplo de item de preenchimento de quadro:

8. Gideon Mantell pensou que o dente podia ter pertencido a diferentes tipos de animais. Completa o quadro, mostrando o que o fez pensar assim.

Tipo de animal	O que o fez pensar assim
Um comedor de plantas	O dente era achatado e com arestas.
Uma criatura gigantesca	[A]
Um réptil	[B]

Fonte: IAVE, 2011, p. 72.

Esse item, aplicado em Portugal no Estudo Internacional de Progresso em Leitura (PIRLS), no ano de 2011, avalia habilidades de leitura de estudantes do 4º ano do ensino fundamental. Para a sua elaboração, foi utilizado um texto informativo, “O mistério do dente GIGANTE”⁵. Nesse caso, solicita-se que o estudante interprete e integre informações⁶ constantes no texto, relacionadas às informações presentes na primeira coluna, de modo a preencher os espaços em branco na segunda coluna.

-
5. Para não apresentarmos excesso de informações e corrermos o risco de confundir nossos leitores, optamos por suprimir o texto de suporte.
6. Interpretar e integrar informações é um processo de compreensão de leitura que solicita ao leitor relacionar seus conhecimentos com o significado do texto. Para isso, o leitor pode deter-se em partes específicas do texto ou no texto como um todo.

Exemplo de item de resposta curta:

Porque é que <<Marta Encontra uma Solução>> também seria um bom título para esta história?

Apresenta uma razão.

Fonte: IAVE, 2016, p. 22.

Novamente, temos um item aplicado no PRILS, em Portugal, no ano de 2016. Aqui, a habilidade avaliada é a capacidade de os estudantes analisarem e avaliarem conteúdo e elementos textuais (IAVE, 2016, p. 9), a partir da leitura do texto literário “Marta e a galinha vermelha”⁷. Para responder ao que foi solicitado, o estudante precisa ler o texto como um todo e avaliar a pertinência do que está sendo proposto, de modo a concluir que, de acordo com a leitura realizada, o título apresentado na pergunta é adequado, pois Marta conseguiu resolver o seu problema.

3.3 Critérios para elaborar itens de resposta construída

Para que possamos elaborar itens de resposta construída de boa qualidade técnica e pedagógica, deve-se atentar, nesse processo, para alguns elementos considerados universais:

- **Conteúdo e demanda cognitiva:** o conteúdo tratado na questão pode ser objetivo, com resposta em termos de certo ou errado, ou subjetivo, cuja resposta exigirá uma avaliação detalhada de um corretor/avaliador. Essa diferença interfere no tipo de correção a ser executada e mesmo nas condições de aplicação.
- **Instrução / comando ao respondente:** pode concretizar-se em uma única frase ou em uma sequência mais longa de instruções sobre a natureza do que se deseja avaliar.

7. Pelo mesmo motivo já apontado anteriormente, optamos por suprimir o texto de suporte.

- **Contextualização / suporte:** diz respeito a textos ou situações-problema empregados como estímulo para a resposta do estudante. De modo semelhante ao que se observa nos itens de resposta selecionada de língua portuguesa, pode se tratar de um texto cujo trecho irá embasar a elaboração do item, ou funcionar como contextualização para orientar a escrita da resposta, como no caso da redação do Enem.
- **Condições de execução:** são definidas por quem elabora os itens e serve para estruturar os testes. Essas condições envolvem, portanto, orientações para realização de atividades e/ou de um teste, que podem dizer respeito, por exemplo, ao tempo de aplicação de uma questão específica ou do teste como um todo. Tais orientações devem ser claras e o mais completas possível, de modo que, ao se realizar a correção, não pairem dúvidas sobre o desempenho.
- **Chaves e critérios de correção:** os critérios de correção e pontuação devem ser estabelecidos quando da elaboração da questão e do teste.

Além desses elementos universais, Haladyna e Rodriguez (2013) apontam algumas diretrizes⁸ para que se possa elaborar itens de resposta construída com o objetivo de se observar as aprendizagens dos alunos, as quais trazemos a seguir.

1) Preocupações de conteúdo

- ✓ Tenha clareza sobre o domínio de conhecimento e habilidades a serem avaliados.
- ✓ Certifique-se de que o formato seja apropriado para a demanda cognitiva pretendida.
- ✓ Procure garantir comparabilidade da construção entre as tarefas avaliativas.

2) Formatação e preocupação de estilo

- ✓ Edite e verifique as instruções dos itens.

8. Tradução e adaptação nossa.

3) Escrever as instruções / comando

- ✓ Defina claramente as orientações, expectativas para o formato de resposta e demandas de tarefas.
- ✓ Forneça informações sobre os critérios de pontuação.
- ✓ Evite exigir suposições implícitas e recursos irrelevantes para a realização das tarefas.

4) Preocupações de contexto

- ✓ Considere a diversidade cultural e a acessibilidade.
- ✓ Certifique-se de que a complexidade linguística é adequada para a etapa de escolaridade à qual a atividade avaliativa se destina.

No item de resposta construída, mesmo que o estudante tenha maior liberdade para executar a tarefa, é preciso conduzi-lo por meio de orientações claras que lhe permitem compreender qual tipo de raciocínio deve ser realizado e qual conhecimento ou habilidade está sendo solicitado pelo item. Portanto, além das diretrizes mencionadas, deve-se ter atenção à construção do comando e aos estímulos apresentados, que podem ser desde uma única frase até um parágrafo mais amplo, além de virem acompanhados de textos não verbais ou multissemióticos.

Além disso, conforme já destacamos, para os itens cujas respostas são de caráter subjetivo, é fundamental que sejam construídas chaves de correção, que devem levar em consideração o conteúdo, o processo cognitivo demandado e, quando se tratar de uma produção textual, devem também considerar a natureza da tipologia textual suscitada pelo enunciado.

Essas chaves de correção são guias para se avaliar e pontuar a tarefa demandada na atividade. Para isso, pode ser realizada uma correção holística ou analítica.

Vejamos o que caracteriza cada um desses tipos de chave de correção:

Chave de correção holística: considera-se o desempenho de modo que todos os componentes da escrita da resposta são compreendidos de forma integrada. Esse tipo de chave de correção, quando aplicado à correção de produções textuais, está mais sujeito à subjetividade do avaliador / corretor da atividade, pois se relaciona muito mais com uma impressão geral do que foi produzido pelo estudante.

Chave de correção analítica: caracteriza-se por trazer um conjunto de elementos que são avaliados de modo independente, que apresentam diferentes graduações de acerto. Um exemplo desse tipo de chave de correção é a matriz de competências para a redação do Enem.

Comparando-se os dois tipos de correção, principalmente quando nos referimos à produção textual, a correção analítica apresenta algumas vantagens. Por descrever cada componente e aspecto a ser avaliado, detalhando-se uma graduação de níveis de acerto com sua respectiva atribuição de pontuação, ela permite minimizar a subjetividade do responsável pela realização da correção da atividade avaliativa.

Acrescente-se, ainda, que o processo de correção detalhada da resposta do estudante passa a ser uma excelente oportunidade para se conhecer seus pontos fortes e as suas fragilidades. Isso permitirá a construção de um diagnóstico dos aspectos que ainda precisam ser recuperados ou aprofundados pelo professor ou pela professora em sala de aula.

Entendemos, portanto, que uma correção analítica se torna mais adequada para atividades aplicadas no contexto de avaliações entendidas como formativas. Vamos conhecer alguns exemplos.

Exemplo de chave de correção analítica para correção de itens de escrita de palavras a partir de um ditado na alfabetização:

Ortografia	
Conceito	Descrição
A	O estudante escreveu a palavra corretamente, configurando uma escrita em nível ortográfico, sendo válida somente a palavra ditada.
B	O estudante escreveu a palavra com desvios, configurando uma escrita em nível alfabetico.
C	O estudante escreveu a palavra com desvios, configurando uma escrita em nível silábico-alfabetico.
D	O estudante escreveu a palavra com desvios, configurando uma escrita em nível silábico (a escrita produzida pelo estudante apresenta relação com a pauta sonora das letras que compõem a palavra).
E	O estudante produziu escrita em nível pré-silábico (apresentou letras aleatórias para representar a palavra, desenhos / garatujas) e/ou grafou outra palavra que não guarda estreita relação com a palavra ditada (provavelmente reproduzida de memória).

Fonte: CAEd/UFJF.

Essa chave de correção, também denominada guia de correção ou rubrica de correção, é destinada à avaliação da escrita de uma palavra ditada. A atividade de escrever palavras é algo corriqueiro no contexto da alfabetização, sendo utilizada para se mensurar a aprendizagem do sistema da escrita alfabetica, tanto no cotidiano das atividades escolares quanto na avaliação educacional em larga escala.

Trata-se de um exemplo de chave de correção que considera a resposta do estudante em uma perspectiva graduada, do mais correto, indicado pelo conceito A, ao menos correto, indicado pelo conceito E. Por meio da descrição de cada um dos conceitos, constatamos que a correção da tarefa de escrita de uma palavra a partir do ditado pauta-se por uma orientação que considera as hipóteses da escrita, conforme apresentado por Ferreiro e Teberosky (1999).

Isto é, a chave de correção considera não apenas se o estudante escreveu corretamente a palavra, revelando dominar os princípios que regem o sistema alfabetico da língua portuguesa no Brasil. Mas, principalmente, a partir da análise da resposta da criança, permite conhecer em que momento desse processo de codificação a criança se encontra, o que permitirá, como já dito anteriormente, orientar as práticas pedagógicas para garantir as aprendizagens no tempo certo.

Exemplo de item de resposta construída e sua respectiva chave de correção analítica para avaliação de compreensão de textos contemplando a realização de inferências:

Item 3**Marta e a Galinha Vermelha**

Porque é que a mãe da Marta sente pena da galinha que está no fim da hierarquia?

Processo: *Fazer inferências diretas*

- Resposta válida

Refere que essa galinha é bicada por todas as outras, ou que ela não pode bicar nenhuma das outras galinhas, ou que é a galinha que fica nas piores condições.

Exemplos:

- Todas as outras galinhas a bicam.
- Ele não pode bicar nenhuma das outras galinhas.
- A chefe das galinhas está sempre a mandar nela.
- É ela que fica com os piores restos de comida.

- Resposta não aceitável

Revela incompreensão relativamente à noção de hierarquia, ou dá uma resposta vaga ou sem relação com o texto, ou repete as palavras da pergunta.

Exemplos:

- Pode ficar doente.
- As outras galinhas sentam-se em cima dela.
- Ainda lhe falta muito até ficar ela a liderar.
- As galinhas gostam de ter um chefe autoritário.

Fonte: IAVE, 2016, p. 42.

Esse item foi aplicado no PIRLS, em Portugal, e teve como suporte o texto literário “Marta e a galinha vermelha”⁹. O objetivo dele é avaliar a capacidade de os estudantes realizarem uma inferência direta.¹⁰

9. Novamente, optamos por suprimir o suporte, para evitarmos excesso de informações.

10. Inferência direta é feita a partir de ideias e informações explícitas no texto. Contudo, a relação entre elas não se encontra na base texto – é preciso processá-lo, tanto de forma localizada quanto globalmente.

Importante observar que, por se tratar de um item cujo objetivo é avaliar um processo de compreensão da leitura, não são contemplados para a correção quaisquer aspectos relacionados à ortografia, morfossintaxe ou coesão. Sendo assim, essa chave de correção, chamada no PIRLS de Guia de Codificação, contempla dois aspectos:

- > Resposta aceitável: recebe o código 1 (um) e traz uma descrição do que se considera como tal, além de exemplos de respostas possíveis para a pergunta feita.
- > Resposta não aceitável: recebe o código 0 (zero), trazendo uma definição de por que a resposta do estudante se caracteriza como tal, bem como exemplos de respostas que não podem ser aceitas para o que foi solicitado.

Para finalizar esta seção, vamos conhecer uma chave de correção de forma mais detalhada, que traz os critérios para a correção de um item de produção textual um pouco mais complexo.

Exemplo de chave de correção analítica para correção de produção textual que solicita a redação de uma narrativa:

CHAVE DE CORREÇÃO POR COMPETÊNCIA

NÍVEIS (NOTAS)	COMPETÊNCIAS			
	C1 REGISTRO	C2 TEMA	C3 TIPOLOGIA TEXTUAL	C4 COESÃO
NÍVEL I (2,0 PTS) ABAIXO DO BÁSI- CO	Nesse nível, a escrita do estudante pode apresentar características tanto do nível alfabético quanto do nível silábico. Também se enquadra nesse nível o texto que apresenta ausência de segmentação ou muitos desvios dessa natureza.	O estudante desenvolve de forma tangencial (superficial) o tema.	Os fatos narrados apresentam-se de forma desarticulada, comprometendo a progressão da narrativa.	O estudante produz texto SEM articulação entre suas partes, configurando uma lista de palavras e/ou frases soltas.

CHAVE DE CORREÇÃO POR COMPETÊNCIA

NÍVEIS (NOTAS)	COMPETÊNCIAS			
	C1 REGISTRO	C2 TEMA	C3 TIPOLOGIA TEXTUAL	C4 COESÃO
NÍVEL II (4,0 PTS) BÁSICO	Nesse nível, o texto do estudante apresenta frequentes desvios de registro que prejudicam a inteligibilidade de grande parte de sua produção. O texto foi escrito COM erros de grafia de palavras (escrita alfabética) e com problemas de pontuação, acentuação ou letra maiúscula (falta ou uso inadequado). O texto pode, ainda, apresentar segmentação inadequada em 4 (quatro) ou mais casos, desde que não configure falta de domínio.	O estudante aborda o tema de modo insuficiente, utilizando-se apenas das ideias do texto provocador.	O texto apresenta personagem(ns) executando ação, que pode ser ou não do narrador. Não apresenta 3 (três) dos elementos seguintes ou apresenta-os sem plausibilidade: sequenciação de fatos e/ou de manutenção do tema; lugar em que a narrativa ocorre; momento (tempo) em que a narrativa ocorre; finalização.	O estudante produz texto com articulação precária e/ou inadequada, apresentando graves e frequentes desvios, assim como apenas um tipo de recurso coesivo.
NÍVEL III (6,0 PTS) INTERMEDIÁRIO	Nesse nível, o texto do estudante apresenta eventuais desvios de registro que ainda podem influenciar significativamente a inteligibilidade de partes do texto. O texto foi escrito COM erros de grafia de palavras (escrita alfabética), mas com uso adequado de pontuação e letra maiúscula (quando possível apreender pelo tipo de letra), com possibilidade de erro de acentuação. Pode apresentar, ainda, segmentação inadequada em até 3 (três) casos.	O estudante desenvolve o tema a partir de ideias previsíveis sem, contudo, detalhá-las.	O texto apresenta personagem(ns) executando ação – o autor do texto ou personagem(ns) criada(s); sequência de fatos; MAS ausência ou não plausibilidade de 2 (dois) dos elementos seguintes: lugar em que a narrativa ocorre; momento (tempo) em que a narrativa ocorre; manutenção da narrativa; finalização.	O estudante produz texto com MUITAS INADEQUAÇÕES na utilização dos recursos coesivos, apresentando uso inadequado dos articuladores em 5 (cinco) ou mais ocorrências.

CHAVE DE CORREÇÃO POR COMPETÊNCIA

NÍVEIS (NOTAS)	COMPETÊNCIAS			
	C1 REGISTRO	C2 TEMA	C3 TIPOLOGIA TEXTUAL	C4 COESÃO
NÍVEL IV (8,0 PTS) ADEQUA- DO	Nesse nível, o texto do estudante apresenta alguns desvios que não afetam a inteligibilidade textual. O texto foi escrito SEM erros de grafia de palavras (escrita ortográfica), mas com problemas de segmentação inadequada em até 2 (dois) casos, pontuação, acentuação ou letra maiúscula (falta ou uso inadequado).	O estudante desenvolve o tema apresentando os fatos de forma contínua e detalhada, atendendo à proposta de produção textual.	O texto apresenta personagem(ns) executando ação – o autor do texto ou personagem(ns) criada(s); sequência de fatos; MAS ausência ou não plausibilidade de 1 (um) dos elementos seguintes: lugar em que a narrativa ocorre; momento (tempo) em que a narrativa ocorre; manutenção da narrativa; finalização.	O estudante produz texto com ALGUMAS INADEQUAÇÕES na utilização dos recursos coesivos, contendo eventuais desvios que não comprometem a estrutura lógico-gramatical e o encadeamento do texto, com uso inadequado em ATÉ 4 (quatro) ocorrências.
NÍVEL V (10,0 PTS) AVANÇA- DO	Nesse nível, o texto do estudante foi escrito SEM erros de grafia de palavras (escrita ortográfica), com uso adequado de pontuação, acentuação gráfica, uso de letra maiúscula (quando possível apreender pelo tipo de letra) e com segmentação adequada em TODO o texto. A presença de 1 (um) desvio permite avaliar o texto dentro desse nível, desde que não apresente recorrência.	O estudante desenvolve o tema a partir de uma narração consistente, apresentando os fatos de forma contínua, atendendo à proposta de produção textual de modo pleno, com introdução de elementos que revelam indícios de autoria, ou seja, que ultrapassam os limites do texto provocador.	O texto apresenta personagem(ns) executando ação – o autor do texto ou personagem(ns) criada(s), mas com plausibilidade de todos seus elementos sequência de fatos; lugar em que a narrativa ocorre; momento (tempo) em que a narrativa ocorre; manutenção da narrativa; finalização plausível.	O estudante produz texto SEM INADEQUAÇÕES na utilização dos recursos coesivos. A presença de 1 (um) desvio permite avaliar o texto dentro desse nível, desde que não apresente recorrência.

Fonte: ESPÍRITO SANTO, 2016, p. 29. Adaptado.

Esse exemplo de chave de correção, denominada também de matriz de correção por competência, contempla quatro competências a serem consideradas quando da correção de um texto narrativo:

Registro: avalia o domínio de um conjunto de regras e convenções da língua portuguesa em sua variedade brasileira, englobando aspectos relacionados à sintaxe, à morfologia, à ortografia, à pontuação e à adequação do vocabulário.

Tema: avalia a adequada compreensão da proposta de produção de texto e seu desenvolvimento, verificando se o estudante compreendeu o que foi solicitado.

Tipologia textual: avalia se o texto redigido atende às características estruturais de um texto narrativo.

Coesão: avalia a utilização de elementos coesivos de modo a se manter a lógica da progressão textual.

Importante observar que cada uma dessas competências se subdivide em cinco níveis de acerto, com cada nível descrevendo uma característica da escrita do estudante para as competências avaliadas. Cada nível, por sua vez, recebe uma pontuação que varia de 2 a 10.

Não há a atribuição de zero, pois essa nota diria respeito a algumas situações observadas no texto do estudante antes ainda de sua análise. Por exemplo, se o estudante:

- deixou tudo em branco;
- escreveu menos linhas do que foi especificado nas orientações para realização da redação;
- rasurou, rabiscou ou escreveu algo que se revela com intenção de anular a redação;
- apenas copiou trechos da proposta de redação;
- escreveu um texto que não atende à tipologia textual solicitada, nesse caso, uma narrativa;
- fugiu do tema, ou seja, elaborou um texto que não se relaciona minimamente com o tema proposto para a redação.

Portanto, ao final, caso o estudante tenha produzido um texto que não se enquadre em nenhuma das situações acima, sua redação passará a ser avaliada, recebendo uma nota para cada competência, de modo que o somatório dessas notas corresponderá à sua nota final na produção textual realizada.

3.4 Itens de resposta construída da etapa de alfabetização, de língua portuguesa e de matemática

Apresentamos, até aqui, alguns diferentes formatos de itens de resposta construída, os quais contemplam, considerando a Taxonomia de Bloom revisada (ANDERSON *et al.*, 2001), desde processos mais simples, como recuperar uma informação explícita, até habilidades mais complexas, como analisar e avaliar conteúdo de um texto para se posicionar de modo coerente. Além disso, demos exemplos de itens de resposta construída associados à produção escrita, que podem estar relacionados aos processos iniciais da alfabetização, como a escrita de palavras e frases, bem como à elaboração de uma redação.

Agora, nos aprofundaremos um pouco mais em exemplos de itens utilizados na avaliação educacional em larga escala, para a etapa de alfabetização e para os componentes de língua portuguesa e matemática.

3.4.1 Itens de resposta construída para a etapa de alfabetização e língua portuguesa

Na etapa da alfabetização, os itens de resposta construída avaliam, em geral, a escrita dos estudantes. Por meio dessa avaliação, é possível compreender as hipóteses que estão sendo construídas pelas crianças sobre como ocorre a representação da fala na escrita. No processo de alfabetização, uma etapa elementar é a capacidade de escrever palavras corretamente, cuja avaliação revela-se de suma importância para que seja possível saber se o estudante está evoluindo da forma e no tempo esperados.

Itens de resposta construída que avaliam a escrita de palavras na alfabetização

O ponto de partida para a elaboração dos itens de resposta construída que avaliam a escrita de palavras é a definição dos objetivos da avaliação, que estão relacionados à etapa de escolarização na qual os itens serão aplicados. Com base nesses objetivos, são definidos os critérios para a escolha do tipo de palavra que será apresentada aos estudantes. Esses critérios são linguísticos, relacionados à estrutura dos vocábulos, e têm a ver com aquilo que se deseja verificar em relação ao desenvolvimento da escrita das crianças.

Sendo assim, se o objetivo é compreender, por exemplo, se os estudantes já consolidaram a hipótese alfabética com relação à escrita (ou seja, se já compreendem que é necessária uma letra para representar cada som da palavra) e se já utilizam as letras convencionais para realizar a representação dos fonemas, deverão ser escolhidas palavras compostas por sílabas canônicas, no padrão consoante vogal. Por outro lado, caso o propósito seja compreender se os estudantes não apenas utilizam a hipótese alfabética em sua escrita, mas já dominam também aspectos da escrita ortográfica, as palavras escolhidas devem possuir outro nível de dificuldade para compor o teste. Por exemplo, palavras que apresentem relações irregulares fonema-grafema; ou com marcas de nasalização, como “tampa”, ou “rã”; ou que apresentam sons que podem ser representados por mais de um grafema, a depender de sua posição na palavra, como “rosa”, onde o “s” tem som de “z”; dentre outras possibilidades.

Com base na clareza quanto ao nível de complexidade dos vocábulos em itens de resposta construída que avaliam a escrita de palavras é que será possível descrever os desempenhos dos estudantes, a partir da qualidade de sua escrita. Tal descrição é fundamental para a elaboração de intervenções pedagógicas que focalizem as dificuldades ou potencialidades observadas, de modo que permitam que os estudantes construam hipóteses mais sofisticadas sobre como a escrita funciona.

A segunda etapa da elaboração dos itens de escrita de palavras é a construção da situação-problema a partir da qual a escrita da palavra será solicitada aos estudantes. A situação-problema pode envolver, por exemplo, a escrita de uma palavra ditada pelo aplicador ou a escrita da palavra mediante a apresentação de uma figura. Ambas as situações apresentam limites e possibilidades.

No caso da escrita da palavra ditada por um aplicador, pode haver interferências relacionadas à pronúncia. Portanto, é importante ficar atento a possíveis variações dialetais da fala do aplicador ou à pronúncia artificializada da palavra com o intuito de “facilitar” a identificação pelo estudante. Isso é comum no caso das variações fonéticas envolvendo vogais que, na fala, têm uma realização fonética diferente da sua representação por grafemas na escrita – por exemplo, “sapatu” ao invés de “sapato”. Em casos como esse, pode ocorrer de o aplicador pronunciar artificialmente a vogal final da palavra, tal como se escreve, o que compromete a realização da tarefa pelo estudante, que poderá produzir uma escrita diferente daquela que produziria sem esse apoio.

Quando estamos diante da escrita a partir de uma imagem oferecida como suporte ao item, pode ocorrer de o estudante fazer uma interpretação da imagem que não seja aquela da palavra pretendida. Por exemplo, em um teste da Avaliação Nacional da Alfabetização (ANA), a partir da imagem de um caminhão, muitos alunos escreveram “Scania”¹¹ ao invés da palavra pretendida. Outra questão a ser considerada é a de elementos que recebem diferentes denominações em regiões do país, como o brinquedo “pipa”, que, a depender da região, pode ser chamado de arraia, morcego, lebreque, bebeu, coruja, tapioca, papagaio etc.

Como se vê, a escolha da imagem deve considerar essas questões, que inviabilizam algumas possibilidades. Esse dado indica a necessidade de cuidados relativos à escolha das imagens que servirão de suporte aos itens, para que a avaliação possa alcançar seus objetivos, especialmente quando os itens são aplicados em diferentes contextos ou regiões do país.

A terceira etapa consiste na elaboração da chave de correção, que guiará a análise das respostas produzidas pelos estudantes. No caso da escrita de palavras, os estudos sobre a psicogênese da língua escrita, desenvolvidos por Emília Ferreiro e Ana Teberosky (1999), lançaram luzes ao processo de alfabetização ao concluírem que os estudantes têm um papel ativo, formulando hipóteses sobre como é possível representar os sons da fala por meio da escrita. Tais hipóteses se manifestam no modo como os estudantes escrevem, estando distantes ou próximos da escrita convencional.

Dessa maneira, nos itens de escrita de palavras, para além de um diagnóstico construído com base em noções de certo e errado, é possível produzir um diagnóstico da “qualidade” do erro, identificando o quanto a resposta do estudante ao item se aproxima ou se distancia da escrita convencional. Para isso, é necessária a elaboração de uma chave de correção, que permita alocar o tipo de escrita produzida pelos estudantes que realizam o teste numa determinada categoria.

11. Marca de uma fabricante de veículos.

A seguir apresentamos o exemplo de um item de resposta construída para a escrita de palavra.

Aplicador, ler SOMENTE a(s) informação(es) que apresenta(m) o desenho de um megafone .

Questão

 **Escreva o nome da figura abaixo.**



Disponível em: <<https://bit.ly/2KMJCTL>>. Acesso em: 11 jul. 2018.

Fonte: CAEd/UFJF.

Nesse primeiro exemplo, o estudante deve escrever a palavra solicitada a partir do estímulo de uma imagem. Trata-se da imagem de uma vela, um objeto provavelmente conhecido por todos os estudantes da etapa à qual se destina o teste e que não recebe outras denominações.

A escolha da palavra a ser escrita está relacionada ao objetivo do teste. Nesse caso, o objetivo é avaliar se os estudantes escrevem corretamente palavras pouco extensas, compostas por sílabas no padrão canônico (consoante e vogal), nas quais existe uma correspondência regular entre as letras e os sons que representam – isto é, um fonema é representado por apenas um grafema¹². Vê-se, portanto, que a escolha da palavra atende ao objetivo de avaliar aspectos da alfabetização inicial.

12. No caso do “L”, embora ele possa representar também o som do “U”, sabemos que isso só ocorre em final de sílaba, antecedido por uma vogal. Portanto, quando o “L” ocorre em início de sílaba, representa sempre o mesmo fonema.

O modelo de chave de correção a seguir é válido para todos os itens que avaliam a escrita de palavras a partir de uma imagem:

Quadro 3: Chave de correção para item de escrita de palavras a partir de imagem

ESCRITA DE PALAVRA A PARTIR DE IMAGEM

Nessa questão, o estudante precisa nomear a imagem apresentada. Além do nome esperado (**palavra exata**), também serão aceitas outras nomeações atribuídas à imagem, ou seja, que pertençam ao mesmo **campo semântico**. Dessa forma, a primeira sinalização referente à resposta do estudante será a **relação da imagem** com a escrita produzida por ele:

RELAÇÃO COM A IMAGEM	
PALAVRA EXATA	Sinalize essa opção quando o estudante escrever exatamente a palavra solicitada, em conformidade com a imagem. Aqui, serão considerados, também, os regionalismos/variações dialetais.
CAMPO SEMÂNTICO	Sinalize essa opção quando o estudante escrever uma palavra dentro do campo semântico suscitado pela imagem. Atenção! Também receberá essa indicação a palavra que, no aspecto ortografia , receber o conceito E.

As imagens apresentadas no teste, suas nomeações exatas e as que se enquadram como campo semântico são:

Uma vez registrada a **situação de correção** e a **relação com a imagem**, proceda à avaliação da **ortografia** do estudante, enquadrando-a em um dos 5 (cinco) níveis a seguir:

ORTOGRAFIA	
A	O estudante escreveu a palavra corretamente, configurando uma escrita em nível ortográfico . Atenção! Caso a palavra escrita não coincida com a nomeação da imagem, somente será válida a resposta que possuir estreita relação com a figura apresentada, ou seja, que pertença ao mesmo campo semântico, como previsto neste Guia de Correção.
B	O estudante escreveu a palavra com desvios, configurando uma escrita em nível alfabético . Exemplos: sapu (Sapo); Elefanti (Elefante); Istrela (Estrela).
C	O estudante escreveu a palavra com desvios, configurando uma escrita em nível silábico-alfabético . Exemplos: Apo (Sapo); Efate (Elefante); Etrela (Estrela).
D	O estudante escreveu a palavra com desvios, configurando uma escrita em nível silábico . Nesse caso, a escrita apresentada compromete a leitura, não sendo possível sua compreensão sem conhecer a imagem. Contudo, a escrita produzida pelo estudante apresenta relação com a pauta sonora das letras que compõem a palavra. Exemplos: sp/ao (Sapo); Elt/Eft (Elefante); Eta/Et/ Etrl (Estrela).
E	O estudante produziu escrita em nível pré-silábico (apresentou letras aleatórias para representar a palavra, desenhos/garatujas) E/OU grafou outra palavra que não guarda estreita relação com a imagem (provavelmente reproduzida de memória).

Observações:

- (1) No caso de repetição de palavras, considerar a que estiver grafada corretamente ou com menos desvios.
- (2) Deve-se considerar a escrita da palavra fora do espaço destinado para sua escrita, mesmo que essa esteja no campo destinado à resposta de outra questão.
- (3) Letra espelhada será contada como um tipo de desvio.
- (4) Diminutivos, aumentativos e plurais, assim como variações dialetais e de gênero, serão aceitos, desde que estejam dicionarizados. Em caso de ocorrência, consultar o dicionário *Caldas Aulete*: <http://www.aulete.com.br/>.
- (5) Não importa o padrão de letra utilizado pelo aluno (cursiva, forma ou mista).
- (6) Acentuação gráfica NÃO será considerada para fins de avaliação nessa questão.
- (7) A segmentação da escrita NÃO SERÁ AVALIADA nessa questão, desde que o aluno não tenha apresentado APENAS letras dispersas ao longo da página na tentativa de compor a palavra. Caso ocorra, deve ser analisado se a ocorrência configura uma escrita pré-silábica.

A chave de correção é um recurso que serve para que as produções dos estudantes sejam avaliadas com base nos mesmos critérios. Desse modo, são dadas orientações aos corretores sobre como analisar as produções e como lidar com as situações que podem ser observadas – no exemplo apresentado, tais orientações se encontram no campo “Observações”.

No caso da escrita de palavras, a chave contempla as hipóteses formuladas pelos estudantes sobre como a escrita se organiza, em conformidade com os estudos da psicogênese da língua escrita de Ferreiro e Teberosky (1999).

Por meio da chave de correção, o item de resposta construída é transformado em um item de múltipla escolha, no qual a produção apresentada pelo estudante será enquadrada, pelo corretor, numa determinada alternativa (no caso dessa chave, A, B, C, D ou E) a depender de suas características.

No exemplo apresentado, também é contemplada a situação em que a escrita produzida pelo estudante, com base na imagem, não a nomeia como foi pensado originalmente no momento da elaboração do item.

Itens de resposta construída que avaliam a produção textual

Pesquisas recentes têm demonstrado que o processo de alfabetização precisa acontecer em concomitância com o que a literatura sobre o tema denomina “letramento”, que se refere aos usos sociais que as pessoas fazem dos conhecimentos sobre a língua em situações reais de escrita. A BNCC afirma que, no ensino fundamental,

[a]s diversas práticas letradas em que o aluno já se inseriu na sua vida social mais ampla, assim como na Educação Infantil, tais como cantar cantigas e recitar parlendas e quadrinhas, ouvir e recontar contos, seguir regras de jogos e receitas, jogar games, relatar experiências e experimentos, serão progressivamente intensificadas e complexificadas, na direção de gêneros secundários com textos mais complexos. (BRASIL, 2018)

Visando ao alcance dessa meta, no eixo de produção de textos, espera-se que os estudantes incorporem estratégias referentes a diferentes gêneros, de modo a ampliar suas práticas de letramento.

Assim como no caso dos itens que avaliam a escrita de palavras, a elaboração de itens que avaliam a produção textual envolve algumas etapas, que devem ser levadas em conta tanto nos anos escolares do ciclo inicial da alfabetização quanto nos anos seguintes, no componente de língua portuguesa.

Novamente, o ponto de partida é a definição dos objetivos da avaliação, que, tal qual nos itens de escrita de palavras, estão relacionados à etapa de escolarização em que os itens serão aplicados. Com base nesses objetivos, será definido o gênero textual que os estudantes deverão produzir. A BNCC orienta quanto aos gêneros que devem ser privilegiados, na produção textual, em cada etapa de escolarização. Também em função dos objetivos da avaliação, serão definidos os aspectos ou dimensões da produção textual que serão objeto da avaliação.

Há vários aspectos envolvidos na produção de um texto: a organização das ideias, os elementos de coesão – que permitem essa organização –, a adequação do texto a um determinado gênero, a ortografia, a sintaxe, a pontuação, dentre outras possibilidades. Por exemplo, pode acontecer que, diante da produção de um bilhete, o estudante saiba organizar o texto para atender ao propósito comunicativo de enviar um recado, mas apresente dificuldades na escrita de algumas palavras; ou, na elaboração de um relato, consiga produzir uma escrita relativamente convencional, mas tenha dificuldade no uso adequado de elementos coesão, repetindo termos em vez de substituí-los por outros.

Com base no exposto, é possível concluir que, assim como no caso da avaliação da escrita de palavras, os itens que avaliam a produção textual permitem identificar o quanto a escrita dos estudantes se aproxima ou se distancia do padrão esperado para a etapa de escolarização em que se encontram. E, ainda, que tipo de intervenção pedagógica deve ser realizada para que as dificuldades sejam sanadas.

É importante destacar que aspectos como a criatividade ou o uso da imaginação não são objeto de avaliação em itens de resposta construída que avaliam a produção textual utilizados em avaliações de larga escala. Isso porque esses critérios são bastante subjetivos, não sendo possível abordá-los, de forma equânime, numa chave de correção. Entretanto, esses podem ser critérios adotados pelo professor ou pela professora em sala de aula. Porém, mesmo nesses casos, é necessário que sejam estabelecidos critérios claros para a correção.

A segunda etapa para a elaboração de itens de produção textual consiste na construção da situação-problema a partir da qual os estudantes deverão produzir o texto. Trata-se da criação de um contexto significativo, que oriente os estudantes para que tenham uma compreensão sobre o que devem escrever. É importante que esse contexto dialogue com as experiências de vida presumidas dos estudantes da etapa avaliada, de modo que possam mobilizar essa experiência para produzir o texto.

Por fim, a terceira etapa prevê a elaboração da chave de correção, que guiará a análise das respostas. Assim como na avaliação da escrita de palavras, os itens que avaliam a produção de textos também permitem um diagnóstico da “qualidade” do erro, identificando o quanto a resposta do estudante ao item, em cada dimensão da produção textual avaliada, se aproxima ou se distancia da escrita convencional. Para isso, é necessária a elaboração de uma chave de correção que permita alocar a escrita produzida pelos estudantes em uma determinada categoria para cada dimensão avaliada.

Sendo assim, pode ocorrer de o texto produzido apresentar uma escrita ortograficamente adequada, mas não corresponder ao gênero solicitado. O contrário também pode ocorrer, isto é, um texto relativamente adequado ao gênero solicitado, mas com muitos problemas relativos à escrita das palavras.

A seguir, apresentamos um item de resposta construída que avalia a produção textual.

Aplicador, ler SOMENTE a(s) informação(ões) que apresenta(m) o desenho de um megafone

Questão 22

E030118G5

Imagine que você descobriu uma passagem secreta no final da sua rua. Escreva uma história contando como foi essa aventura. Não se esqueça de contar as coisas que você viu e fez.

Fonte: CAEd/UFJF.

Nesse exemplo, o gênero textual a partir do qual os estudantes deverão produzir uma narrativa pode ser um relato de experiência ou ficcional. Para isso, é construído um contexto no qual essa produção adquire sentido, solicitando-se ao estudante que imagine uma determinada situação. A orientação para que o aluno conte “as coisas que viu e fez” direciona a produção textual para o gênero relato de experiência, a ser construído em primeira pessoa. A dimensão ficcional é suscitada pelo uso da forma verbal “Imagine”, a qual introduz o contexto para a produção textual.

Para a correção dessa produção textual, utiliza-se a chave de correção apresentada a seguir.

Quadro 4: Chave de correção para item de produção textual a partir de sequência de imagens ou situação motivadora

PRODUÇÃO DE TEXTO: NARRATIVA A PARTIR DE SEQUÊNCIA DE IMAGENS (TIRINHA) OU SITUAÇÃO MOTIVADORA

A última questão apresentada no teste solicita ao estudante a produção textual baseada na situação motivadora pré-determinada. Nessa questão, a escrita do estudante será avaliada em 7 (sete) aspectos distintos, sendo que o conceito atribuído a cada um deles será independente, ou seja, não existe correlação que exija a atribuição de um mesmo conceito a todos, pois cada aspecto observa a escrita do estudante sob uma perspectiva específica. As propostas foram:

Inicie a avaliação da escrita do estudante sinalizando a **situação de correção** na qual se encaixa sua resposta, sendo que, para essa questão, há 4 (quatro) situações previstas:

SITUAÇÃO DE CORREÇÃO	
NORMAL	Sinalize essa situação de correção se a resposta do estudante estiver passível de ser corrigida, de acordo com os critérios deste guia.
EM BRANCO	Sinalize essa situação de correção se a questão estiver em branco, ou seja, se não houver nenhuma resposta.
PRÉ-SILÁBICO	Sinalize essa situação de correção se a escrita apresentada pelo estudante encontra-se em nível pré-silábico.

Uma vez indicada a **situação de correção**, proceda à avaliação do estudante nos **aspectos linguístico-textuais** descritos a seguir:

ASPECTO 1 – ADEQUAÇÃO À PROPOSTA	
Esse aspecto avalia a coerência textual, ou seja, se o estudante comprehende a proposta de produção textual, desenvolvendo o tema. A narrativa (ou relato) deve referir-se ao suporte motivador, sendo possível o aluno elaborar diversos contextos esse, desde que plausíveis.	
A	O estudante escreveu uma história coerente em relação à situação motivadora.
B	O estudante escreveu uma história com pouca plausibilidade em relação à situação motivadora. Isto é, o estudante desenvolve o tema de forma tangencial (faz uma abordagem parcial/superficial do tema).
C	O estudante escreveu um texto que não é coerente com a sequência de imagens/situação motivadora, ou seja, fuga ao tema (aborda algo completamente diferente do proposto) ¹ .
ASPECTO 2 – TIPOLOGIA TEXTUAL	
Nesse aspecto, observa-se o domínio da estrutura textual narrativa, ou seja, a tipologia textual a partir da qual o aluno deve desenvolver o tema proposto. A tipologia textual refere-se à forma que o texto pode apresentar, visando responder à intenção comunicativa prevista. É possível o estudante apresentar sequências de outras tipologias também, desde que sejam empregadas como recurso narrativo.	
A	O estudante escreveu um texto que apresenta todos os elementos constitutivos de uma narrativa: narrador, lugar/espaco, tempo, personagens praticando ação/ações (o aluno pode adotar um narrador-personagem) e enredo (situação inicial, situação central e desfecho).
B	O estudante escreveu um texto com ausência de 1 (um) ou mais elementos constitutivos de uma narrativa: narrador, lugar/espaco, tempo, personagens praticando ação/ações (o estudante pode adotar um narrador-personagem) e enredo (situação inicial, situação central e desfecho).
C	O estudante apenas enumerou (descreveu) aspectos relativos a uma personagem e/ou fato(s).
D	O estudante escreveu um texto que não se configura como uma narrativa , ou seja, não atendeu à tipologia textual, elaborando outra estrutura textual. Exemplos: bilhete, receita, poema, lista etc.
ASPECTO 3 – USO DA PÁGINA ²	
Esse aspecto avalia o domínio do estudante quanto ao uso adequado do espaço delimitado para sua escrita, seguindo as regras de uso da página, ou seja, respeitando as direções da escrita (de cima para baixo e da esquerda para a direita), as margens direita e esquerda, linha e sequência da escrita – mudança de linha.	
A	O estudante escreveu o texto usando o espaço delimitado para sua escrita de forma adequada , seguindo todas as regras de uso da página, ou seja, respeitando: as direções da escrita (de cima para baixo e da esquerda para a direita), as margens direita e esquerda, linha e sequência da escrita – mudança de linha.
B	O estudante escreveu o texto sem seguir, pelo menos, uma das regras de uso da página.
C	O estudante escreveu o texto sem seguir as regras de uso da página.
ASPECTO 4 – ORTOGRAFIA	
Esse aspecto avalia o domínio dos mecanismos linguísticos próprios da variedade brasileira da Língua Portuguesa por parte do estudante: ORTOGRAFIA. Observa-se, também, a estruturação morfossintática do texto.	
A	O estudante escreveu o texto sem desvios, configurando uma escrita em nível ortográfico .
	Atenção! A presença de até 3 (três) desvios permitirá avaliar o texto dentro desse nível, desde que não apresente recorrência.
B	O estudante escreveu o texto com desvios, configurando uma escrita em nível alfabético .
C	O estudante escreveu o texto com desvios, configurando uma escrita em nível silábico-alfabético , sendo possível compreender a escrita sem maior cooperação.
D	O estudante escreveu o texto com desvios, configurando uma escrita em nível silábico , pois apresenta relação com a pauta sonora das letras que compõem a palavra.

ASPECTO 5 – PONTUAÇÃO	
Esse aspecto avalia o domínio dos sinais de pontuação, recurso linguístico empregado para atribuir expressividade ao que foi escrito (entonação e pausas, por exemplo). Nessa etapa de escolaridade, será avaliado o emprego dos seguintes sinais de pontuação: ponto final, exclamação, interrogação e travessão.	
A	O estudante pontuou adequadamente nas vezes em que utilizou os sinais de pontuação em seu texto. Atenção! A presença de 1 (um) desvio permitirá avaliar o texto dentro desse nível, desde que não apresente recorrência e o estudante tenha empregado 2 (dois) ou sinais de pontuação em seu texto.
B	O estudante pontuou inadequadamente na maioria das vezes em que utilizou os sinais de pontuação em seu texto.
C	O estudante utilizou apenas 1 (um) ponto para concluir o texto.
D	O estudante não fez uso de qualquer sinal de pontuação.
ASPECTO 6 – COESÃO	
Esse aspecto avalia a coesão textual, que é o domínio dos mecanismos linguísticos necessários para a construção da narrativa, ou seja, a conexão, ligação e harmonia entre os elementos linguísticos de um texto (a transição de ideias entre as frases e os parágrafos). Nessa etapa de escolaridade, não se espera o domínio de estruturas complexas previstas para domínio em etapas subsequentes, mas o conhecimento do estudante sobre a necessidade de interligar as partes do seu texto de forma linear, do ponto de vista linguístico.	
A	O estudante produziu texto articulando as ideias por meio do uso de conexões, substituições e/ou marcas linguísticas. Atenção! A presença de até 1 (um) desvio permitirá avaliar o texto dentro desse nível, desde que não apresente recorrência e tenha feito uso de(s) elemento(s) coesivo(s) mais de 1 (uma) vez.
B	O estudante produziu texto articulando as ideias com algumas falhas no uso dos conectivos.
C	O estudante produziu texto com muitas falhas no uso dos conectivos, em relação ao texto como um todo, comprometendo a compreensão de sua escrita.
D	O estudante produziu texto sem articulação entre suas partes, configurando uma de lista de palavras e/ou frases soltas.
ASPECTO 7 – SEGMENTAÇÃO	
Esse aspecto avalia a compreensão do estudante sobre o reconhecimento do alinhamento da escrita, bem como a unidade palavra em frases. Para essa correção, o parâmetro avaliativo serão os desvios de hiposegmentação e hipersegmentação.	
A	O estudante segmentou adequadamente todas as palavras do texto.
B	O estudante apresentou poucos desvios de segmentação em relação ao texto como um todo.
C	O estudante apresentou desvios de segmentação ao longo de todo o texto.

Observações:

- (1) No Aspecto 4 – Ortografia, observe que:
 - no caso de repetição de palavras, considerar a que estiver grafada corretamente ou com menos desvios;
 - será aceito o texto (ou fragmento do texto) que for escrito fora do espaço delimitado para sua composição;
 - diminutivos, aumentativos e plurais, assim como variações de gênero, serão aceitos, desde que estejam dicionarizados (em caso de ocorrência, consultar o dicionário *Caldas Aulete*: <http://www.aulete.com.br/>);
 - letra espelhada será contada como um tipo de desvio;
 - não importa o padrão de letra utilizado pelo aluno (cursiva, forma ou mista);
 - acentuação gráfica e o emprego/ausência de letra maiúscula NÃO serão considerados para fins de avaliação nessa questão;
 - concordância verbal NÃO será observada para fins de avaliação.
Segmentação será avaliada apenas no Aspecto 7.
- (2) Para a correção do Aspecto 7 – Segmentação, observe se o texto do estudante apresenta desvios de hiposegmentação e hipersegmentação, assim definidos:
 - **hiposegmentação** – ocorre quando o alfabetizando junta uma palavra na outra em uma frase ou texto;
 - **hipersegmentação** – ocorre quando o alfabetizando separa letra ou sílaba de uma mesma palavra.

Como é possível observar nessa chave de correção, o item de resposta construída de produção textual desse segundo exemplo desdobra-se em sete outros itens, uma vez que são sete os aspectos ou dimensões a serem avaliados. Para cada um dos aspectos, há possibilidades de caminhos a serem seguidos – alternativas A, B, C... –, cabendo ao corretor assinalar aquela que corresponde ou que melhor se alinha ao texto produzido pelo estudante. Observa-se, portanto, que os itens de resposta construída que avaliam a produção textual podem oferecer um diagnóstico bastante detalhado, contemplando diferentes aspectos da escrita dos estudantes, de modo que é possível compreender quais já alcançaram um desenvolvimento satisfatório e quais precisam, ainda, receber maior apoio em sala de aula.

Vamos, agora, analisar um item de resposta construída de língua portuguesa aplicado logo após o ciclo inicial da alfabetização. Como você já sabe, e conforme se apresenta em documentos basilares do PIRLS, o 4º ano do ensino fundamental é uma etapa escolar de transição importante no processo de desenvolvimento da leitura, pois, nesse ponto, “os estudantes aprenderam a ler e agora estão lendo para aprender” (MULLIS; MARTIN, 2019, p. 1, tradução nossa). Em face dessa colocação, as crianças

também já devem ser capazes de fazer uso da escrita para produzir textos de tipologias e gêneros textuais compatíveis com a etapa de escolaridade e com os campos de atuação estabelecidos pela BNCC. Isso pressupõe, portanto, que os estudantes, a partir do 4º ano do ensino fundamental, já sejam capazes, além de ler e compreender o que leem, de registrar, por meio da escrita, as tarefas propostas por itens de resposta construída.

O exemplo a seguir, voltado ao 5º ano do ensino fundamental, mostra um item que propõe a escrita de uma redação.

PRODUÇÃO DE TEXTO

(E050001H6) Leia o texto abaixo e inspire-se na imagem para realizar a produção textual solicitada.



“Um dia, ao anoitecer, peguei meu cachorro e fomos visitar uma velha casa abandonada que ficava no fim da minha rua. Assim que entramos na casa, um vento forte soprou e fechou a porta, deixando a gente preso dentro dela.”

Imagen disponível em: <https://br.freepik.com/vetores-premium/casa-assombrada-na-floresta-a-noite_1265177.htm>. Acesso em: 4 jul. 2018.

Continue esse texto criando uma história a partir das informações apresentadas, imaginando que você é o narrador, ou seja, a pessoa que está contando essa história. Não se esqueça de contar as coisas que você viu e fez durante essa aventura.

Fonte: CAEd/UFJF.

Essa atividade avaliativa enquadra-se no tipo de resposta construída subjetiva, pois, a partir da compreensão do que está sendo proposto para a produção textual, o estudante deve dar continuidade, por escrito, ao parágrafo trazido como texto motivador, o que será realizado conforme a subjetividade de cada um, de acordo com o seu repertório linguístico e o seu domínio da tipologia narrativa.

Esse tipo de proposta de produção textual – “continue a história” – é bastante recorrente em atividades de escrita no contexto escolar. Além do texto motivador, também é apresentada uma imagem que visa contribuir para a identificação do cenário, o ambiente em que a história vai se desenrolar, criando um clima de suspense e estimulando o estudante na realização da atividade.

Neste outro exemplo, temos um item de resposta construída voltado ao final da segunda etapa do ensino fundamental e início do ensino médio.

(R100003G5) Leia os textos abaixo e, em seguida, redija um texto de acordo com a proposta solicitada.

Texto 1
Responsabilidade ambiental
Responsabilidade ambiental é um conjunto de atitudes, individuais ou coletivas, voltadas para o desenvolvimento sustentável. O desenvolvimento sustentável consiste em atender às necessidades do presente sem comprometer a capacidade das gerações futuras em atender às suas próprias necessidades. Praticamos a responsabilidade ambiental sempre que utilizarmos de forma racional os recursos colocados a nossa disposição, sempre que evitarmos o desperdício, sempre que cuidarmos da destinação correta do lixo que geramos, e sempre que procurarmos reutilizar os materiais evitando que estes sejam descartados desnecessariamente, buscando reduzir ou evitar possíveis riscos e danos à natureza. [...]
Disponível em: < http://www.gruposeculo.com.br/crescer/?p=736 >. Acesso em: 1 fev. 2016. Fragmento.
Texto 2
Entenda a tragédia de Mariana, em Minas Gerais
[...] Duas barragens da mineradora Samarco se romperam na cidade de Mariana (MG) [...]. Nessas barragens havia lama, rejeitos sólidos e água. Esses detritos são resultado da mineração na região. [...] Pelo menos 128 residências foram atingidas pela onda de lama e dejetos. [...] Os prejuizos são calculados em mais de R\$ 100 milhões [...]. Especialistas dizem que pode levar séculos para o ambiente se recuperar. A lama que se espalhou por Minas Gerais e Espírito Santo impede que matéria orgânica cresça. Uma das consequências que a lama está causando é o assoreamento, o acúmulo de sedimentos na calha do rio, causando impactos socioeconômicos e ambientais. Segundo o Ibama, houve alterações nos padrões de qualidade da água. Um dos impactos é a morte de animais, terrestres e aquáticos, por asfixia. [...]
BOGNAR, Ariadne. Disponível em: < http://meexplica.com/2015/11/entenda-a-tragedia-de-mariana-em-minas-gerais/ >. Acesso em: 1 fev. 2016. Fragmento.

Imagen disponível em: <<http://migre.me/u6Wqt>>. Acesso em: 15 jun. 2016.

Com base na leitura desses textos e nos conhecimentos construídos ao longo de sua formação, escreva um texto dissertativo-argumentativo, na modalidade culta da Língua Portuguesa, sobre o tema **O PAPEL INDIVIDUAL E O COLETIVO NO CAMPO DA RESPONSABILIDADE AMBIENTAL**, discutindo sobre as responsabilidades dos diversos setores da sociedade na área ambiental, apresentando proposta de intervenção sobre o tema a qual respeite os direitos humanos. Organize o seu texto de forma coesa e coerente, com argumentos e fatos para a defesa de seu ponto de vista.

Fonte: CAEd/UFJF.

Nessa atividade, solicita-se a elaboração de texto dissertativo-argumentativo com base em um tema fundamentado em dois textos motivadores, pertencentes a gêneros textuais distintos, mas com abordagens complementares. Os textos utilizados como suporte devem ser apenas subsídio para a discussão a ser realizada na redação, não podendo ser reproduzidos ou copiados.

Essa tarefa, assim como a do exemplo anterior, se situa no contexto de itens de resposta construída com resposta subjetiva, pois, para a sua realização, cada estudante coloca em ação seus conhecimentos textuais e linguísticos, bem como as suas experiências de leitura que possam vir a incrementar a resposta. Para isso, pressupõe-se a capacidade de cada estudante ler e compreender o tema, relacionando os textos e extraíndo-lhes os aspectos em comum e aquilo que os diferencia, permitindo construir a sua argumentação.

Como podemos observar, a proposta solicita que o estudante emita sua opinião sobre uma temática controversa, como demandam os textos da ordem do argumentar, e que suscita diferentes posicionamentos, conforme o ponto de vista adotado. Por sua vez, a chave de correção apresenta os critérios para que o avaliador possa analisar a resposta do estudante nesse item, de acordo com o que se vê no quadro a seguir.

Quadro 5: Chave de correção para item de produção de texto argumentativo

CHAVE DE CORREÇÃO POR COMPETÊNCIA					
NÍVEIS (NOTAS)	COMPETÊNCIAS				
	C1 REGISTRO	C2 TEMA / TIPOLOGIA TEXTUAL	C3 COERÊNCIA	C4 COESÃO	C5 PROPOSTA DE INTERVENÇÃO
NÍVEL 0 (0,0 PT) INADEQUADO	Demonstra desconhecimento da modalidade escrita formal da língua portuguesa.	-----	Apresenta informações, fatos e opiniões não relacionados ao tema (ligado a assuntos suscitados pela temática) e/ou sem defesa de um ponto de vista.	Não articula as informações.	O estudante não apresenta proposta de intervenção.
NÍVEL I (2,0 PTS) ABAIXO DO BÁSICO	Demonstra domínio precário da modalidade escrita formal da língua portuguesa, de forma sistemática, com diversificados e frequentes desvios gramaticais, de escolha de registro e convenções da escrita.	Apresenta o assunto tangenciando o tema ou com domínio precário do texto dissertativo-argumentativo, com traços constantes de outros tipos textuais.	Apresenta informações, fatos e opiniões pouco relacionados ao tema ou incoerentes e sem defesa de um ponto de vista.	Articula as partes do texto de forma precária.	O estudante elabora proposta de intervenção tangencial ao tema ou a deixa subentendida no desenvolvimento da argumentação.
NÍVEL II (4,0 PTS) BÁSICO	Demonstra domínio insuficiente da modalidade escrita formal da língua portuguesa, com muitos desvios gramaticais, de escolha de registro e convenções da escrita.	Desenvolve o tema recorrendo à cópia de trechos dos textos motivadores ou apresenta domínio insuficiente do texto dissertativo-argumentativo, não atendendo à estrutura prototípica com proposição, argumentação e conclusão.	Apresenta informações, fatos e opiniões relacionados ao tema, mas desorganizados, contraditórios e/ou limitados aos argumentos dos textos motivadores, em defesa de um ponto de vista.	Articula as partes do texto, de forma insuficiente, com muitas inadequações, e apresenta repertório limitado de recursos coesivos.	O estudante elabora, de forma precária, uma proposta de intervenção relacionada ao tema, que não se articula com a discussão desenvolvida no texto ou com desenvolvimento insuficiente dos meios para sua realização.

CHAVE DE CORREÇÃO POR COMPETÊNCIA					
NÍVEIS (NOTAS)	COMPETÊNCIAS				
	C1 REGISTRO	C2 TEMA / TIPOLOGIA TEXTUAL	C3 COERÊNCIA	C4 COESÃO	C5 PROPOSTA DE INTERVENÇÃO
NÍVEL III (6,0 PTS) INTERMEDIÁRIO	Demonstra domínio mediano da modalidade escrita formal da língua portuguesa e de escolha de registro, com alguns desvios gramaticais e de convenções da escrita.	Desenvolve o tema por meio de argumentação previsível e com domínio mediano do texto dissertativo-argumentativo, apresentando proposição, argumentação e conclusão.	Apresenta informações, fatos e opiniões relacionados ao tema, limitados aos argumentos dos textos motivadores e/ou pouco organizados em defesa de um ponto de vista.	Articula as partes do texto de forma mediana, com inadequações e repertório pouco diversificado de recursos coesivos.	O estudante elabora, de forma limitada, proposta de intervenção relacionada ao tema, mas pouco articulada à discussão desenvolvida no texto.
NÍVEL IV (8,0 PTS) ADEQUADO	Demonstra bom domínio da modalidade escrita formal da língua portuguesa e de escolha de registro, com poucos desvios gramaticais e de convenções da escrita.	Desenvolve o tema por meio de argumentação consistente e apresenta bom domínio do texto dissertativo-argumentativo, com proposição, argumentação e conclusão.	Apresenta informações, fatos e opiniões relacionados ao tema, de forma organizada, com indícios de autoria em defesa de um ponto de vista.	Articula as partes do texto com poucas inadequações, apresentando repertório diversificado de recursos coesivos.	O estudante elabora proposta de intervenção clara, relacionada à tese e bem articulada com a discussão desenvolvida no texto, explicitando os meios para realizá-la, mas sem marca de ineditismo.
NÍVEL V (10,0 PTS) AVANÇADO	Demonstra excelente domínio da modalidade escrita formal da língua portuguesa e de escolha de registro. Desvios gramaticais ou de convenções da escrita são aceitos somente em caráter excepcional quando não apresentam reincidência.	Desenvolve o tema por meio de argumentação consistente, a partir de um repertório sociocultural produtivo, e apresenta excelente domínio do texto dissertativo-argumentativo.	Apresenta informações, fatos e opiniões relacionados ao tema proposto, de forma consistente e organizada, configurando autoria em defesa de um ponto de vista.	Articula bem as partes do texto e apresenta repertório diversificado de recursos coesivos.	O estudante elabora proposta de intervenção clara e inovadora, relacionada à tese e bem articulada com a discussão desenvolvida no texto, explicitando os meios para realizá-la.

Fonte: CAEd/UFJF.

Para cada uma dessas competências, há cinco níveis de desenvolvimento, do inadequado ao avançado, com suas respectivas notas. Sendo assim, a pontuação do estudante nesse item terá como base o nível de desempenho em cada uma das cinco competências avaliadas. Mas, para além da nota final, o nível alcançado em cada uma das competências fornece informações importantes sobre o estágio de desenvolvimento do estudante em competências fundamentais relacionadas à produção textual, permitindo que a professora ou o professor adote práticas em sala de aula com foco nas principais dificuldades de cada um.

Em síntese, itens de resposta construída que avaliam produção textual devem oferecer condições para identificar a adequação da produção dos estudantes ao gênero proposto, no que concerne a aspectos de forma e conteúdo. Na etapa da alfabetização, as propostas devem privilegiar os gêneros que circulam na vida cotidiana, enquanto, à medida que o estudante avança no seu percurso escolar ao longo do ensino fundamental e médio, devem entrar em cena textos mais complexos, como narrativas mais sofisticadas e textos argumentativos que tratem de temas mais desafiadores e que exijam a capacidade de o estudante reunir fatos para embasar sua tese e lidar com diferentes visões de mundo.

No que diz respeito à correção, a análise da qualidade da produção textual deve estar, nos anos iniciais, relacionada a aspectos que permitam que os textos sejam coesos e coerentes e, portanto, compreensíveis. Ao avançarmos nas demais etapas escolares, novos critérios devem ser considerados, como o domínio do gênero e da tipologia textual e de suas estratégias. Conclui-se, então, que itens de resposta construída que avaliam a produção textual contemplam uma ampla gama de aspectos, que devem ser devidamente discriminados na chave para a correção desses itens.

Sendo assim, a elaboração de itens de resposta construída em língua portuguesa requer do elaborador: (i) conhecimentos sobre a língua, de modo que sejam produzidos instrumentos consistentes, que permitam avaliar o que realmente se pretende; (ii) conhecimentos sobre como os estudantes se apropriam da língua escrita na etapa de ensino correspondente, de modo a estabelecer critérios claros de correção; (iii) conhecimentos sobre os processos de ensino, de modo a saber que tipo de intervenção sugerir diante dos resultados da avaliação.

3.4.2 Itens de resposta construída de matemática

Em matemática, itens de resposta construída permitem observar habilidades como a capacidade de fazer construções geométricas com régua e compasso, construir argumentações e elaborar propostas de intervenção na realidade por meio de objetos de conhecimento próprios do componente curricular. Além disso, mesmo quando se trata de habilidades que são passíveis de observação por meio de itens de resposta selecionada, os itens de resposta construída criam condições objetivas para que seja traçado um diagnóstico mais preciso da aprendizagem, revelando a natureza do erro cometido pelo estudante na realização da tarefa proposta, o que permite uma intervenção mais focada e personalizada.

Apresentamos, a seguir, um exemplo de um item de resposta construída de matemática, destinado a estudantes do 5º ano do ensino fundamental, que avalia a capacidade de “efetuar divisão entre números naturais, utilizando algoritmos, com até dois algarismos no divisor”.

Efetue a divisão de 370 por 25 e registre o raciocínio empregado.

Gabarito comentado: Espera-se que os estudantes consigam efetuar a divisão, utilizando algum algoritmo, conforme o algoritmo clássico, exemplificado a seguir.

$$\begin{array}{r}
 37'5 \quad | \quad 25 \\
 - 25 \downarrow \\
 \hline
 125 \\
 - 125 \\
 \hline
 0
 \end{array}$$

Outra possibilidade seria o estudante responder 15, acertando o item, e apresentar a seguinte justificativa:

$$300 + 50 + 25 = 375$$

Essa justificativa sugere que o estudante comprehende o significado da divisão e consegue implementar uma estratégia coerente de cálculo. No entanto, fica evidente que ele não faz associação entre a multiplicação e a divisão, razão pela qual não emprega a multiplicação para acelerar suas estimativas parciais. O resultado é uma estratégia pouco eficiente, pois se o quociente for grande, por exemplo, 93, esse procedimento se revelará pouco viável, sendo o estudante obrigado a listar 93 parcelas.

Uma terceira possibilidade seria ele responder 15 e justificar da seguinte forma:

$$\begin{aligned}
 10 \times 25 &= 250 \\
 4 \times 25 &= 100 \\
 \underline{1} \times 25 &= \underline{25} \\
 15 \times 25 &= 375
 \end{aligned}$$

Essa resposta sugere que o estudante comprehende o significado da divisão, e que a considera como a operação inversa da multiplicação, ao implementar uma estratégia de cálculo coerente com esse significado.

Todas essas justificativas podem ser consideradas corretas, sendo que há procedimentos que são mais eficientes que outros. O importante é que o estudante disponha de várias estratégias distintas, mas é igualmente relevante que ele saiba reconhecer qual é a mais adequada, o que pode ser feito por meio da execução de outros exemplos, resolvidos por estratégias diferentes, para que, ao final, se compare a eficiência dessas estratégias.

O próximo exemplo, adequado a estudantes do 1º ano do ensino médio, é um item adaptado do PISA¹³. Como se pode verificar, na chave de correção, “crédito integral” deve ser compreendido como valor total ou conceito máximo, conforme for o caso. “Crédito parcial” deve ser compreendido como pontuação parcial (parte do valor máximo da questão) ou conceito intermediário. “Nenhum crédito” deve ser entendido como sem pontuação ou pontuação igual a zero, ou conceito mínimo.

CONCENTRAÇÃO DE MEDICAMENTOS

Uma mulher hospitalizada recebe uma injeção de penicilina. A penicilina é decomposta contínua e progressivamente, de modo que, uma hora após a injeção, somente 60% da penicilina estará ativa. Esse padrão se repete: ao final de cada hora, somente 60% da penicilina que estava presente no final da hora anterior permanece ativa.

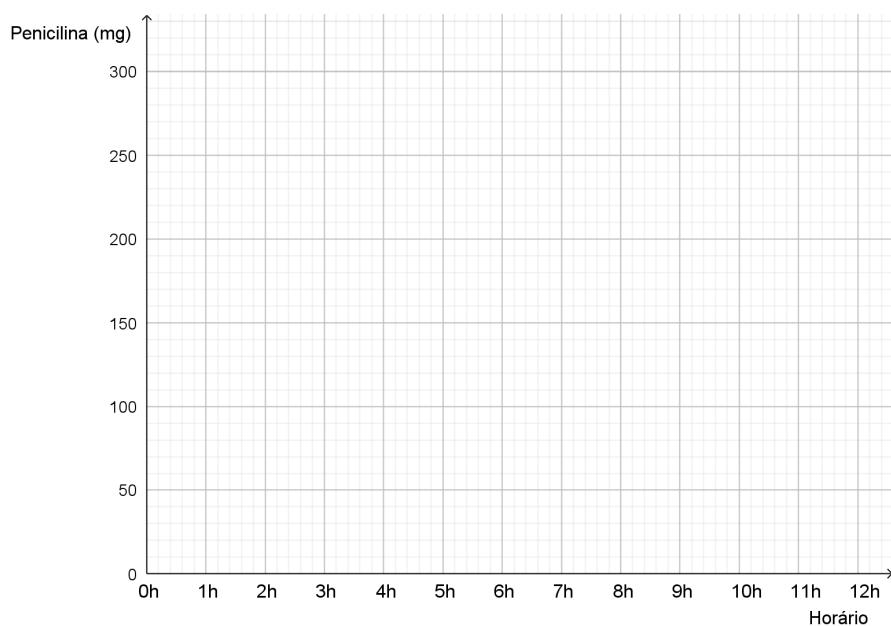
13. PISA, [s.d.], p. 44-45.

Suponha que seja administrada, a esta mulher, uma dose de 300 miligramas de penicilina às 8 horas da manhã.

- A) Complete o quadro abaixo, mostrando a quantidade de penicilina que permanecerá ativa no sangue da mulher em intervalos de uma hora, no período das 8h às 11h da manhã.

Horário	8h	9h	10h	11h
Penicilina (mg)	300			

- B) No plano cartesiano, marque os pontos que representam os dados do quadro acima e esboce a curva que representa a concentração de penicilina presente no corpo dessa mulher a partir do instante em que ela recebeu a injeção com 300 mg de penicilina.



Chave de correção:

- A) Crédito integral: preenche corretamente os campos do quadro.

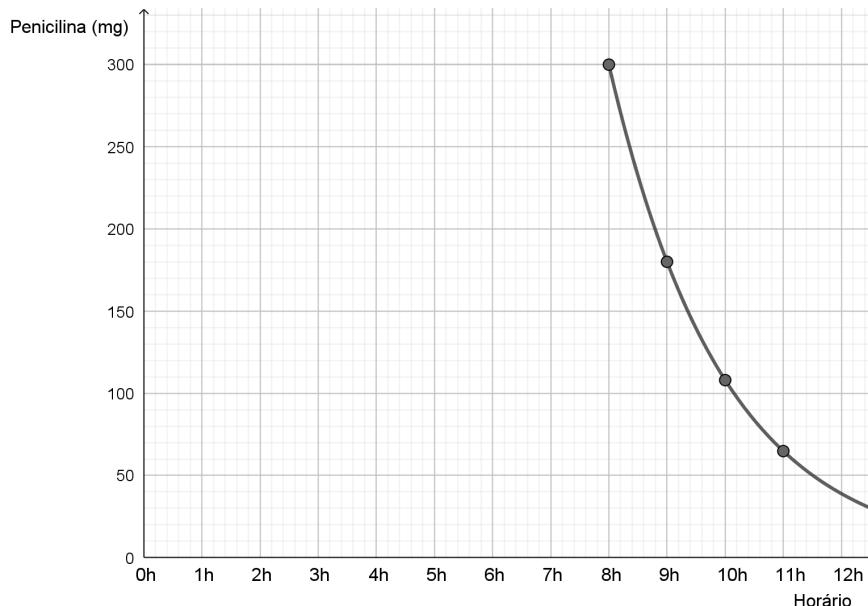
Horário	8h	9h	10h	11h
Penicilina (mg)	300	180	108	64,8

Observe que, nesse contexto, seria aceitável que, na última célula do quadro, o estudante preenchesse com 65, em lugar de 64,8, pois, nesse caso, ele teria feito as contas corretamente e apresentado uma quantia aproximada.

Crédito parcial: preencher o quadro com um ou dois valores corretos.

Nenhum crédito: Não preencher o quadro ou preenchê-lo somente com valores incorretos.

B) Crédito integral: marca corretamente os quatro pontos correspondentes aos quatro pares ordenados extraídos do quadro e esboça uma curva exponencial contínua, iniciando no ponto de coordenadas (8, 300).



Crédito parcial 1: representa os pontos obtidos no quadro corretamente, traça a curva pelos pontos representados, mas não se atenta ao fato de que a curva tem início no ponto de coordenadas (8, 300), pois o gráfico se estende também para a esquerda desse ponto.

Crédito parcial 2: representa os pontos obtidos no quadro corretamente, mas não traça a curva pelos pontos OU representa os pontos obtidos no quadro corretamente, mas une os pontos por segmentos de reta OU tenta traçar uma reta pelos quatro pontos representados.

Nenhum crédito: representa os pontos obtidos no quadro de forma incorreta OU não representa os pontos no plano cartesiano.

Note que esse item avalia várias habilidades. Na primeira parte, o estudante é convidado a preencher um quadro envolvendo cálculo de porcentagem, em que os cálculos subsequentes dependem dos anteriores. Portanto, a habilidade avaliada na questão (a) é “aplicar cálculo de porcentagem na resolução de problema”.

Já na questão (b), é solicitado que o estudante represente os dados do quadro por ele preenchido em um plano cartesiano, considerando os pares ordenados criados a partir desse quadro. Aqui, a habilidade que está sendo avaliada é “representar pares ordenados obtidos a partir de uma tabela como pontos em um plano cartesiano”. Ao final da questão (b), o estudante ainda deve esboçar a curva que representa a situação descrita no texto, o que, nesse caso, demanda a percepção de que a concentração de penicilina no corpo decresce exponencialmente e que a curva solicitada deve ser parte do gráfico de uma função exponencial – isto é, uma curva do tipo exponencial iniciando no ponto de coordenada (8, 300).

Sendo assim, nessa parte final, o que está sendo avaliado é a capacidade de o estudante “esboçar uma curva exponencial que representa um fenômeno descrito textualmente”. Observe que essa situação poderia ser explorada de outras formas, caso propuséssemos, por exemplo, uma questão (c) na qual poderia ser solicitada a lei de formação da função que descreve a concentração de penicilina presente no corpo dessa pessoa, a partir do instante em que 300 mg de penicilina são aplicadas em seu organismo.

3.5 Considerações finais sobre o uso de itens de resposta construída

Para se escolher o formato da atividade avaliativa a ser trabalhada na sala de aula, é fundamental que o procedimento seja bem definido pelo professor ou pela professora, que pode se orientar por algumas perguntas-guia, tais como:

- Quais resultados desejo observar com esse item ou com a avaliação como um todo?
- Quanto tempo os estudantes terão para responder ao item ou à avaliação como um todo?
- Quais conteúdos podem ser ou serão escolhidos para serem avaliados?
- Será uma atividade individual ou em grupo?
- A que materiais, recursos, os alunos terão acesso para executar a tarefa a ser proposta?

Responder a esses questionamentos irá ajudar a traçar estratégias adequadas para o desenvolvimento de atividades avaliativas que propiciem informações relevantes sobre as aprendizagens dos estudantes.

Até aqui, conhecemos diferentes tipos e formatos de itens de resposta construída. Para finalizarmos, retomamos o que dissemos no início deste capítulo: itens de resposta selecionada e de resposta construída podem avaliar habilidades semelhantes relacionadas à compreensão leitora, mas exigem posturas distintas do leitor em face da tarefa proposta.

No caso de itens de resposta selecionada, como já anuncia o nome, o estudante escolhe, dentre as alternativas apresentadas, aquela que ele entende que corresponda à resposta correta, ou seja, a alternativa que atenda ao que foi perguntado no comando do item. Para isso, o estudante precisa ler o texto dado como suporte ou a situação-problema, construir um significado para o que foi apresentado nesses estímulos, ler e compreender a tarefa solicitada no comando para resposta e, então, selecionar a alternativa que atenda ao que foi pedido.

Já para responder aos itens de resposta construída, além dos processos de compreensão elencados para o outro tipo, conforme apresentados anteriormente, os estudantes precisam elaborar uma representação em linguagem das ideias criadas, da compreensão do que foi lido, o que vai desde o nível da palavra, passando pela frase até chegar ao texto, além de fazer a integração da ação motora ao conhecimento dos princípios que regem o sistema de escrita alfabética, para finalmente produzir sua resposta escrita.

Isso nos remete, mais uma vez, ao fato de que itens de resposta construída não só podem abordar habilidades mais complexas, mas também envolvem processos cognitivos mais complexos.

4. O que você aprendeu até aqui?

Neste capítulo, você realizou um percurso que lhe permitiu conhecer e aprender sobre itens de resposta construída, ampliando e aprofundando seus conhecimentos a respeito desse tipo de instrumento comumente utilizado em atividades avaliativas no cotidiano escolar, mas também presente em avaliações em larga escala.

Sendo assim, você aprendeu sobre as vantagens e as desvantagens da aplicação de itens de resposta construída, além das características gerais e dos desafios em torno do seu uso. Isso envolve desde a questão do tempo para que os itens sejam devidamente corrigidos, bem como a construção de suas respectivas chaves de correção, que devem trazer as informações necessárias para que o professor ou a professora identifique o nível de desenvolvimento do estudante com base em sua resposta.

Também foi possível conhecer um pouco mais a respeito dos processos cognitivos passíveis de serem avaliados por meio desse tipo de instrumento, incluindo tanto os que se adequam mais a esse modelo de item quanto aqueles comuns aos itens de resposta selecionada. Para isso, você foi apresentado a diferentes formatos de itens de resposta construída, com exemplos e tipos comumente utilizados em avaliações em larga escala, em âmbitos nacional e internacional.

Tais quais os instrumentos de resposta selecionada, os itens de resposta construída também demandam que sejam consideradas algumas diretrizes para a sua elaboração. Portanto, você pôde verificar, aqui, os principais critérios a serem seguidos, de modo a se garantir qualidade técnica e pedagógica.

Por fim, este capítulo também discorreu sobre duas possibilidades diferentes para a correção de um item de resposta construída, denominadas como chave de correção holística e chave de correção analítica. Com isso, foi possível conhecer características, vantagens e desvantagens de cada uma, por meio de diferentes exemplos de chaves de correção de itens utilizados em avaliações em larga escala. Afirmamos, também, que a correção analítica tende a ser mais adequada para fins de uma avaliação de caráter formativo, por apresentar e descrever níveis detalhados de desenvolvimento para diferentes percursos cognitivos.

5. Referências

- ANDERSON, L. W.; KRATHWOHL, D. R.; AIRASIAN, P. W.; CRUIKSHANK, K. A.; MAYER, R. E.; PINTRICH, P. R.; RATHS, J.; WITTRICK, M. C. (ed.). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. 2. ed. New York: Longman, 2001.
- BRASIL. Ministério da Educação. *Base Nacional Comum Curricular*, 2017.
- DOWNING, S. M.; HALADYNA, Thomas M. *Handbook of test development*. Mahwah, New Jersey/London: Lawrence Erlbaum Associates Publishers, 2006.
- ESPÍRITO SANTO. Secretaria de Estado da Educação. *PAEBES 2016: Revista do Professor – Produção Textual*. v.1, jan./dez. 2016. Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd, 2016.
- FERREIRO, E.; TEBEROSKY, Ana. *A psicogênese da língua escrita*. Porto Alegre: Artmed, 1999.
- HALADYNA, Thomas M.; RODRIGUEZ, Michael C. *Developing and Validating Test Items*. New York: Routledge, 2013.
- HOGAN, T. P. *Introdução à prática de testes psicológicos*. Rio de Janeiro: LTC, 2006.
- IAVE. *PIRLS 2016: literacia de leitura & literacia de leitura online* – Unidade de avaliação. Lisboa: IAVE, 2016. Disponível em: https://iave.pt/wp-content/uploads/2019/08/Unidades_AvaliacaoPIRLS_ePIRLS_2016.pdf. Acesso em: 17 fev. 2022.
- IAVE. *PIRLS 2011: Textos e itens disponibilizados ao público*. IAVE, 2011. Disponível em: https://iave.pt/wp-content/uploads/2019/08/PIRLS_2011_Itens_Libertos_Finali_Publica.pdf. Acesso em: 17 fev. 2022.
- MULLIS, Ina V. S.; MARTIN, Michael O. (ed.) *PIRLS 2021 Assessment Framework*. Chestnut Hill: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College And International Association for the Evaluation of Educational Achievement (IEA), 2019.
- OSTERLIND, S. J. *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats*. 2. ed. Dordrecht: Kluwer Academic Publishers, 2002.
- PISA. *Itens liberados de matemática*. [s.d.]. Disponível em: https://download.inep.gov.br/download/internacional/pisa/Itens_Liberados_Matematica.pdf. Acesso em: 19 mar. 2022.
- STUDART, D. C.. Conhecendo a experiência museal das crianças por meio de desenhos. In: MASSARANI, Luisa (ed.). *Ciência e criança: a divulgação científica para o público infanto-juvenil*. Rio de Janeiro: Museu da Vida/Casa de Oswaldo Cruz/Fiocruz, 2008. Disponível em: http://www.museudavida.fiocruz.br/images/Publicacoes_Educacao/PDFs/cienciaecrianca.pdf. Acesso em: 11 mar. 2022.

6. Testando o que você aprendeu

Atividades de conhecimentos gerais sobre itens de resposta construída

1) A respeito dos itens de resposta construída, são feitas as seguintes afirmativas:

- I. Os itens de resposta construída têm aplicabilidade restrita por exigirem a elaboração de chaves de correção.
- II. Os itens de resposta construída podem ter respostas objetivas e respostas subjetivas.
- III. Os itens de resposta construída podem avaliar as mesmas habilidades de um item de múltipla escolha.
- IV. Itens de resposta construída de resposta subjetiva devem ser evitados no ensino fundamental.

Estão corretas apenas as afirmativas

- A) I e II.
- B) I e III.
- C) II e III.
- D) II e IV.
- E) III e IV.

2) Associe as duas colunas, relacionando os termos “vantagem” e “desvantagem” às características dos itens de resposta construída.

- 1. Vantagem
- 2. Desvantagem

- () Permitem avaliar habilidades mais complexas.
- () Podem levar o estudante a desenvolver um raciocínio criativo.
- () Exigem mais tempo para serem corrigidos.
- () Podem ter a correção influenciada pela subjetividade do corretor.
- () Dificultam que o estudante acerte ao acaso.

A sequência correta, de cima para baixo, dessa associação é

- A) 1, 1, 1, 2, 1.
- B) 1, 1, 2, 2, 1.
- C) 1, 2, 1, 1, 2.
- D) 2, 1, 2, 2, 1.
- E) 2, 2, 1, 2, 2.

3) Para a elaboração de itens de resposta construída, são estabelecidas algumas diretrizes. Classifique-as como verdadeiras (V) ou falsas (F).

- () A definição de critérios de pontuação deve ser algo facultativo.
- () A escolha do formato do item deve ser compatível com a demanda cognitiva pretendida.
- () A aplicação desse item, na avaliação em larga escala, ocorre logo após sua elaboração.
- () A linguagem deve ser adequada aos estudantes, evitando-se ambiguidades.

A sequência de classificação correta, de cima para baixo, é

- A) F, F, F, V.
- B) V, V, F, V.
- C) F, V, V, F.
- D) F, V, F, V.
- E) V, F, V, F.

Atividades sobre itens da etapa de alfabetização.

4) Veja o item abaixo para responder às questões 4.1 a 4.4

👉 Veja a figura abaixo.



Disponível em: <<https://bit.ly/3KgR97z>>. Acesso em: 4 mar. 2022.

👉 Escreva o nome dessa figura.

- 4.1) Qual habilidade esse item avalia?
- 4.2) A palavra escolhida como suporte está adequada para estudantes do 1º ou 2º anos? Justifique sua resposta.
- 4.3) Quais critérios linguísticos definem essa palavra?
- 4.4) A habilidade está sendo avaliada a partir de uma imagem oferecida como suporte. Essa imagem alcança o objetivo da avaliação da habilidade? Justifique.

Veja agora a resposta do estudante e leia a chave de correção apresentada na sequência.

👉 Veja a figura abaixo.



Disponível em: <<https://bit.ly/3KgR97z>>. Acesso em: 4 mar. 2022.

👉 Escreva o nome dessa figura.

COPU

RELAÇÃO COM A IMAGEM	
PALAVRA EXATA	Sinalize essa opção quando o estudante escrever exatamente a palavra solicitada, em conformidade com a imagem. Aqui serão considerados, também, os regionalismos/variações dialetais, desde que estejam dicionarizados ou listados no Vocabulário Ortográfico da Língua Portuguesa.
CAMPO SEMÂNTICO	Sinalize essa opção quando o estudante escrever uma palavra dentro do campo semântico suscitado pela imagem, mas também empregada para nomear essa imagem. Atenção! Também receberá essa indicação a palavra que, no aspecto ORTOGRAFIA, receber o conceito E.
ORTOGRAFIA	
A	O estudante escreveu a palavra corretamente, configurando uma escrita em nível ortográfico . Atenção! Caso a palavra escrita não coincida com a nomeação da imagem, somente será válida a resposta que possuir estrita relação com a figura apresentada, ou seja, que pertença ao mesmo campo semântico.
B	O estudante escreveu a palavra com desvios, configurando uma escrita em nível alfabético .
C	O estudante escreveu a palavra com desvios, configurando uma escrita em nível silábico-alfabético .
D	O estudante escreveu a palavra com desvios, configurando uma escrita em nível silábico . Nesse caso, a escrita apresentada compromete a leitura, não sendo possível sua compreensão sem conhecer a imagem. Contudo, a escrita apresenta relação com a pauta sonora das letras que compõem a palavra.
E	O estudante produziu escrita em nível pré-silábico (apresentou letras aleatórias para representar a palavra, desenhos/garatujas) e/ou grafou outra palavra que não guarda estreita relação com a imagem (provavelmente reproduzida de memória).

- 4.5) Com base na resposta dada pelo estudante, qual das alternativas da chave de correção você marcaria? Por quê?
- 4.6) Que estratégia didática você proporia para que esse estudante desenvolvesse totalmente a habilidade avaliada?
- 4.7) Escreva um exemplo de como seria uma resposta que você atribuiria ao nível E da chave de correção.

- 5) **Veja o item e a sua chave de correção abaixo para responder às questões 5.1 e 5.2.**

Veja a cena abaixo.



✍ Escreva uma frase contando o que acontece nessa cena.

ASPECTO 1 – PLAUSIBILIDADE COM A IMAGEM

	O estudante escreveu uma frase (ou um texto) plausível em relação à cena. Atenção! A A avaliação desse aspecto não considera se a frase possui, ou não, desvios ortográficos, de segmentação ou qualquer outro de caráter linguístico. Observa-se, apenas , a coerência presente na descrição realizada. Para atribuição do conceito “A”, podem ser consideradas plausíveis estruturas oracionais (com presença de verbo) ou frases nominais (que não exprimem ação ou algo semelhante, uma vez que não apresentam verbos), desde que se relacionem de modo coerente com a imagem.
B	O estudante escreveu uma frase (ou um texto) com pouca plausibilidade em relação à cena.
C	O estudante escreveu uma frase (ou um texto) incoerente em relação à cena, configurando fuga ao tema . Atenção! Enquadram-se, aqui, os casos nos quais o estudante escreveu apenas o nome de 1 (um) dos elementos que compõem a cena.

ASPECTO 2 – ORTOGRAFIA	
	O estudante escreveu ortograficamente uma frase (ou um texto), respeitando os critérios de segmentação depalavras.
A	Atenção! A frase que não contiver verbo será avaliada, automaticamente, com o conceito D nesse aspecto, independentemente da avaliação atribuída no Aspecto 1. Será considerada em nível ortográfico a frase que apresentar apenas 1 (um) tipo de desvio não recorrente,exceto de segmentação.
B	O estudante escreveu uma frase (ou um texto) com desvios, configurando uma escrita em nível alfabético .
C	O estudante escreveu uma frase (ou um texto) com desvios, configurando uma escrita em nível silábico-alfabético .
D	O estudante escreveu uma frase (ou um texto) com desvios, configurando uma escrita em nível silábico . Atenção! Também receberão esse conceito as frases nominais, ou seja, que não apresentam verbo (elencam um ou mais elementos da cena) OU uma frase cuja leitura está comprometida, mas que apresenta relação com apauta sonora de palavras que nomeiam elementos da cena.
E	O estudante apresenta escrita em nível pré-silábico (letras aleatórias e/ou desenhos/garatujas).

- 5.1) Considerando os três níveis do Aspecto 1 – Plausibilidade com a imagem –, registre como seria um exemplo de resposta para cada um desses níveis.

ASPECTO 1 – PLAUSIBILIDADE COM A IMAGEM	
A	
B	
C	

- 5.2) Seria possível um estudante encontrar-se no nível A do Aspecto 1 – Plausibilidade –, mas situar-se no nível E do Aspecto 2 – Ortografia? Justifique sua resposta.

Atividades sobre itens de resposta construída de língua portuguesa.

6) Leia com atenção os dois textos a seguir para responder às questões 6.1 a 6.5.

Texto 1



Disponível em: <https://saiasjustas.wordpress.com/tag/charge-reforma-ortografica/>. Acesso em: 17 mar. 2022.

Texto 2

“Um conhecido conto popular retrata que um ladrão foi surpreendido pelas palavras de Rui Barbosa ao tentar roubar galinhas em seu quintal:

— Não o interpelo pelos bicos de bípedes palmípedes, nem pelo valor intríngueco dos retrocitados galináceos, mas por ousares transpor os umbrais de minha residência. Se foi por mera ignorância, perdão-te, mas se foi para abusar da minha alma prosopopéia, juro pelos tacões metabólicos dos meus calçados que dar-te-ei tamanha bordoada no alto da tua sinagoga que transformarei sua massa encefálica em cinzas cadavéricas.

O ladrão, todo sem graça, perguntou:

— Mas como é, seu Rui, eu posso levar o frango ou não?

MAIA, Márcio Barbosa. *Rui Barbosa, o ladrão de galinhas e o juridiquês*. Disponível em: <https://www.conjur.com.br/2010-mai-23/rui-barbosa-ladro-galinhas-juridiques-decano-unb>.

Acesso em 17 mar. 2022.

A) Esses textos pertencem ao mesmo gênero textual? Justifique sua resposta, elencando pelo menos 2 características que os aproximam ou diferenciam.

B) No Texto 1, as personagens fazem uso de uma mesma linguagem, característica do meio virtual, também identificada como “internetês”. Explique, com base no Texto 1, em que consiste esse tipo de linguagem.

C) No Texto 2, por que o ladrão faz a pergunta ao final?

- 6.1) Analise os dois textos e aponte os elementos que os aproximam e que os diferenciam.
- 6.2) Esses itens de resposta construída são de que tipo: resposta objetiva ou resposta subjetiva? Justifique sua resposta.
- 6.3) Indique a habilidade avaliada em cada uma das atividades acima.
- 6.4) Qual percurso cognitivo o estudante precisa percorrer para responder à atividade C?
- 6.5) Elabore uma chave de correção analítica para o item A.

Atividades sobre itens de resposta construída de matemática

- 7) Considere a seguinte habilidade, própria do 5º ano do ensino fundamental, para responder às questões 7.1 e 7.2.**

Habilidade: Efetuar divisão entre números naturais, com divisor sendo formado por até dois algarismos.

7.1) Elabore um item de resposta construída para avaliar essa habilidade.

7.2) Elabore uma chave de correção detalhada para o item elaborado.

- 8) Reconstrua o item que você propôs na Atividade 7, agora em formato de resposta selecionada, com quatro alternativas de resposta, e compare a qualidade de informação que as duas estruturas de item podem oferecer ao professor ou à professora, em termos de diagnóstico do desenvolvimento do estudante.**

- 9) Veja o item abaixo para responder às questões 9.1 a 9.3.**

Ana e Beatriz foram a uma papelaria para comprar lápis e canetas. Ana comprou dois lápis e uma caneta e pagou R\$ 12,00. Beatriz comprou três lápis e duas canetas, idênticas às compradas por Ana, e pagou R\$ 21,00.

Quais são os preços desses tipos de lápis e de caneta nessa papelaria?

9.1) Resolva esse item de forma detalhada.

9.2) Indique qual habilidade esse item está avaliando.

9.3) Elabore uma chave de correção detalhada para esse item, contemplando as etapas intermediárias envolvidas no processo de sua resolução.

LEITURA E INTERPRETAÇÃO DE RESULTADOS DE AVALIAÇÕES EM LARGA ESCALA¹⁴

Luis Antônio Fajardo

Joaquim José Soares Neto

Manuel Palacios Cunha Melo

Wellington Silva

1. O que você vai aprender aqui?

- Introdução à análise pedagógica e estatística dos itens
- Definição do propósito de um teste
- Considerações sobre a construção dos testes educacionais
- Teoria Clássica dos Testes aplicada à avaliação educacional
- Fidedignidade e validade dos testes de proficiência
- Uso da Teoria da Resposta ao Item na interpretação dos resultados das avaliações
- Critérios de ancoragem de itens na construção de escalas de proficiência
- Padrões de desempenho estudantil para nortear as interpretações pedagógicas das escalas de proficiência

14 Parcialmente adaptado da disciplina Introdução à Estatística Aplicada à Educação do Módulo de Nivelamento do curso de especialização em Estatísticas e Avaliação Educacional (CAEd, 2020).

2. Iniciando a construção do conhecimento

O propósito deste capítulo é fornecer um conjunto de informações que permitam a compreensão e a apropriação dos resultados das avaliações em larga escala, de modo que seja possível utilizá-los na sua prática profissional, seja como docente em sala de aula, seja como gestor educacional, seja na pesquisa acadêmica.

Quando se trata da compreensão, da apropriação e da utilização dos resultados das avaliações, é necessário perceber também os seus respectivos limites. Isso porque os resultados tanto de uma avaliação externa quanto interna são parte de um contexto muito mais amplo do fenômeno educacional. Outro ponto relevante é o fato de que as escolas e os estudantes avaliados não costumam se encontrar em pé de igualdade, do ponto de vista das condições de aprendizado, sendo possível citar inúmeros fatores que influenciam ou podem influenciar os resultados educacionais.

Por exemplo, a condição socioeconômica dos estudantes e de suas respectivas famílias costuma estar fortemente associada ao desempenho escolar. Também as condições das escolas – como a infraestrutura e os recursos didáticos – igualmente têm seu peso no desempenho educacional, além do papel dos professores, do clima em sala de aula, do tipo de liderança escolar, das expectativas de todos em relação ao sucesso educacional etc.

Ainda assim, os resultados das avaliações externas proporcionam suporte valioso para o progresso educacional. Por meio deles, é possível observar – de modo bastante preciso e abrangente – uma ampla gama de desempenhos dos estudantes, que, se não são completamente exaustivos, ainda assim cobrem um espectro consideravelmente largo e importante do currículo e da realidade educacional.

E, por meio desse conhecimento, descortina-se uma série de vantagens para os usuários dessas informações. Com elas, por exemplo, é possível conhecer, dentro de certos limites, o real estado de aprendizado dos alunos de uma determinada escola ou sistema educacional. Também se podem observar quais são os pontos fortes do ensino em questão, por meio da verificação de quais habilidades já estão sendo suficientemente dominadas pelos estudantes, bem como os pontos fracos e os desafios existentes, relacionados às partes do currículo que ainda não foram devidamente cobertas ou assimiladas. Todas essas informações, por sua vez, além de poderem ser utilizadas pelos profissionais do “chão da escola” – como professores, coordenadores, gestores

escolares etc. –, também podem ser úteis em instâncias superiores da administração educacional, como no caso dos técnicos e gestores das secretarias municipais e estaduais de educação.

Dessa forma, a compreensão dos resultados educacionais, conforme comumente divulgados pelas avaliações em larga escala, mostra-se muito importante para os profissionais comprometidos com a melhoria da educação ofertada nas escolas. E é exatamente sobre isso que nos deteremos neste capítulo.

3. Construindo o conhecimento

3.1 A definição do propósito de um teste

Um teste exige que seus propósitos estejam muito bem delimitados, a fim de ser adequadamente concebido, produzido, aplicado e interpretado. Afinal de contas, é exatamente a partir da delimitação dos objetivos de um teste que este se realiza, o que por sua vez, pode também acarretar consequências para os diversos atores envolvidos em suas diferentes etapas.

Há uma gama de propósitos que um teste educacional pode possuir, como a divisão entre avaliações somativas e formativas. Nesse caso, enquanto as primeiras se preocupam com a mensuração de resultados para fins de aferição do desempenho após um ciclo escolar – por exemplo, com o propósito de determinar como os alunos se saíram frente a um determinado conteúdo curricular, de modo a orientar o (re) planejamento do próximo ciclo –, as avaliações formativas têm como propósito básico produzir um diagnóstico sobre o desenvolvimento dos estudantes em relação às habilidades previstas no currículo, a fim de fornecer subsídios para o seu respectivo aperfeiçoamento ao longo do ano letivo.

Daí também se depreende que alguns testes têm uma função de verificação do desempenho, às vezes atrelada a possíveis consequências, como a aprovação dos estudantes, ao mesmo tempo que outras avaliações possuem mais um caráter de diagnóstico e de orientação para a formulação ou ajuste das práticas pedagógicas ou das políticas educacionais.

Quanto às consequências dos testes, elas costumam se dividir, *grosso modo*, entre altas ou baixas (como os termos em inglês, bastante utilizados na literatura internacional, *high stakes* e *low stakes*, respectivamente). Os testes de altas consequências são aqueles que, conforme o próprio nome indica, têm seus respectivos resultados capazes de gerar um considerável impacto na vida dos estudantes que os realizam, ou dos profissionais das escolas envolvidas. Entre os testes de altas consequências mais comuns no Brasil, pode-se citar, por exemplo, o Exame Nacional do Ensino Médio (Enem), que é normalmente utilizado como um critério de ingresso em cursos ou carreiras específicas da educação superior. Outro exemplo de testes de altas consequências pode ser encontrado nos Estados Unidos, onde certos estados, ocasionalmente, preveem a possível interferência pública na gestão de escolas cujos resultados se mostram sistematicamente baixos ao longo de diversas avaliações externas. E isso poderia acarretar, inclusive, a designação de novos membros da gestão escolar, como diretores e professores, com o consequente afastamento dos profissionais da gestão anterior.

Já os testes de baixas consequências – ao menos para os estudantes que os fazem – podem ser exemplificados, no Brasil, por provas como as do Sistema Nacional de Avaliação da Educação Básica (Saeb) e por vários outros testes congêneres aplicados pelos diversos estados brasileiros. Em casos assim, o resultado da prova tem, geralmente, pouca implicação para o aluno que participa, de modo que redes e sistemas educacionais precisam lançar mão de outros esforços para obter engajamento por parte dos estudantes e garantir a validade dos testes e de seus resultados. Por outro lado, muitas dessas avaliações externas no Brasil são atreladas a benefícios e bonificações aos profissionais envolvidos, obtidos de acordo com os resultados alcançados, o que pode fazer com que suas consequências sejam altas para esses atores.

Deve-se ter em mente, também, que, quando atrelados a altas consequências, os testes passam a ter um maior potencial de estimular um comportamento inadequado – ou mesmo antiético – por parte dos atores educacionais, com vistas à obtenção de melhores resultados. Isso pode implicar, por exemplo, a adoção de estratégias de “ensinar para o teste”, ou seja, de treinarem excessivamente seus estudantes para responder ao conteúdo e à forma típicos das avaliações em larga escala, deixando, assim, de se preocupar com uma formação mais profunda e diversificada. Porém, um caso certamente pior nesse sentido seria a pura e simples ocorrência de fraudes.

Outra possível divisão dos testes é quanto à unidade de análise, que se desdobra em diversas possibilidades. Nas avaliações em larga escala mais conhecidas, as maiores unidades costumam corresponder a sistemas educacionais inteiros, como uma rede federal, estadual ou municipal de educação, e as menores unidades, às escolas.

Por sua vez, a escolha dos alunos como a menor unidade de análise costuma ser menos comum, embora os testes dessas avaliações sejam aplicados a eles habitualmente. Em parte, isso ocorre por questões gerenciais, que tendem a se deter em unidades maiores de interesse, tendo em vista que, na maioria das vezes, são as administrações centrais que se ocupam da análise dos resultados. E, em parte, também decorre do problema psicométrico pelo qual, para se medir a proficiência individual de um dado aluno com muita precisão, seria preciso lhe aplicar testes muito mais longos do que aqueles que se costuma utilizar hoje em dia. Porém, em avaliações formativas, geralmente aplicadas em diferentes momentos do ano letivo e cujo foco maior é o professor ou a professora, os resultados dos estudantes tendem a receber mais destaque, pois é com base neles que o profissional tomará decisões relacionadas ao ensino do currículo em sala de aula.

Sendo assim, independentemente do tipo de avaliação que se deseja aplicar, há dois pontos importantes a serem considerados. Um deles diz respeito à qualidade dos testes, que deve ser suficientemente elevada e adequada aos propósitos da avaliação – segundo alguns critérios a serem vistos mais adiante, como a validade dos testes e sua fidedignidade. O segundo, por sua vez, relaciona-se ao fato de que os resultados de uma avaliação devem ser analisados com certa cautela, tendo em vista suas limitações intrínsecas e levando em conta o contexto no qual foi aplicada.

3.2 Considerações sobre a construção dos testes educacionais

Definido o propósito do teste, o próximo passo a seguir é prepará-lo para aplicação. Essa etapa, por sua vez, envolve diversos procedimentos, tais como:

- 1) *Definir que tipo de item ou itens se deseja utilizar, com base nos propósitos do teste já estabelecidos.* Neste momento, deve-se fazer a escolha a respeito do formato dos itens, cujas possibilidades já abordamos nos Capítulos 2 e 3 deste livro. A escolha entre itens de resposta selecionada e de resposta construída irá impactar nos procedimentos referentes desde a aplicação até a correção e a interpretação dos resultados. Sabemos que os itens de resposta selecionada apresentam, entre seus pontos fortes, maior fidedignidade, rapidez e economia de correção. Por outro lado, entre suas limitações, está o fato de que, com eles, torna-se mais difícil mensurar algumas competências e habilidades mais complexas, como a criatividade. Já os itens de resposta construída permitem uma maior abrangência da avaliação, porém trazem também alguns problemas e desafios, pelos quais as correções tendem a

ser menos fidedignas, mais lentas e mais caras. Naturalmente, é possível aumentar a fidedignidade de correção dos itens de resposta construída, conforme já apontamos no Capítulo 2, por meio da criação de uma chave de correção, do treinamento dos corretores e da supervisão das suas atividades. Porém, como já adiantamos, isso tende a deixar o processo mais lento e dispendioso.

- 2) *Elaborar os itens em quantidade e variedade suficientes para cobrir todo o teste.* Neste caso, normalmente se treina uma equipe de especialistas, que deverão produzir itens que sejam, preferencialmente, 2 ou 3 vezes mais numerosos do que a quantidade N estipulada de itens para comporem o teste. O motivo disso é que, na fase seguinte, a da pré-testagem, muitos itens costumam ser descartados por apresentarem qualidades psicométricas abaixo do desejado.
- 3) *Aplicar um pré-teste dos itens.* O pré-teste deve, preferencialmente, ser aplicado em uma amostra de alunos que seja típica da população-alvo do teste. A partir de seus resultados, são obtidas as estatísticas de cada item, como as suas medidas de ajuste, graus de dificuldade etc., as quais deverão ser levadas em conta na etapa seguinte, a da montagem dos testes.
- 4) *Selecionar os itens que irão compor a avaliação.* Aqui, geralmente se opta por itens que possuem maior fidedignidade, o que, em termos técnicos, corresponde aos itens que apresentam maior correlação com o desempenho geral no teste. Porém, nesse ponto, entram em jogo alguns fatores. Um deles é o fato de que, no pré-teste, diversos itens podem apresentar comportamento estatístico aquém do desejado, com fraco ajuste ou baixa discriminação. Itens assim devem ser rejeitados em favor daqueles que possuem melhor desempenho estatístico. Ao mesmo tempo, se o número de itens no pré-teste for muito pequeno – igual ou somente um pouco maior do que o número N de itens previstos para compor o teste –, isso pode obrigar os elaboradores dos testes a manterem certo número de itens ruins, na falta de outros, o que pode comprometer a qualidade geral. Soma-se a isso o fato de que a seleção e a consequente eliminação de itens após o pré-teste impõem o desafio de se ter que cobrir todo ou boa parte do currículo previsto. Portanto, conforme afirmado anteriormente, o mais recomendável é aumentar consideravelmente o número de itens a serem produzidos e pré-testados, embora a tendência seja tornar o processo mais lento e dispendioso.

Outro procedimento relevante na escolha dos itens é dosar o seu nível médio de dificuldade, de modo a fazer com que essa dificuldade se ajuste, da melhor maneira possível, àquilo que se pretende fazer com o teste. Isso porque, conforme comentaremos com mais detalhes posteriormente, um item tem sua capacidade máxima de discriminação quando seu respectivo parâmetro de dificuldade é igual ou semelhante à proficiência do estudante que a ele responde. Dessa forma, caso se deseje avaliar alunos de menor proficiência – conforme mensurada em testes anteriores –, melhor seria utilizar um teste mais fácil, e o contrário se sucederia caso se desejasse avaliar alunos de proficiência mais elevada.

Ainda nessa fase, é recomendável excluir itens que apresentam valores significativos de funcionamento diferencial – *Differential Item Functioning* (DIF). Trata-se de casos em que os itens apresentam diferenças consideráveis em seus respectivos parâmetros – como o de dificuldade – em sujeitos ou amostras de mesma proficiência média, porém pertencentes a grupos diferentes, segundo um critério qualquer, como raça/etnia, nível socioeconômico etc. Nesses casos, é mais recomendável eliminar os itens com altos valores de DIF, devido a critérios de equanimidade, visto que o sucesso ao responder a eles – ou a probabilidade desse sucesso acontecer – deveria ser tão somente uma função da habilidade ou proficiência do aluno, e não de um grupo – racial, social etc. – ao qual esse aluno pertence.

- 5) *Montar as provas, a partir dos itens selecionados.* Na montagem das provas, é comum utilizar-se um método chamado Blocos Incompletos Balanceados (BIB), pelo qual se obtêm diferentes versões das provas, compostas de diferentes combinações de blocos de itens. Por exemplo, pode haver 10 blocos de 13 itens cada, nomeados de A até J, de tal modo que cada prova se componha de 3 blocos ou 39 itens, segundo padrões tais como: ABC, ABD etc. Esse procedimento tem a vantagem de permitir a utilização de um maior número e variedade de itens numa mesma prova, o que contribui para se cobrir um espectro mais amplo do conteúdo a ser avaliado. Porém, isso é conseguido sem tornar as provas muito longas ou cansativas para os estudantes. Por outro lado, conforme se comentou anteriormente, isso pode comprometer a fidedignidade dos testes no nível do aluno, pois todas as habilidades não serão contempladas no nível individual, apenas agregado. Sendo assim, isso não será um problema para níveis maiores de agregação, tais como escolas, municípios, estados etc., visto que, nessas situações, o número de casos costuma ser elevado o suficiente para produzir estatísticas mais precisas.

- 6) *Aplicar a avaliação de modo suficientemente padronizado.* A padronização na aplicação dos testes tem como objetivo garantir que todos os estudantes os realizaram sob as mesmas condições, de modo que nenhum seja prejudicado ou beneficiado. Esse processo é essencial tanto para a fidedignidade quanto para a validade da avaliação, para que os resultados alcançados pelos estudantes representem, de fato, o nível de desenvolvimento deles.
- 7) *Producir e informar os resultados.* Os resultados das avaliações costumam ser publicados em boletins de resultados, boletins pedagógicos, resumos executivos, plataformas digitais, entre outros. Quanto a essa etapa, embora ela apareça como a última desta sequência, na verdade, precisa estar devidamente estruturada antes mesmo de todo o processo começar, visto que influencia todas as etapas anteriores, as quais devem se moldar a ela, e não o contrário. Isso porque todo o planejamento em torno da avaliação deve levar em conta a forma como o seu público – professores e gestores escolares e educacionais – serão informados do seu resultado, bem como as estratégias adotadas para a apropriação e uso dos dados. Por exemplo, se a rede ou o sistema educacional julgar necessária uma capacitação a gestores e docentes para a compreensão dos indicadores e uso dos resultados, é essencial que essa formação esteja articulada à própria publicação dos resultados.

3.3 A Teoria Clássica dos Testes aplicada à avaliação educacional

A Teoria Clássica dos Testes (TCT) surgiu a partir dos trabalhos de Spearman e outros autores, principalmente a partir do início do século XX. Ela se constitui de um arcabouço estatístico convencional, pelo qual se definem e aplicam diversos procedimentos matemáticos ao problema da testagem psicológica. A TCT difere-se da Teoria da Resposta ao Item (TRI) – que também será apresentada posteriormente neste capítulo – tanto do ponto de vista da concepção dos meios de estimar e interpretar os diversos parâmetros de uma avaliação – como a proficiência dos examinados, a dificuldade e a discriminação dos itens etc. –, quanto da complexidade dos cálculos envolvidos nesses processos. Tal complexidade costuma ser consideravelmente maior no caso da TRI, razão pela qual ela só obteve uma maior difusão de uso com o advento da era computacional, ao longo das últimas décadas.

Entretanto, apesar das consideráveis diferenças entre a TCT e a TRI, pode-se também dizer que ambas em geral se preocupam com os mesmos temas. Isso porque, independentemente da teoria que se utiliza para se trabalhar com os testes, há alguns pontos que são praticamente incontornáveis. Exemplos disso, a serem vistos com mais detalhes em breve, são as questões da fidedignidade e da validade dos testes, além da normalização dos escores de uma dada avaliação, a qual, por sua vez, é empregada para se produzir e interpretar os seus respectivos resultados.

Portanto, começaremos esta seção, do ponto de vista da TCT, abordando um tópico bastante relevante no presente contexto, referente à normalização dos testes para se avaliar o desempenho dos examinados.

Quando aplicamos um teste a alguém com o propósito de mensurar um traço psicológico ou a proficiência em qualquer componente curricular, invariavelmente, após a correção, obtemos alguma medida de valor. Em muitos casos, essa medida é expressa na forma de um número. Esse número pode ser, por exemplo, a quantidade de vezes que o sujeito investigado responde “sim” a uma série de perguntas, do tipo: “Você gosta de estudar matemática?”, “Você faz o dever de casa de matemática?”, “O professor corrige o dever de casa de matemática?” etc.; ou pode representar uma nota correspondente ao número de acertos no teste.

Vamos, agora, examinar uma situação em que a medida expressa na forma de um número requer atenção. Se um sujeito acerta seis itens de matemática de dez apresentados, o que de fato podemos afirmar sobre a sua proficiência no componente curricular considerado? Acertar 60% das questões num teste constituído de perguntas muito fáceis não é nada excepcional; na verdade, pode até ser considerado um resultado medíocre, dependendo do grau de facilidade do teste. Ao mesmo tempo, atingir tal proporção de acerto numa prova extremamente difícil pode ser considerado um feito notável.

Outra situação: uma pessoa realiza um teste psicológico para mensurar o seu grau de extroversão, respondendo afirmativamente a 16 das 40 perguntas feitas sobre se ela tem ou não características típicas de tal comportamento. Mais uma vez, o que esse número significa? Ou, em outras palavras, quão extrovertida é essa pessoa?

Devido a problemas como esses, os testes, para serem adequadamente interpretados, necessitam de uma *norma*, isto é, uma espécie uma “régu” capaz de trazer algum tipo de interpretação para o resultado obtido. De modo geral, podemos dizer que a utilização de uma norma permite compararmos o resultado obtido por uma pessoa em particular com o resultado obtido por um grupo maior e mais abrangente de indivíduos

ao qual a pessoa examinada também pertence, o que serve como referência para tal julgamento.

Nesse sentido, há diversas maneiras de normalizarmos os resultados de um teste. Aqui, discorreremos sobre uma das principais delas: os *percentis*.

3.3.1 Definição de percentis

Para compreendermos como funcionam os percentis, começemos com um exemplo simples. Suponha que um teste de matemática com dez questões foi aplicado a 600 alunos. Os resultados da aplicação estão sintetizados no exemplo a seguir.

Tabela 1: Exemplo de cálculo de resultados de um teste

Nota	N	%	%_acumulado
0	18	3,0	3,0
1	21	3,5	6,5
2	30	5,0	11,5
3	72	12,0	23,5
4	85	14,2	37,7
5	122	20,3	58,0
6	98	16,3	74,3
7	55	9,2	83,5
8	42	7,0	90,5
9	37	6,2	96,7
10	20	3,3	100,0
TOTAL	600	100	

Nesse quadro, em “Nota”, temos o número de acertos na prova. Como no exemplo havia dez itens, a nota pôde variar de 0 (nenhum item acertado) a 10 (todos os itens acertados).

Em “N”, por sua vez, estão listados os quantitativos de respondentes para cada nota obtida. Ou seja, é possível ver que 18 alunos zeraram a prova, 21 alunos acertaram apenas 1 item, 30 alunos acertaram 2 itens etc., até que observamos, na última linha, que 20 alunos fecharam a prova, isto é, acertaram todos os 10 itens.

Na coluna “%”, é indicado o percentual que o número da coluna anterior (N) representa em relação à totalidade de alunos avaliados (600). Dessa forma, por exemplo, vemos que os 18 alunos que zeraram a prova correspondem a 3% de todos os alunos que fizeram o teste; no outro extremo, os 20 alunos que acertaram tudo correspondem a 3,3% do total de respondentes.

Por fim, em “%_acumulado”, temos a acumulação dos percentuais de alunos, à medida que se progride do acerto nulo para o acerto total (ou seja, à medida que “descemos” na tabela). É possível observar, na primeira linha e penúltima coluna (%), que 3% dos alunos tiraram nota 0. Descendo mais uma linha, nessa mesma coluna, vemos que 3,5% dos alunos acertaram um único item (N=1). Na última coluna (%_acumulado), o valor corresponde à soma dos dois percentuais: $3\% + 3,5\% = 6,5\%$. Ou seja, 6,5% dos estudantes acertaram até um item na prova (isto é, ou acertaram só um item ou não acertaram nenhum). E, assim, à medida que descemos nessa última coluna, também vamos sucessivamente somando os demais valores percentuais da coluna anterior (%), até que, na categoria das maiores notas (10 acertos), o resultado chega a 100%.

Ao fazermos essa última operação, conseguimos dispor todos os estudantes avaliados ao longo de uma sequência que nos permite facilmente ver, por exemplo, quantos por cento deles tiraram, no máximo, uma determinada nota. Essa é precisamente a ideia dos percentis.

Passemos agora, então, a “ler” algumas informações desse último exemplo utilizando o conceito dos percentis, pois isso nos permite obter uma compreensão bem nítida do que eles significam. Algumas observações possíveis são:

- › Quem, por exemplo, no teste, tirou a nota 1 encontra-se no percentil 6,5. Isso porque quem obteve tal nota alcançou desempenho acima de – ou, na pior das hipóteses, igual a – 6,5% do total de estudantes avaliados. Outra forma de dizer isso é que o percentil 6,5 no teste corresponde à nota 1.
- › Quem acertou 5 questões encontra-se no 58º percentil, ou, de outra forma, o 58º percentil corresponde à nota 5, e assim por diante.

Um teste como o do exemplo anterior, cuja interpretação leva em conta a distribuição dos sujeitos ao longo dos próprios resultados, é chamado *teste referenciado por norma*. A ideia por trás disso é muito simples: o teste se comporta como uma “competição” entre os sujeitos que o fazem, de modo que, utilizando uma variável de interesse do teste, como a nota obtida, é possível então classificar os sujeitos que o fizeram, indo do resultado mais baixo para o mais alto.

Entretanto, um ponto importante precisa, aqui, ser observado: embora um teste referenciado por norma possa ser muito útil, já que ele nos permite ver o desempenho de um dado sujeito relativamente aos demais, os seus resultados podem ser enganosos. Vejamos outro exemplo, para entendermos melhor o motivo.

Na escala nacional de proficiência em Matemática do Saeb, o valor 250 pontos corresponde à média de todos os estudantes do país no componente curricular em 1997, para o 9º ano do ensino fundamental, com desvio padrão igual a 50 pontos.¹⁵

Suponha, então, que um município hipotético tenha dez escolas de ensino médio – nomeadas por cores, conforme a Tabela 2. Ao se aplicar um teste de matemática para estudantes do 3º ano do ensino médio, de cada uma dessas escolas, temos os resultados indicados na tabela a seguir, dispostos em ordem crescente.

Tabela 2: Resultados de escolas de um município fictício em um teste de matemática

Escola	Média matemática 3º ano EM	%	%_acumulado
Azul	180	10	10
Verde	183	10	20
Rosa	191	10	30
Amarela	199	10	40
Cinza	200	10	50
Lilás	204	10	60
Roxa	229	10	70
Vermelha	251	10	80
Laranja	255	10	90
Marrom	270	10	100

Com base nos resultados, o que pode ser dito do desempenho da escola vermelha? Segundo uma referência por norma, uma possibilidade é dizer que a escola vermelha tem a sua respectiva média (251 pontos) correspondendo ao 80º percentil. Em outras palavras, podemos assegurar que o desempenho da escola vermelha está superior – ou, na pior das hipóteses, igual – ao de 80% do conjunto de todas as escolas, o que parece ser um resultado bastante positivo.

15 Aqui, cabe destacar que a escala de proficiência, embora esteja “centrada” na média do 9º ano do ensino fundamental, estende-se para mais e para menos, de modo a englobar tanto o 3º ano do ensino médio quanto o 5º ano do ensino fundamental.

Por outro lado, também pode ser dito que, embora a escola vermelha tenha ficado acima da maioria das escolas, a sua média para o 3º ano do ensino médio – 251 pontos – é algo típico de alunos do 9º ano do ensino fundamental, cuja média de desempenho na escala é 250 pontos. Ou seja, o desempenho da escola vermelha em termos relativos às demais escolas é muito bom, porém, em termos absolutos, é, na verdade, um tanto quanto decepcionante.

Outro exemplo pode apontar para um problema semelhante, porém, na direção oposta. Imagine um conjunto de escolas bem-sucedidas em proporcionar aos seus estudantes uma excelente aprendizagem, atingindo médias elevadas em uma avaliação específica. Quando calculamos os percentis para os resultados, embora todas as médias de proficiência sejam altas, algumas serão ainda mais altas do que outras. E, assim, a escola desse conjunto que tiver a menor média – ou uma das menores médias – em relação às demais poderia ser, de certa forma, “injustiçada”, pois estaria recebendo uma classificação de “perdedora”, embora tenha tido um ótimo desempenho em termos absolutos.

3.3.2 Testes referenciados por norma e critério e dificuldade dos itens

Problemas como esses nos fazem perceber que é preciso ter cuidado quando observamos os resultados de uma avaliação conforme dispostos dos menores para os maiores números – o dito ranqueamento ou *ranking*. Na verdade, sempre que utilizamos resultados apresentados dessa forma, corremos o risco de incorrermos em graves problemas de interpretação. Resultados apresentados assim, porém, são comuns quando observamos os dados das avaliações externas em educação.

Uma forma de se resolver tal problema é pensar no chamado *teste referenciado por critério*. Nesse caso, a avaliação do desempenho baseia-se não no conjunto de sujeitos que participam da prova, mas sim em uma série de critérios costumeiramente definidos antes mesmo de se aplicar o teste. De tal forma que, com base na escala obtida, é possível avaliar o desempenho dos sujeitos e, consequentemente, criar ou definir *padrões de desempenho*, classificando, de acordo com estes, os sujeitos observados.

No exemplo ilustrado, caso tivéssemos nos amparado da interpretação por critério, poderíamos, *a priori*, ter definido quais resultados – na escala Saeb – seriam considerados baixos, médios ou altos para o 3º ano do ensino médio.

Para isso, poderíamos, por exemplo, ter determinado os chamados *níveis de proficiência*, que são os sucessivos intervalos na escala adotada, de modo a cobrir toda a faixa de proficiência. Caso tivéssemos feito isso, certamente nenhuma das escolas listadas na ilustração estaria bem classificada, visto que, no máximo, somente algumas delas conseguiram ter o seu desempenho médio comparável ao do 9º ano do ensino fundamental, embora todas pertencessem ao ensino médio.

Outra consideração importante sobre o resultado de uma avaliação educacional diz respeito à dificuldade dos itens que compõem o teste. Como o próprio nome diz, a dificuldade de um item em um teste se associa ao fato de o item se mostrar mais fácil ou mais difícil para os sujeitos avaliados.

Naturalmente, a dificuldade do item depende também da forma como ele é corrigido. Em alguns casos, como redações ou itens dissertativos, por exemplo, é comum que a “nota” do examinado possa teoricamente variar ao longo de um intervalo, como de 0 a 10 pontos. Por outro lado, também é bastante comum, principalmente em testes de múltipla escolha, que os itens sejam corrigidos dicotomicamente, ou seja, na forma de “certo” (quando a resposta marcada corresponde ao gabarito) e “errado” (quando se marca qualquer outra opção que não o gabarito).

Aqui, trataremos particularmente do caso dicotômico, embora ele também possa ser estendido à situação mais geral dos acertos parciais. Sendo assim, suponha que uma prova de matemática, constituída de seis itens de múltipla escolha, é feita por 1.000 alunos. Se estamos interessados em mensurar a dificuldade dos itens que compõem essa prova, uma vez realizada a correção dela, poderíamos obter, de uma maneira relativamente fácil e direta, a proporção ou porcentagem de acertos em cada um dos seis itens. A tabela a seguir mostra os resultados obtidos para esse exemplo.

Tabela 3: Percentuais de acerto de estudantes por itens em um teste fictício

Item	N [Acertos (em 1000 alunos)]	Proporção de acertos (N/1000)	%_acertos
1	522	0,522	52,2
2	265	0,265	26,5
3	956	0,956	95,6
4	832	0,832	83,2
5	670	0,67	67
6	118	0,118	11,8

Conforme observamos na tabela, o item 1 foi acertado por 522 dos 1.000 alunos que participaram da avaliação. Portanto, a proporção de acertos desse item foi de 0,522 ($=522/1.000$), sendo o percentual de acerto igual a esse último valor multiplicado por 100. Esse cálculo é então repetido para os demais itens.

Dessa forma, se tivéssemos, ainda que vagamente, que separar os itens quanto à aparente dificuldade – entre baixa, média e alta –, poderíamos talvez dizer que os itens 3 e 4 têm baixa dificuldade, isto é, são fáceis, visto que foram acertados pela maior parte dos alunos, 95,6% e 83,2%, respectivamente.

Os itens 1 e 5 poderiam ser classificados como de dificuldade média – 52,2% e 67% de acerto, respectivamente. Por sua vez, os itens 2 e 6 seriam considerados difíceis, ou de dificuldade alta – 26,5% e 11,8%, de acerto, respectivamente –, visto que, nesses, os acertos foram escassos.

Esse modo de calcular objetivamente a dificuldade de um item tem algumas consideráveis vantagens, como a relativa facilidade de ser realizado e também de ser compreendido. Entretanto, também se trata de um procedimento que guarda alguns inconvenientes. Isso porque as proporções ou porcentagens de acerto, além de dependerem dos próprios itens, igualmente dependem de quem responde a estes.

Digamos, por exemplo, que os dados da Tabela 3 façam referência a um teste de matemática aplicado a estudantes de escolas “de elite” de um determinado município ou estado e que essas escolas possuem um rigoroso processo de seleção para admitir os seus respectivos discentes, os quais, na maioria das vezes, são oriundos de famílias socioeconomicamente mais favorecidas. Por outro lado, se esse mesmo teste for depois aplicado em escolas rurais ou da periferia das cidades, que tradicionalmente atendem a crianças e jovens de famílias socialmente desfavorecidas, é provável que os percentuais de acerto se modifiquem para menos, dando a impressão de que os itens ficaram mais difíceis. Dessa forma, o que era um item de dificuldade baixa, poderia ser então classificado como de dificuldade média, enquanto um de dificuldade média passaria a ser de dificuldade alta, visto que os percentuais de acerto seriam menores.

Sendo assim, a proporção ou porcentagem de acerto de um item é uma medida clássica de sua respectiva dificuldade, portanto, comumente empregada na TCT. Conforme destacamos, trata-se uma medida amplamente utilizada e facilmente compreendida, embora também possua as limitações que acabamos de apontar. E é exatamente devido a limitações como essas que surgiram abordagens alternativas para lidar com esse problema – como a Teoria da Resposta ao Item (TRI), que ainda será apresentada, posteriormente, de forma detalhada.

3.3.3 Discriminação dos itens

Na linguagem comum, a palavra *discriminação* costuma ter conotação negativa, como algo associado a um preconceito racial, social, religioso, entre outros. Entretanto, em psicometria, discriminação tem um sentido mais neutro, pois se refere à capacidade, por exemplo, de um item *discriminar* ou *distinguir* bem os respondentes que possuem maior medida de um certo construto de interesse – por exemplo, os estudantes de maior proficiência numa determinada disciplina – daqueles que têm menos dessa característica – isto é, dos estudantes de menor proficiência.

Quando seguimos por esse caminho, percebemos que o termo *discriminação* passa até a se revestir de uma conotação positiva. Isso porque o termo se refere à capacidade de um item ou de um teste ser *justo*, uma vez que, à medida que o item ou o teste possua uma boa discriminação, ele então consegue distinguir quem sabe mais de quem sabe menos.

Uma maneira clássica de se aferir a capacidade de discriminação de um item dentro de seu respectivo teste é levar em conta, precisamente, o desempenho dos estudantes nessas duas medidas – item e teste. Ou seja, um item terá uma boa discriminação num teste somente se ele tiver uma correlação elevada – ou, pelo menos, razoável – com o próprio teste do qual ele faz parte.

Nesse sentido, um procedimento relativamente simples ajuda a ilustrar a ideia que está por trás disso. Consideremos ainda o exemplo da Tabela 3, em que tínhamos 1.000 estudantes respondendo a um teste com seis questões de matemática. Uma vez corrigido esse teste, é possível calcular a nota de cada aluno. Outro procedimento comum, em casos como esse, é ordenar os resultados dos estudantes, do menor para o maior, o que nos permite obter os percentis da distribuição de resultados, conforme vimos anteriormente.

Mas, uma vez tendo chegado à distribuição das notas de todos os estudantes no teste, uma forma de tentarmos aferir o poder de discriminação de cada item é observar como estão se comportando as taxas de acerto nos diferentes itens, para, pelo menos, dois grupos de alunos: (1) o grupo dos que alcançaram os melhores resultados gerais no teste inteiro; e (2) o grupo dos que tiraram os piores resultados no teste inteiro.

O raciocínio por trás disso é simples: se dividimos, por exemplo, os 1.000 estudantes em duas metades, sendo o grupo 1 com os 500 que obtiveram as menores notas no teste e o grupo 2 com os outros 500 alunos que alcançaram as maiores notas, podemos, para cada item que compõe o teste, observar a sua respectiva taxa de acerto em cada um desses grupos.

Suponhamos, então, que realizamos todo esse percurso, obtendo os resultados mostrados na Tabela 4 a seguir.

Tabela 4: Percentual de acerto de itens entre grupos de estudantes

Item	Nº de acertos por grupo de alunos			Proporção de acertos por grupo			Percentual de acertos por grupo			Dif. de %
	1	2	Total	1	2	Total	1	2	Total	
1	52	470	522	0,104	0,94	0,522	10,4	94	52,2	83,6
2	89	176	265	0,178	0,352	0,265	17,8	35,2	26,5	17,4
3	458	498	956	0,916	0,996	0,956	91,6	99,6	95,6	8
4	348	484	832	0,696	0,968	0,832	69,6	96,8	83,2	27,2
5	189	481	670	0,378	0,962	0,67	37,8	96,2	67	58,4
6	36	82	118	0,072	0,164	0,118	7,2	16,4	11,8	9,2

Para entendermos melhor o que a tabela nos informa, consideremos, por exemplo, o item 1. Lembre-se de que, agora, temos 500 alunos no grupo 1, correspondente às menores notas no teste, e outros 500 alunos no grupo 2, correspondente às maiores notas.

Observemos que o item 1 foi acertado por apenas 52 alunos do grupo 1 e por 470 alunos do grupo 2. O total de acertos nesse item é, naturalmente, a soma desses dois totais parciais, de modo que chegamos a 522 acertos, conforme já havíamos mostrado na tabela anterior.

Porém, uma vez conhecendo os totais parciais de acerto em cada grupo, fica fácil obter as suas respectivas proporções de acerto: basta dividir o número de acertos por 500, que é o número de alunos em cada grupo. Fazendo isso, constatamos que essas proporções são, respectivamente, 0,104 e 0,94. Por sua vez, a proporção total, de 0,522, é obtida dividindo o número total de acertos, que é 522, pelo número total de alunos (1.000), conforme já havíamos feito na tabela anterior.

Outra forma de considerar esses mesmos dados descritos é multiplicar por 100 as proporções já obtidas, a fim de obter os valores percentuais, que estão nas colunas mais à direita na Tabela 4.

Por fim, a última coluna (Dif. de %) mostra simplesmente o resultado obtido quando se subtrai o percentual de acerto no grupo 1 do percentual no grupo 2. Ou seja, esse valor corresponde à diferença, cujo cálculo se dá da seguinte forma: "% de acerto no grupo 2" – "% de acerto no grupo 1".

Em relação, especificamente, ao item 1, percebemos que o percentual de acerto no grupo de alunos com as notas mais baixas é igual a 10,4%, enquanto no grupo de alunos com as notas mais altas, o percentual de acerto corresponde a 94%. Sendo assim, a diferença entre os dois é, portanto, 83,6%.

Tudo isso aponta para o fato de que o item 1 possui um elevado poder de discriminação. Ou seja, o item 1 consegue, com bastante sucesso, distinguir os estudantes com maior proficiência daquelas com menor proficiência, visto que tende a ser acertado principalmente por quem alcançou os melhores resultados no teste.

Por outro lado, consideremos agora o item 3. Conforme já havíamos visto, trata-se do item mais fácil do teste, pois foi acertado por 956 alunos, de um total de 1.000. Mas, se quase todo mundo acertou esse item, então ele foi quase que inteiramente acertado tanto no grupo 1 quanto no grupo 2 de estudantes, o que podemos confirmar a partir dos percentuais de acerto nesses dois grupos, que correspondem a 91,6% e 99,6%, respectivamente.

Dessa forma, subtraindo um desses percentuais do outro, chegamos tão somente a uma diferença de 8%. Em outras palavras, podemos dizer que o item 3 apresenta um baixo poder de discriminação, logo, não trará muitas informações na hora da análise dos resultados.

Outro item bastante ilustrativo nessa tabela é o de nº 6. Trata-se do item mais difícil de todos nesse exemplo, visto que o seu percentual total de acerto é de somente 11,8%, o menor do conjunto de seis itens. Observando os dados da tabela, vemos que o item 6 foi muito pouco acertado tanto no grupo 1 quanto no grupo 2. Dessa forma, como os dois percentuais parciais são muito próximos entre si, sendo igualmente baixos, a diferença entre ambos é também bastante pequena, de modo a valer apenas 9,2%.

Resultados assim nos permitem chegar a uma conclusão importante sobre o poder de discriminação dos itens: tanto quando são muito difíceis para o grupo de sujeitos ao qual são aplicados, como no caso contrário, quando são muito fáceis, a discriminação de um item tende a ser baixa. Isso porque todo mundo, independentemente de ter uma proficiência maior ou menor, acaba tendo um desempenho similar em relação a eles. Ou seja, **itens extremamente difíceis são resolvidos de forma errada por quase todo mundo, não apenas pelos estudantes de proficiência menor, mas também pelos de proficiência maior. No outro extremo, itens extremamente fáceis tendem a ser acertados tanto pelos estudantes de maior proficiência quanto pelos de menor proficiência.** Sendo assim, tanto num caso quanto no outro, com base no resultado

obtido nesse item específico, não conseguimos distinguir bem o grupo de melhor desempenho do grupo de pior desempenho,.

Por outro lado, os itens de maior poder de discriminação são precisamente aqueles cujos percentuais globais de acerto situam-se próximos dos 50%, como é o caso do item 1 da Tabela 4, que tinha um percentual total de acerto igual a 52,2%. Em casos assim, o item tende a “separar” ou “discriminar” melhor os dois grupos de estudantes, de modo a ser mais acertado no de maior proficiência em relação ao outro.

3.4 A fidedignidade dos testes de proficiência

A fidedignidade de um teste relaciona-se com a sua capacidade de produzir resultados consistentes, quando ele tenta mensurar o construto ou dimensão para o qual foi concebido.

A seguir, apresentamos dois exemplos mais “concretos” para ilustrar melhor a questão da fidedignidade de um teste.

Exemplo 1

Em uma avaliação externa de um estado brasileiro, com resultados calculados na escala do Saeb, os estudantes tiveram de realizar um teste de língua portuguesa, para o qual havia quatro cadernos diferentes, identificados pelas cores amarelo, azul, cinza e rosa. Nessa edição, havia questões diferentes entre os cadernos de cores diferentes. Suponha que Paulo tenha feito o caderno cinza e alcançado 265 de proficiência. Caso tivesse respondido a um caderno de outra cor, Paulo teria obtido essa mesma nota?

Exemplo 2

Maria fez uma redação em sua escola, que foi corrigida pela professora Vera, de língua portuguesa, a qual lhe atribuiu nota 8. Caso a sua redação tivesse sido corrigida pela professora Marta, que leciona a mesma disciplina em outra turma da escola, a sua nota teria sido a mesma?

Exemplos assim dizem respeito, precisamente, à questão da fidedignidade de um teste. Como é possível perceber, a fidedignidade é uma questão bastante relevante, visto que testes produzidos e/ou corrigidos de maneira não fidedigna podem acarretar consideráveis imprecisões – ou até mesmo injustiças – na determinação dos seus respectivos resultados.

No exemplo 1, se Paulo tirasse, ao menos aproximadamente, a mesma nota na prova de língua portuguesa, independentemente da cor do caderno de itens, poderíamos dizer que a avaliação aplicada tem, aparentemente, elevada fidedignidade. Já no exemplo 2, analogamente, também poderíamos dizer que a correção seria fidedigna caso as notas atribuídas pelas duas professoras, cujo trabalho é independente, fossem iguais ou semelhantes.

Porém, em geral, não é possível verificar a fidedignidade tal qual indicamos. O estudante que participa da avaliação externa responde a um único caderno de questões. E, principalmente no contexto escolar, a redação de um aluno, muitas vezes, é corrigida por apenas um professor, e não por dois ou mais.

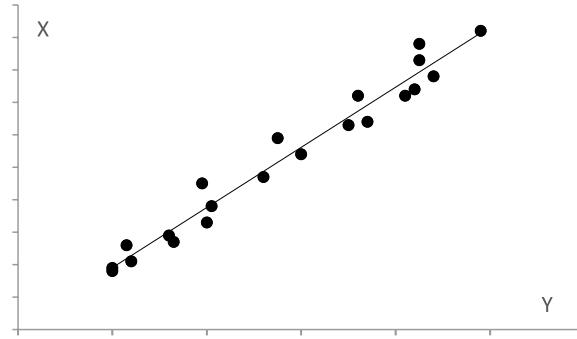
Portanto, é preciso conceber maneiras de lidar com a fidedignidade de um teste por meio de procedimentos padronizados e objetivos. E, nesse ponto, a estatística, mais uma vez, vem nos auxiliar. Um instrumento poderoso é o estudo das chamadas distribuições bivariadas, que abordaremos sucintamente a seguir.

3.4.1 Distribuições bivariadas

Em estatística, de um modo geral, muitas vezes interessa-nos saber como duas variáveis se comportam quando são consideradas ao mesmo tempo. Isso nos leva, precisamente, a definir as chamadas distribuições bivariadas.

Suponha que, mediante a aplicação de questionários e provas a um grupo de estudantes, tenha sido possível mensurar, para cada um deles, o seu respectivo índice socioeconômico (X) e também a sua proficiência em matemática (Y). Uma forma de representar esses resultados simultaneamente seria fazer um gráfico com dois eixos, X e Y, em que cada ponto representado corresponderia a um estudante, de acordo com os valores para ambas as variáveis. Suponha, então, que o resultado disso tenha sido o seguinte:

Gráfico 1: Relação positiva entre duas variáveis

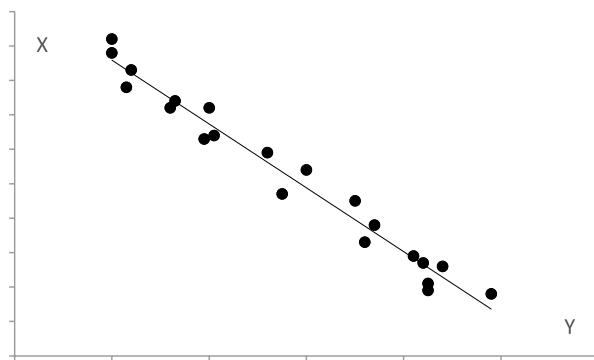


Nesse caso, percebemos claramente que os pontos se distribuem ao longo de uma linha diagonal – hipotética e, aqui, traçada –, de tal modo que, quanto maiores os valores de X ou da condição socioeconômica – representados nos pontos mais à direita no gráfico –, maiores também tendem a ser os valores de Y ou da proficiência em matemática – representados nos pontos mais acima no gráfico.

Em casos assim, a reta que sintetiza o comportamento geral dos casos apresenta uma clara inclinação para cima, ou positiva. *Dizemos então que há uma associação positiva entre essas duas variáveis, pois o aumento de uma delas tende a se fazer acompanhar de um aumento também da outra.* Quando essa associação simultânea se dá de uma forma ao menos aproximadamente linear, dizemos que existe uma considerável associação linear entre as variáveis.

Às vezes, as associações lineares existem e são fortes, sendo, porém, negativas – isto é, o aumento em uma determinada variável se faz acompanhar de um decréscimo na outra. Um exemplo disso é dado no gráfico a seguir.

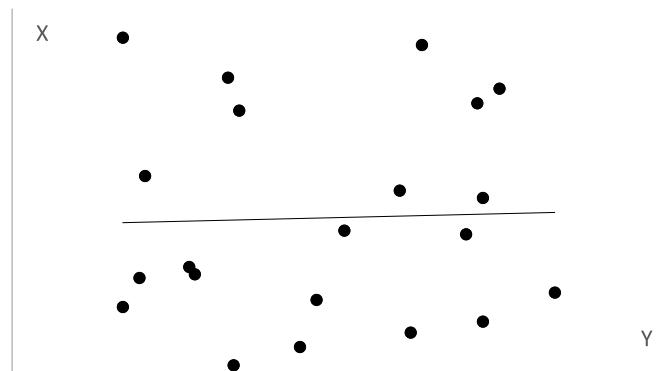
Gráfico 2: Relação negativa entre duas variáveis



Um caso como esse poderia corresponder a estudantes, representados pelos pontos do gráfico, classificados de acordo com o número de horas de estudo de um determinado tópico no eixo X, horizontal, e a quantidade de erros cometidos sobre esse mesmo tópico numa prova, no eixo Y, vertical. É possível observar que os maiores valores de horas de estudo, presentes na parte mais à direita no gráfico, se fazem acompanhar dos menores valores de erros cometidos nesse tópico, presentes na parte mais abaixo no gráfico. Nesse caso, dizemos que as *duas variáveis em questão estão linear e negativamente associadas*.

Por outro lado, também pode haver duas variáveis que não parecem possuir uma forte associação entre si, nem no sentido positivo, nem no negativo. Imagine, por exemplo, um gráfico em que o eixo X representa o número de quilômetros que separa a casa dos estudantes até a escola onde eles estudam e o Y representa a nota dos alunos em química:

Gráfico 3: Relação não significativa entre duas variáveis



O fato de o aluno morar mais perto da escola – pontos mais à esquerda – ou mais longe da escola – pontos mais à direita no gráfico – em nada parece se correlacionar com a nota tirada em química. Dessa forma, a reta que melhor sintetiza a associação entre essas duas variáveis é horizontal (ou seja, ela não sobe nem desce). *Em casos assim, dizemos que as duas variáveis não se encontram significativamente associadas entre si*.

Uma maneira matemática de se medir a associação entre duas variáveis é através do chamado coeficiente de correlação r de Pearson. Não nos deteremos aqui nos detalhes de seu cálculo, mas sim na respectiva interpretação.

O coeficiente r , teoricamente, pode variar entre -1 e 1, sendo que, nesses extremos, a correlação é perfeita. Isto é, caso representássemos os casos num gráfico de modo análogo ao que fizemos nos dois primeiros exemplos, os resultados seriam tais que

os pontos ficariam todos exatamente ao longo de uma reta ascendente (para $r=1$) ou descendente (para $r=-1$). Por outro lado, em uma situação como a do Gráfico 3, r seria igual a zero ou a um valor muito próximo disso, visto que não há associação significativa entre a distância entre a casa dos estudantes e a escola em relação ao desempenho em química – como podemos ver, a reta não sobe e nem desce.

Importante ressaltar que, na prática, sempre existe uma certa dispersão dos pontos em torno da reta que sintetiza a associação, como em todas as ilustrações apresentadas. Dessa forma, em situações reais, o valor de r nunca chega a 1 ou -1, podendo, não obstante, aproximar-se bastante disso.

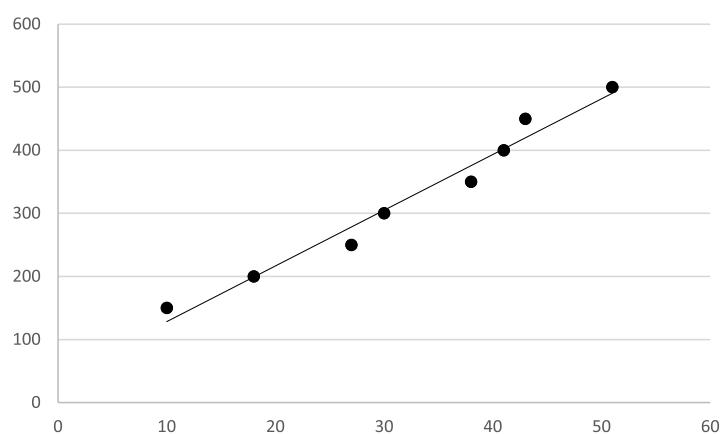
O estudo das correlações é importante para as considerações sobre a fidedignidade de um teste porque *um teste fidedigno deve, necessariamente, possuir uma elevada correlação com o construto que ele se propõe a mensurar.*

Vamos a um exemplo?

Suponha que, de algum modo, tenha sido possível estimar o conhecimento de física de um grupo de estudantes, sendo que, para cada um deles, temos uma medida confiável de sua respectiva proficiência nessa disciplina.

Essa medida é apresentada na escala vertical do Gráfico 4 a seguir, dentro de uma escala de algumas centenas de pontos. Por outro lado, temos um Teste A de física, que produz notas em uma nova escala, de algumas dezenas de pontos, sendo que tais resultados correspondem ao eixo horizontal do mesmo gráfico.

Gráfico 4: Relação entre duas medidas fidedignas



Ora, numa situação como essa, os dados parecem apontar para uma elevada correlação entre os resultados do Teste A e a proficiência dos alunos. Isso, por sua vez, permite-nos concluir que o Teste A possui uma elevada fidedignidade para se mensurar a proficiência em física dos alunos.

Se dispuséssemos também de um Teste B de física, que tivesse igualmente esse bom comportamento, isto é, esse elevado nível de fidedignidade, então qualquer que fosse o teste a ser aplicado a um aluno ou a um grupo deles – o A ou o B –, ficaríamos confiantes de que os resultados seriam semelhantes, ressalvadas transformações de escalas entre os dois testes. Ou seja, os estudantes que tivessem um bom desempenho no Teste A possivelmente também se sairiam bem no Teste B, e vice-versa, embora não seja preciso que todos os alunos façam ambos os testes para sabermos disso.

3.4.2 Fatores que afetam a fidedignidade de um teste

Há diversos fatores que podem afetar a fidedignidade dos instrumentos avaliativos. Conhecer esses fatores é importante para elaboradores de itens e de testes, pois, na maioria dos casos, é possível modificar ou corrigir as características dos itens, de modo a aumentar a fidedignidade do teste.

Apresentaremos aqui alguns dos fatores que costumam ser considerados mais relevantes.

Número de itens e conteúdo dos testes

Em geral, a fidedignidade aumenta com o número de itens de um teste. Se desejamos mensurar a proficiência em matemática de um grupo de estudantes, um teste com 30 itens possivelmente será mais fidedigno do que um teste de 10 ou 5 itens – se mantidas inalteradas as demais circunstâncias ou características.

Isso porque um teste muito curto não consegue, em geral, cobrir um espectro muito grande de habilidades ou conhecimentos, o que é particularmente verdadeiro no campo da avaliação educacional. Dessa forma, se um determinado estudante domina apenas parte do conteúdo de uma certa disciplina e participa de um teste sobre ela com poucos itens, sua nota poderá variar consideravelmente a depender de a prova avaliar mais ou menos as habilidades que o estudante domina.

Por outro lado, testes mais longos costumam ser mais abrangentes, de modo que essa dependência relacionada à sorte tende diminuir, visto que todos ou quase todos os tópicos de interesse possivelmente serão cobertos num único teste. Portanto, a quantidade de itens se associa também à chamada *fidedignidade de conteúdo* do teste.

Correção dos testes

Outro fator crucial para a determinação da maior ou menor fidedignidade de um teste é o modo como a correção é feita. Isso porque dois corretores, avaliando um mesmo item ou teste, podem chegar a conclusões diferentes acerca de seu respectivo valor, por diversos motivos. Desse modo, quando os itens são de resposta construída, por exemplo, para se garantir uma maior fidedignidade, é preciso observar uma série de procedimentos, conforme já destacamos no Capítulo 3, como criar uma chave de correção, com gabarito, para cada item; treinar os corretores; fazer com que uma mesma prova seja corrigida por dois ou mais corretores, para uma posterior comparação etc.

Procedimentos como esses, embora muito úteis e necessários, costumam também ser caros e demorados. Esse é um dos motivos pelos quais muitas das avaliações em larga escala evitam itens de resposta construída, privilegiando, em seu lugar, os de múltipla escolha, que podem ser corrigidos eletronicamente, com mais rapidez e precisão. Nos últimos tempos, porém, com avanços na tecnologia, principalmente no campo da inteligência artificial, possibilidades de correção eficiente de itens de resposta construída têm surgido e podem vir a mudar esse cenário.

Aplicação do teste

Os resultados de um teste podem variar de acordo com fatores associados à sua aplicação, como o tempo para que os itens sejam respondidos, a presença ou não de instruções claras para se fazer o teste, a possibilidade de se usar ou não equipamentos diversos – como calculadoras ou textos de consulta –, a existência ou não de condições adequadas para a testagem – como um ambiente silencioso e confortável. Por tudo isso, um importante aspecto na aplicação de um teste é a sua *padronização*, a partir da qual se garante que a sua realização seguirá determinados protocolos, de modo a se assegurar que todos os examinados, na medida do possível, respondam ao teste em condições de igualdade.

Condições pessoais

Por melhores que sejam os procedimentos de padronização para a correta aplicação de uma avaliação, ainda restam fatores que dependem basicamente do estudante. Uma pessoa pode realizar um teste, por exemplo, depois de ter tido uma boa noite de sono, ou depois de ter passado a noite em claro cuidando de um parente enfermo; também pode estar muito bem de saúde, ou padecendo de uma gripe. Naturalmente, isso pode fazer com que a mesma pessoa obtenha resultados diferentes para o mesmo conjunto de itens, caso faça o mesmo teste em dias ou condições diferentes. Problemas assim, naturalmente, fogem ao controle de quem aplica a avaliação. Porém, isso tampouco impede, por outro lado, que essas questões também possam ser consideradas por aqueles que se debruçam sobre o tema das avaliações.

3.4.3 A validade dos testes de proficiência

Tão importante quanto tratarmos da fidedignidade é tratarmos da validade de um teste. A validade relaciona-se à capacidade de o teste mensurar, de fato, aquilo que ele se propõe a medir.

Como um exemplo disso – ou melhor, um contraexemplo –, digamos que tenha sido aplicado um teste de matemática com itens contendo enunciados relativamente complexos, nos quais aparecem palavras cujo significado possa ser desconhecido para boa parcela dos estudantes avaliados. Em casos assim, é possível que alguns alunos tirem notas baixas. Porém, a razão disso, ao menos em parte, pode não ser necessariamente a sua baixa proficiência em matemática – que deveria ser o foco da avaliação –, mas sim dificuldades de leitura e interpretação de texto. Portanto, podemos dizer que, em um caso como esse, a validade do teste de matemática ficou comprometida.

Na verdade, há diversos tipos de validade de um teste, sendo que alguns dos mais importantes são a *validade de conteúdo* e a *validade instrucional*.

A validade de conteúdo foi ilustrada no exemplo apresentado. E, como facilmente percebemos, ela é de extrema importância na avaliação educacional. Tal qual apresentamos no Capítulo 2, há uma série de diretrizes relacionadas ao conteúdo de um item que devem ser levadas em conta na hora de sua elaboração. Uma delas é deixar o enunciado o mais claro e objetivo possível, evitando que elementos possam desviar o foco da avaliação para algo que não seja, precisamente, o objetivo do teste.

Por outro lado, a validade instrucional relaciona-se ao fato de o conteúdo do teste ter sido ou não ministrado àqueles que o farão. Trata-se igualmente de um assunto relevante em avaliações em larga escala, em que é preciso garantir que todos os estudantes tenham recebido aulas sobre todos os tópicos constantes da matriz de referência utilizada para a construção do teste. Por outro lado, caso haja uma variação na rede ou no sistema de ensino quanto à apresentação de determinados tópicos curriculares – com escolas ensinando conteúdos diferentes ou em momentos distintos –, naturalmente o teste terá problemas de validade instrucional.

Ainda sobre a validade, cabe ressaltar a sua respectiva dependência da fidedignidade. Isto porque, se um teste não for fidedigno – por um ou mais de um dos motivos apresentados anteriormente –, tampouco será válido. O inverso, por sua vez, não necessariamente acontece. Um teste pode ser válido – caso, por exemplo, tenha um conteúdo adequado –, porém, pode não ser fidedigno – caso, por exemplo, as condições ou protocolos de testagem não sejam padronizadas.

3.5 O uso da Teoria da Resposta ao Item na interpretação dos resultados das avaliações

Conforme já adiantamos, a Teoria da Resposta ao Item (TRI) adota uma abordagem conceitual diferente da empregada pela Teoria Clássica dos Testes (TCT). A TRI se desenvolve a partir da ideia das escalas cumulativas, porém ainda sob o ponto de vista clássico, por meio da chamada escala de Guttman.

3.5.1 Escalas cumulativas: o caso da escala de Guttman

Uma diferença fundamental entre uma escala somativa e uma cumulativa é: enquanto a primeira se atenta à soma dos resultados das diversas perguntas que compõem um questionário ou teste, a segunda se preocupa mais com a maneira como essas respostas se acumulam, ao longo de um determinado *continuum* que mensura, precisamente, o construto-alvo do questionário ou do teste. Passemos para um exemplo, a fim de tornar mais clara a nossa percepção sobre o assunto.

Suponha que alguém elabore um pequeno teste para mensurar a proficiência em matemática de um grupo de estudantes, conforme indicado a seguir. Na Tabela 5, temos cinco itens ordenados por nível de dificuldade, em ordem crescente. Para cada um, o estudante precisa calcular o valor da incógnita a .

Tabela 5: Itens ordenados por nível de dificuldade

Item (nº)	1	2	3	4	5
Item	$a = 4 + 7$	$a = 3 + 2 \times 5$	$a^2 - a = 0$	$a = \log(1000 \times 100) - \log 1$	$a = d(x^2 + 9x - 4)/dx; x = 3$
Pedro	1	0	0	0	0
Paula	1	1	1	0	0
Francisco	1	1	1	1	1

Como se pode verificar, o item 1 diz respeito a um tipo de operação aritmética, a soma; o item 2 envolve dois tipos de operações básicas, multiplicação e soma; o item 3 apresenta uma equação de 2º grau; o item 4 traz um problema de logaritmo; e o item 5, um problema de derivadas.

Para a nossa discussão, não é necessária a resolução dos problemas matemáticos, pois não estamos interessados nas respostas propriamente ditas, mas sim no seu respectivo padrão quanto ao acerto ou erro. Nesse exemplo, as respostas já foram devidamente corrigidas, sendo adotada a seguinte convenção: 1 = acerto; 0 = erro.

Antes de prosseguirmos na exploração dos dados, segundo a perspectiva da escala de Guttman, cabe fazer duas observações com base no que já vimos até agora sobre alguns procedimentos psicométricos comuns à TCT.

Um deles diz respeito ao fato de que, num exemplo como esse, também é possível adotar o procedimento somativo para calcular os escores dos indivíduos. Nesse caso, bastaria obter a quantidade de vezes em que o avaliado acerta o item, ou então calcular o seu respectivo percentual de acerto, de modo a obter, assim, uma medida de proficiência do aluno.

Observando a tabela, vemos que Pedro acertou uma questão em cinco, ou seja, alcançou 20% de acerto. Paula, por sua vez, apresentou três respostas corretas, das cinco possíveis, acertando 60%. Já Francisco respondeu corretamente a todos os itens, isto é, 100% do teste.

Uma segunda observação é considerar os totais de coluna, de modo que, no exemplo, possamos observar como cada item do teste está se comportando, a fim de encontrarmos o seu respectivo grau de dificuldade. Dessa forma, percebemos que os itens para os quais houve as maiores quantidades de acertos – sinalizados com o número 1 – são aqueles que têm a maior soma. O primeiro item, portanto, é o mais fácil de ser respondido, já que, nesse exemplo, todos o acertaram. Ou seja, os itens cujas medidas de dificuldade são as maiores são aqueles que possuem a menor soma de acertos.

Sendo assim, observando como os itens se comportam em relação à sua respectiva dificuldade, é possível perceber que, à medida que caminhamos para a direita na tabela, os itens vão se tornando mais difíceis, visto que cada vez menos pessoas tendem a acertá-los. Com isso, percebemos que todos os estudantes, no exemplo apresentado, sabem quanto é $4+7$, no item 1, enquanto apenas um deles respondeu corretamente o problema de derivadas, no item 5. Esse último item, dessa forma, é considerado o mais difícil; e o primeiro, o mais fácil de todos os apresentados.

Feitas essas considerações, passemos a abordar esse problema do ponto de vista da escala de Guttman. De acordo com ela, se quisermos avaliar os indivíduos quanto ao construto-alvo adotado, não precisamos somar os resultados de suas respectivas respostas; basta nos atentarmos a um único ponto da escala: aquele no qual o indivíduo muda o padrão de suas respostas, deixando, por exemplo, de acertar e passando a errar os itens.

Na Tabela 5, observamos que Paula acertou a equação de 2º grau, mas errou o problema de logaritmo. É previsível, então, que Paula erre itens mais difíceis do que esse último, como o problema de derivadas. Por outro lado, como Paula acertou a equação de 2º grau, é de se esperar que ela também acerte itens mais fáceis, como os de operações aritméticas, nos itens 1 e 2.

Os demais estudantes também se comportam segundo essa lógica. Contudo, cabe observar um ponto relevante: para que uma escala de Guttman funcione, ela precisa ser consistente, significando que, para um determinado indivíduo, os acertos fiquem no lado mais fácil e os erros, no mais difícil. O limite entre esses dois tipos de resposta – se mais para o lado do fácil ou do difícil – estará associado ao escore do indivíduo.

Por outro lado, haverá inconsistência nessa escala se, por exemplo, um determinado indivíduo errar uma pergunta fácil ao mesmo tempo que acerta uma pergunta considerada difícil.

No nosso exemplo, a consistência desse caso se manifesta no fato de que, quando olhamos o padrão de respostas, podemos pensar numa linha diagonal imaginária, dividindo o quadro de respostas em duas partes: uma inferior à esquerda, apenas formada de números 1 – acertos –, e outra superior à direita, apenas formada de números 0 – erros. Entretanto, se houvesse inconsistências, teríamos casos com um ou mais números 1 dentro da região dos números 0, e vice-versa.

A maior ou menor observância desse padrão associa-se ao chamado problema de consistência interna entre os itens, condição fundamental para o bom funcionamento e para a própria existência de uma escala de Guttman.

Por essa razão, utilizamos uma medida para as escalas de Guttman chamada de *replicabilidade*, responsável por mensurar precisamente o nível de consistência. Geralmente, adotamos escalas de Guttman com replicabilidade de, no mínimo, 90%, isto é, na qual há um máximo de 10% de discordância entre os valores previstos pelo modelo e os realmente obtidos.

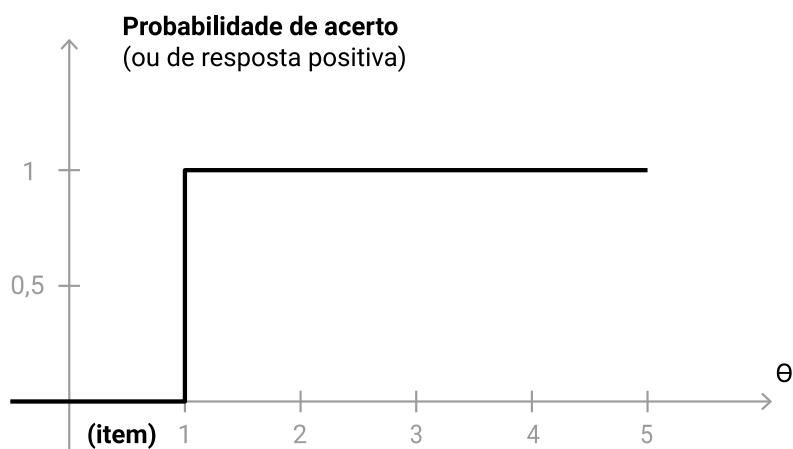
No exemplo, outro ponto de relevância é o fato de que a proficiência do estudante pode ser numericamente estimada como correspondendo ao ponto da escala em que ele deixa de acertar os itens e passa a errá-los. Com esse raciocínio, podemos, então, obter, para cada aluno, as medidas de proficiência segundo a escala de Guttman.

3.5.2 Função degrau unitário: definição e relevância para a TRI

Os itens de uma escala de Guttman podem ser graficamente representados segundo o que denominamos *função degrau unitário*. Trata-se de um gráfico com dois eixos, em que o horizontal expressa a dificuldade do item e o vertical, a probabilidade de um sujeito acertá-lo.

Vejamos, então, como é apresentada a função para o item mais fácil da escala de proficiência em matemática, com base no exemplo anterior.

Gráfico 5: Probabilidade de acerto do item 1



Poderíamos interpretar esse gráfico da seguinte forma:

O eixo horizontal mensura um construto – genericamente identificado com a letra grega teta (θ) – que, nesse caso, corresponde a uma escala de proficiência em matemática. Dessa forma, quanto mais um aluno acertar itens mais difíceis, mais ele possui dessa medida θ e, consequentemente, mais à direita ele se encontra no eixo horizontal.

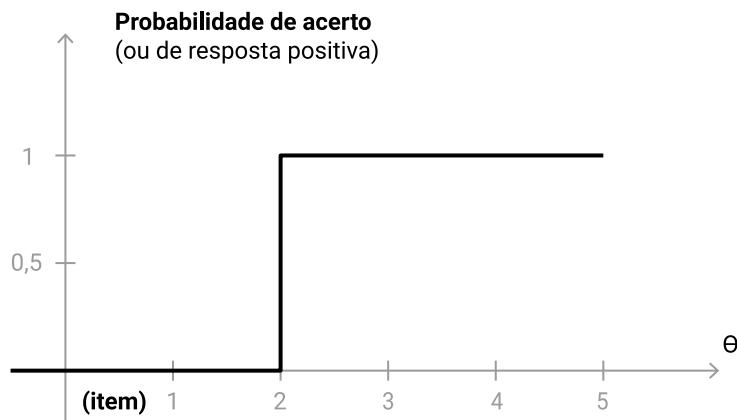
O eixo horizontal, além de permitir que nele encaixemos os indivíduos, também permite que ali situemos os próprios itens do teste, segundo o seu relativo grau de dificuldade. Dessa forma, o gráfico mostra que o item 1 do teste, que solicita ao estudante a resolução da soma $4+7$, tem a sua inclinação, ou “degrau”, mais à esquerda do que todos os demais itens do teste.

A interpretação disso é simples: como o item 1 é o mais fácil do teste, ele também é o item que, para ser acertado, demanda dos indivíduos menos do construto-alvo do teste, isto é, menos proficiência em matemática. Em outras palavras, indivíduos com relativa baixa proficiência em matemática, situados mais à esquerda no gráfico, terão, ainda assim, um grau suficiente de proficiência para acertarem esse item.

Por outro lado, o eixo vertical indica a probabilidade de o respondente acertar um item. Conforme percebemos, na função degrau unitário, há uma espécie de “tudo ou nada”: se o indivíduo não tem o suficiente de um determinado construto, a sua chance de acertar o item em questão é nenhuma – probabilidade igual a 0. Por outro lado, se ele tiver, ao menos, o valor mínimo suficiente do construto para acertar o item, então ele seguramente o fará – probabilidade igual a 1.

À medida que os itens vão ficando mais difíceis, os seus respectivos degraus vão se deslocando mais à direita do gráfico. Apresentamos, a seguir, o degrau correspondente ao próximo item.

Gráfico 6: Probabilidade de acerto do item 2



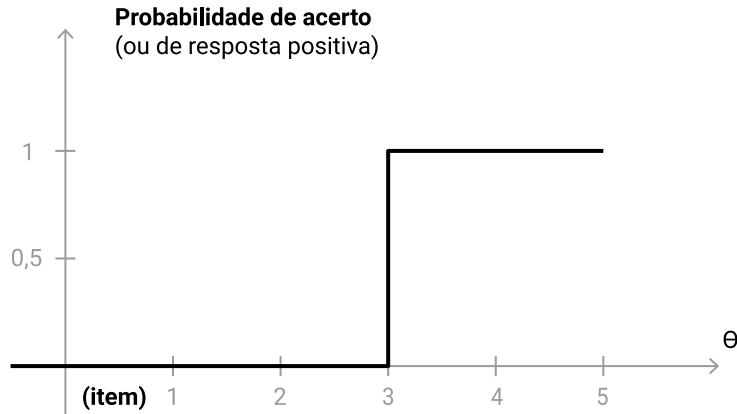
Observe, nesse exemplo, que o degrau se deslocou para a direita, em relação ao Gráfico 5. Isso quer dizer que, para esse item ser acertado, é preciso que o respondente tenha uma maior proficiência em matemática do que no caso anterior – ou seja, é preciso que o seu respectivo valor de θ ande mais para a direita no gráfico, em relação ao primeiro caso.

3.5.3 Modelo de um parâmetro ou de Rasch

Até aqui, compreendemos que, no âmbito da Teoria Clássica dos Testes (TCT), é possível criar uma escala cumulativa do tipo Guttman, para nela dispormos os resultados de um teste mensurando um determinado construto-alvo. Também vimos que uma representação gráfica disso poderia se apresentar na forma da função degrau unitário.

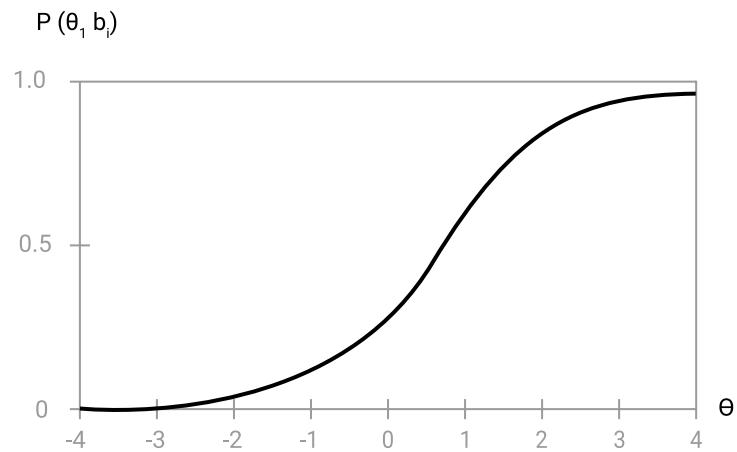
Agora, vamos observar a função degrau unitário para um dado item de um teste de matemática. No Gráfico 7, a função representa, especificamente, o comportamento do item 3 no exemplo dado anteriormente: uma equação de 2º grau, de tal modo que o aluno poderia acertá-la ou não.

Gráfico 7: Probabilidade de acerto do item 3



Em seguida, observe a chamada *curva de Rasch*, que é utilizada para representar o *modelo de um parâmetro* da TRI.

Gráfico 8: Curva de Rasch do item 3



Antes de analisarmos as diferenças entre essas duas representações, vejamos o que elas têm em comum.

Semelhanças entre a função degrau unitário e a curva de Rasch

Em ambas as representações, o eixo horizontal expressa a medida de dificuldade dos itens e, ao mesmo tempo, a medida de proficiência ou habilidade dos indivíduos avaliados.

Por “proficiência” ou “habilidade”, referimo-nos ao chamado “traço latente”, ou, nas representações aqui adotadas, ao theta (θ), como medida do construto-alvo.

No exemplo que apresentamos, o traço latente é a proficiência em matemática; já em um questionário de atitudes, o traço latente pode ser o grau de aceitação – por exemplo, se favorável a determinado posicionamento político ou comportamental, entre uma infinidade de outras possibilidades.

Nos dois gráficos, quanto mais caminhamos da esquerda para a direita, maior fica a medida de proficiência ou habilidade. Em outras palavras, quanto mais, no eixo horizontal, avançamos para a direita, mais o item é difícil e mais o indivíduo é proficiente.

Outra semelhança entre as representações está no eixo vertical, que corresponde à probabilidade de o indivíduo acertar um determinado item. Essa probabilidade, por sua vez, encontra-se em um intervalo de 0 (nenhuma chance de acerto) a 1 (100% de chance de acerto).

Entretanto, ainda que façamos as considerações sobre semelhanças da função de grau unitário e da curva de Rasch, precisamos destacar algumas diferenças relevantes entre as duas representações.

Diferenças entre a função de grau unitário e a curva de Rasch

No caso da função de grau unitário, correspondente à escala de Guttman, só há dois valores possíveis para a probabilidade de acerto: 0 ou 1.

Dessa forma, a escala de Guttman é *determinística*: Se o indivíduo possuiu, quanto a uma dada habilidade, menos do que deveria para acertar o item, então ele certamente o “errará” ($P=0$, sendo P a probabilidade de acerto). Por outro lado, se o indivíduo tem uma habilidade que é minimamente suficiente para acertar o item, então ele certamente o acertará ($P=1$).

Por sua vez, no caso da curva de Rasch, não existem apenas dois valores possíveis para a probabilidade de acerto. Primeiramente, é possível perceber que esses dois extremos ($P=0$ ou $P=1$), teoricamente, são inatingíveis. Isso porque, à medida que se desloca mais para a esquerda, a curva vai abaixando cada vez mais, porém sem nunca tocar o eixo horizontal, onde a probabilidade de acerto P seria nula. Por outro lado, avançando-se para a direita, a curva vai subindo à medida que se aproxima da altura $P=1$, mas sem nunca atingir esse limite.

Sendo assim, na curva de Rasch, os valores $P=0$ e $P=1$ correspondem a duas assíntotas da função representada. Essa função, por sua vez, graficamente é chamada de *curva característica do item*, ou CCI.

Aparentemente, essa diferença quanto ao fato de a função incluir ou não os valores 0 e 1 para a probabilidade de acerto poderia soar como um detalhe sem grande importância. Porém, há uma diferença fundamental entre os dois casos: na situação tratada pelo modelo de um parâmetro da TRI, teoricamente, um examinando sempre tem uma probabilidade de acertar um item, por menor que seja a sua habilidade. E, por outro lado, a sua probabilidade de acertar um item nunca é igual a 100%, por maior que também seja sua habilidade. Por isso, a abordagem da TRI é chamada de *probabilística*.

No caso da TRI, a probabilidade de acerto, na verdade, assume valores fracionários entre 0 e 1, como 0,12 (ou 12%), 0,50 (ou 50%), 0,87 (87%) entre, literalmente, uma infinidade de valores possíveis.

Sobre a Curva Característica do Item (CCI)

Podemos dizer que uma CCI deve ter duas características adicionais: (1) ser monotônica, isto é, não ter “picos” nem “vales”, que, caso existissem, implicariam pontos de inflexão; e (2) ser não decrescente, isto é, à medida que se caminha para a direita, na direção das maiores proficiências ou habilidades, a probabilidade de se acertar o item tende a aumentar, ou, na pior das hipóteses, a se manter constante, porém, jamais a diminuir.

Outra diferença importante entre as duas representações é a sua respectiva inclinação. No caso da função degrau unitário, a inclinação é infinita, porque o degrau é vertical. É precisamente devido a isso que podemos verificar o aspecto binário dessa função: no limiar entre seus dois extremos, a probabilidade P passa, subitamente, de 0 para 1.

Por outro lado, no caso da curva de Rasch, percebemos que, além da inclinação não ser infinita em nenhum ponto, ela é, na verdade, variável: mais à esquerda no gráfico, a curva começa subindo pouco; depois, aproximadamente do meio do gráfico, ela sofre uma inflexão, tornando-se mais inclinada; e, finalmente, ao se aproximar da direita do gráfico, volta a se inclinar menos. Esse comportamento, por sua vez, também possui importantes implicações psicométricas, conforme veremos mais à frente.

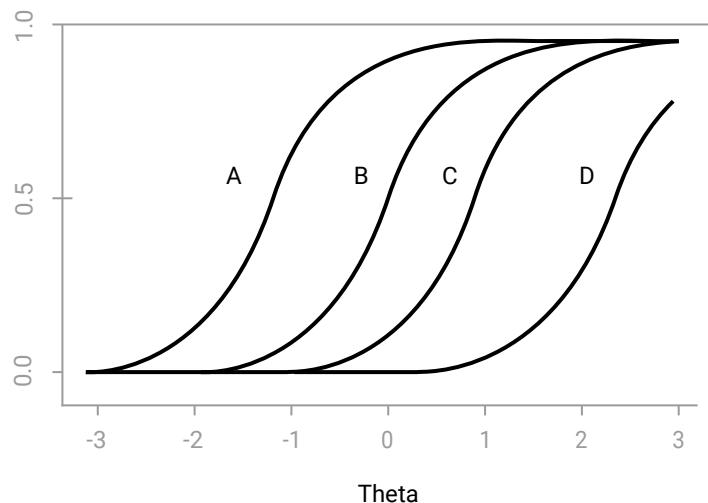
O modelo de um parâmetro da TRI, ou modelo de Rasch, é assim definido porque distingue em geral, um único parâmetro, qual seja, o da dificuldade dos itens. Entretanto, na TRI, pode haver também outros parâmetros de interesse, como o da discriminação dos itens. Tratemos, a seguir, desses dois parâmetros com um pouco mais de detalhes.

3.5.4 Parâmetro de dificuldade dos itens no modelo de Rasch

O modelo de Rasch também é chamado de modelo de um parâmetro porque, nele, todas as curvas são idênticas entre si, excetuando-se por um parâmetro, o da dificuldade, ao qual historicamente costumamo-nos referir como o parâmetro b .

Veja o Gráfico 9 a seguir, para identificar o parâmetro b .

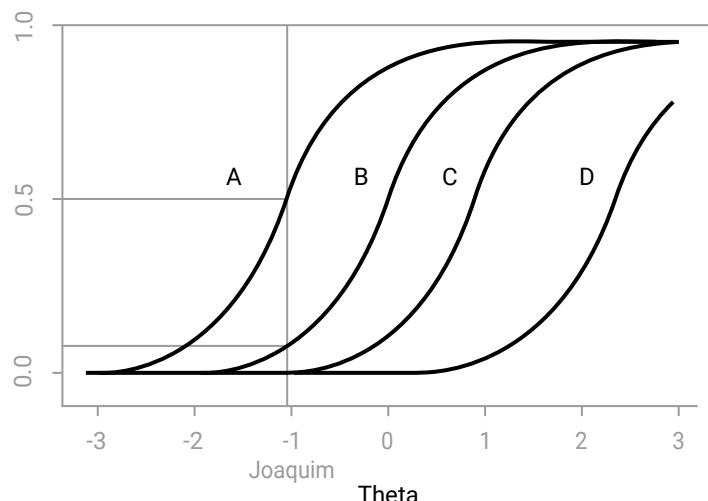
Gráfico 9: Curva característica dos itens A, B, C e D



Nesse exemplo, são representadas as CCIs de quatro itens – A, B, C e D –, com exatamente a mesma forma, apenas se diferindo umas das outras no seu respectivo deslocamento lateral. Já vimos que, quando consideramos a escala horizontal do gráfico, quanto mais caminhamos para a direita, mais aumenta o grau de dificuldade dos itens. Dessa forma, podemos dizer que, nessa representação, o item A é mais fácil do que o B, que é mais fácil do que o C, que, por sua vez, é mais fácil do que o D.

Continuemos com esse mesmo exemplo, porém agora com a indicação de como se comportam as chances de acerto nos diferentes itens, dependendo da proficiência dos avaliados.

Gráfico 10: Curva característica de itens e a probabilidade de acerto de Joaquim



Suponha que Joaquim, um estudante que respondeu aos itens A, B, C e D, indicados no exemplo, tem uma proficiência igual a -1 na escala arbitrária aqui utilizada.¹⁶

Para entendermos, no gráfico, qual é a probabilidade de acerto de Joaquim em cada um dos itens que compõem o teste exemplificado, precisamos traçar uma linha vertical sobre a sua respectiva proficiência (-1) e observar os pontos em que essa reta vertical corta as diferentes CCIs. As alturas das interseções nos dão, precisamente, a probabilidade de acerto de Joaquim em cada um desses itens.

Fazendo isso, percebemos, de imediato, que Joaquim tem 50% de probabilidade de acertar o item A. Isso porque, a partir do ponto de intercessão entre a reta vertical da proficiência de Joaquim e a curva do item A, podemos traçar agora uma reta horizontal, até encontrarmos o valor 0,5 (ou 50%) no eixo vertical, que nos fornece, precisamente, esse valor.

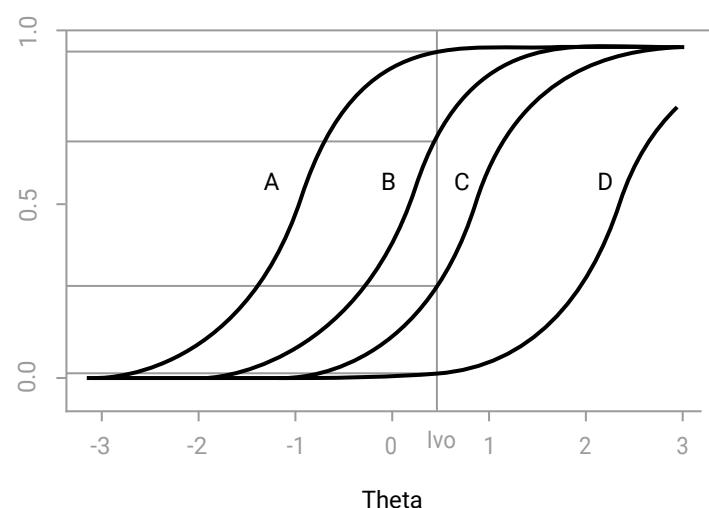
Ao realizarmos esse mesmo procedimento com o item B, percebemos que a probabilidade de Joaquim o acertar é consideravelmente menor, correspondendo a algo inferior a 10% (ou 0,1 na escala adotada). Isso, uma vez mais, deixa claro que o item B é, de fato, mais difícil do que o item A, visto que, mantendo constante a proficiência de Joaquim em -1, as suas chances de acerto diminuem consideravelmente quando passamos do item A para o B.

16 Cabe destacar que uma proficiência com valor negativo, embora possa, à primeira vista, parecer estranha, não tem, a princípio, nada de errado, visto que se trata de uma simples convenção.

Por outro lado, quando observamos os itens C e D, vemos que eles se encontram tão deslocados para a direita em relação à proficiência de Joaquim, que a interseção da linha vertical com as curvas dos dois itens quase ocorre num ponto de altura zero. Isso nos leva a considerar que a chance de Joaquim acertar esses dois itens mais difíceis é praticamente nula, embora não seja exatamente igual a zero, mas que podemos considerar como marginalmente superior a zero.

Considere, agora, o caso de Ivo, que possui uma proficiência igual a 0,5, portanto, bastante superior à de Joaquim.

Gráfico 11: Curva característica de itens e a probabilidade de acerto de Ivo



Ao traçarmos uma linha vertical sobre a sua respectiva proficiência (0,5) e observarmos os pontos em que essa reta vertical corta as diferentes CCIs, como demonstramos anteriormente, podemos perceber que a probabilidade de Ivo acertar o item A é alta, quase igual a 100% ou 1; a de acertar o item B é ainda elevada, sendo aproximadamente igual a 60% ou 0,6; por sua vez, a do item C é baixa, correspondente a aproximadamente 20% ou 0,2; finalmente, a probabilidade de acerto do item D, o mais difícil do grupo, é desprezível.

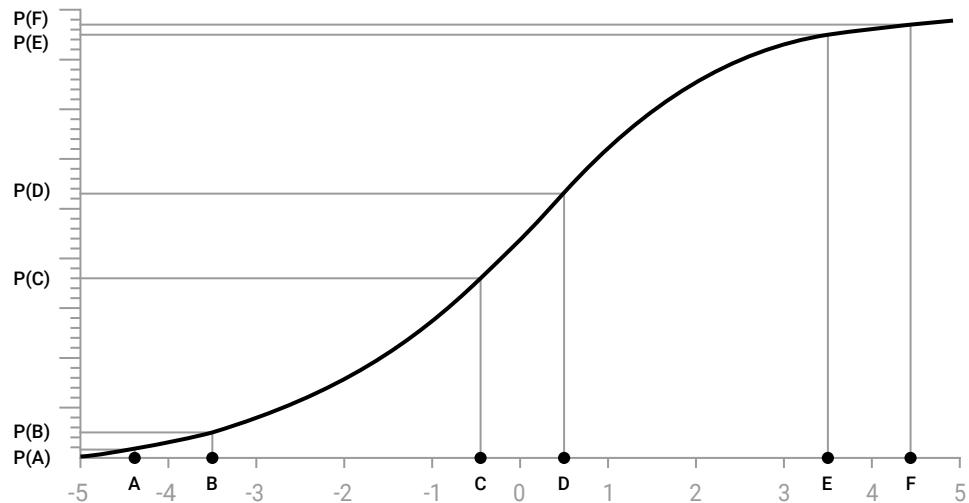
3.5.5 A discriminação dos itens no modelo de Rasch

Conforme comentamos anteriormente, no modelo de um parâmetro da TRI, a curva característica do item possui diferentes valores de inclinações, dependendo do ponto da curva considerado para análise.

Geometricamente, essas inclinações correspondem a retas tangentes à curva nos pontos de interesse. Outro aspecto relevante em relação à inclinação é o fato de que, quanto maior ela for, maior é a capacidade de discriminação do item. Além disso, destacamos que uma importante característica de um item é a sua capacidade de discriminação. Isto é, bons itens devem ser capazes de diferenciar quem sabe mais de quem sabe menos, de tal modo que sua tendência é ser acertado pelos que sabem mais e, ao mesmo tempo, ser errado pelos que sabem menos.

Vejamos mais um exemplo para esclarecermos como isso ocorre. O Gráfico 12 a seguir apresenta uma curva de Rasch em cujo eixo horizontal são marcados seis indivíduos, identificados pelas letras de A a F.

Gráfico 12: Proficiência de indivíduos na curva de Rasch



Os indivíduos exemplificados, quanto à sua respectiva proficiência ou habilidade, podem ser divididos em três duplas, a saber:

- > A e B, indivíduos de menor proficiência, cujos valores correspondem a, respectivamente, -4,5 e -3,5.
- > C e D, indivíduos de proficiência média, cujos valores correspondem a, respectivamente, -0,5 e 0,5.
- > E e F, indivíduos de maior proficiência, cujos valores correspondem a, respectivamente, 3,5 e 4,5.

Um dado relevante é que, dentro de cada dupla, a distância do indivíduo de menor para o de maior proficiência é igual a 1 ponto na escala arbitrada, conforme observamos nos valores apresentados: $4,5-3,5 = 0,5-(-0,5) = (-3,5)-(-4,5) = 1$.

Apesar de os indivíduos de maior e menor desempenho, dentre as duplas, estarem separados por distâncias iguais quanto à proficiência, percebemos que essa igualdade já não se repete quando levamos em conta a probabilidade de acerto no item representado pela curva do gráfico, ou seja, quando consideramos o eixo vertical da figura.

No eixo vertical, as probabilidades de acerto dos indivíduos A, B, C, D, E e F estão indicadas pelas respectivas alturas dos pontos P(A), P(B), P(C), P(D), P(E) e P(F).

É possível perceber que a distância vertical entre P(A) e P(B) é muito pequena, e, no outro extremo, na parte superior do eixo vertical, a distância vertical entre P(E) e P(F) também é muito pequena. Por outro lado, quando nos atentamos para a dupla do meio do gráfico – alunos C e D –, percebemos que a distância vertical entre P(C) e P(D) é consideravelmente elevada.

Dessa forma, concluímos que o item representado na figura, embora tenha uma única curva associada a ele, não possui igual discriminação em todos os pontos. Ao contrário, o poder de discriminação varia de acordo com o ponto observado. Nesse caso, a discriminação é menor se próxima às extremidades inferior e superior da proficiência, conforme medida no eixo horizontal. Isso ocorre porque, mesmo havendo um salto de um ponto na escala nessas regiões – de A para B, ou de E para F –, a probabilidade de acerto do item quase não aumenta, ou aumenta muito pouco.

Por outro lado, nesse exemplo, a maior discriminação do item ocorre na sua parte central. Nesse caso, quando fazemos a proficiência saltar de uma unidade, ao passarmos de C para D, percebemos um significativo aumento na probabilidade de acerto do item.

Para reforçar nossa compreensão, pensemos em uma segunda hipótese, utilizando o mesmo exemplo. Suponhamos que o item seja de matemática, aplicado por meio de um teste do 9º ano do ensino fundamental, no qual estão matriculados os estudantes C e D. Dessa forma, uma diferença de 1 ponto na escala de proficiência adotada faz com que os estudantes do 9º ano passem a ter probabilidades muito diferentes quanto ao acerto no item, que é próprio do seu ano de escolaridade.

Agora, imaginemos que os alunos A e B, também representados no Gráfico 12, estejam matriculados no 5º ano do ensino fundamental. Ora, isso faz com que suas respectivas proficiências sejam insuficientes para lhes darem uma grande probabilidade de acerto. Quanto a essa dupla de alunos, percebemos que o estudante B se encontra 1 ponto acima da escala em relação ao A. Seguindo por essa linha, caso aplicássemos a esses dois estudantes um item típico do ano em que eles se encontram – o 5º ano do ensino fundamental –, então a probabilidade de o aluno B acertar esse novo item seria muito maior que a do aluno A. Porém, como a esses dois estudantes foi apresentado um item de uma turma mais avançada – 9º ano do ensino fundamental –, ambos acabaram sendo “nivelados por baixo”, de modo que a probabilidade de acerto desse item para cada um acabou ficando quase igual – e quase nula.

De modo similar, poderíamos pensar que os alunos E e F, na extremidade direita, correspondem, por exemplo, a dois estudantes matriculados em um curso superior de Engenharia. Dessa forma, embora ambos também estejam separados por 1 ponto na escala de proficiência, como o item é muito fácil para eles, quase não há diferença entre suas respectivas probabilidades de acerto, que são quase iguais a 100%.

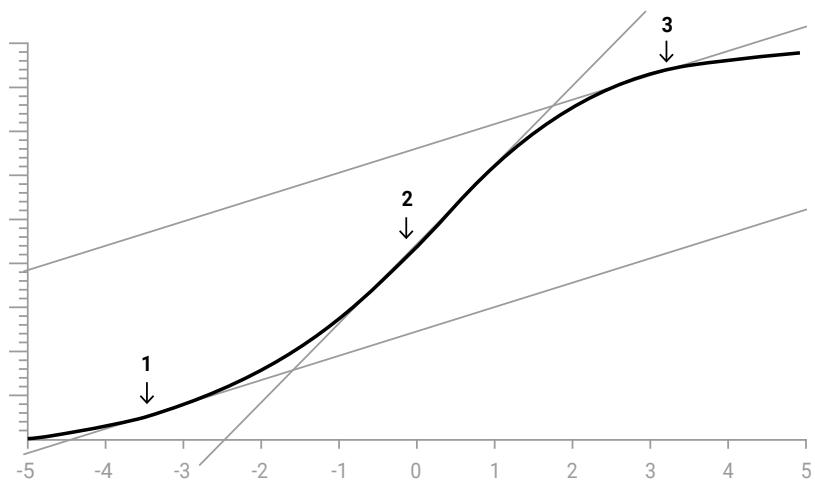
TRI e capacidade de discriminação do item

Tendo em vista tudo isso que apresentamos e discutimos, algumas considerações a respeito da capacidade de discriminação de um item segundo a TRI devem ser destacadas:

- 1) Para um mesmo item, a capacidade de discriminação varia de quase zero, em alguns trechos próximos aos dois extremos de sua respectiva curva, para um valor máximo, próximo ao centro da curva.
- 2) Em função disso, a capacidade de discriminação real de um item depende não apenas dele mesmo, mas também da proficiência do examinado, sendo tão maior quanto mais próxima do centro da curva estiver a proficiência dele.

Outra forma de perceber essa variação na capacidade de discriminação de um item é pela observação das retas tangentes à CCI em seus diferentes pontos, como indica o Gráfico 13 a seguir.

Gráfico 13: Capacidade de discriminação de um item de acordo com retas tangentes à CCI

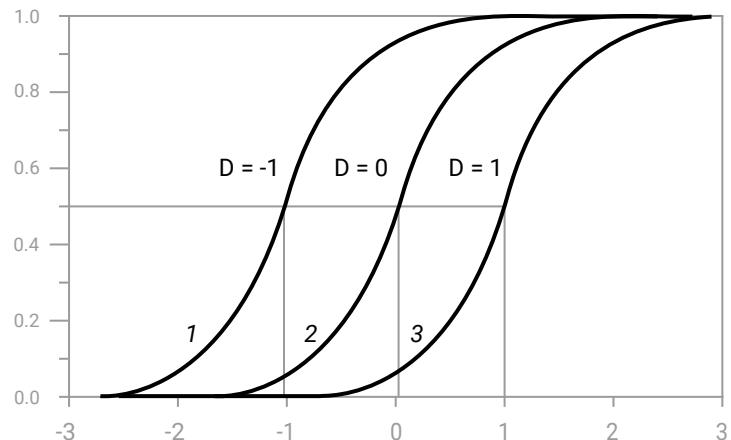


Temos, aqui, a mesma curva mostrada anteriormente. Vejamos que, na região mais à esquerda (1), a reta tangente à curva é pouco inclinada, quando comparada à reta tangente na região central (2). E a tangente volta a ficar mais baixa na região mais à direita (3).

Portanto, a medida geométrica da discriminação de um item é dada pela inclinação da reta tangente à sua curva característica. Essa inclinação, conforme vimos, está variando ao longo de toda a curva. Porém, na TRI, quando falamos do parâmetro de discriminação de um item qualquer, estamos sempre nos referindo ao máximo parâmetro que o item pode apresentar, como na situação 2, do Gráfico 13.

Num modelo de um parâmetro, ou de Rasch, cabe dizer que todos os itens têm o mesmo parâmetro de discriminação, apenas diferindo uns dos outros pelo parâmetro da dificuldade. E o parâmetro da dificuldade, por definição, no modelo de Rasch, corresponde ao valor da proficiência ou habilidade para a qual existe 50% de chance de se acertar o item, conforme se pode identificar na figura a seguir, na qual D corresponde ao parâmetro de dificuldade, também representado por b .

Gráfico 14: Parâmetro de dificuldade de itens no modelo Rasch

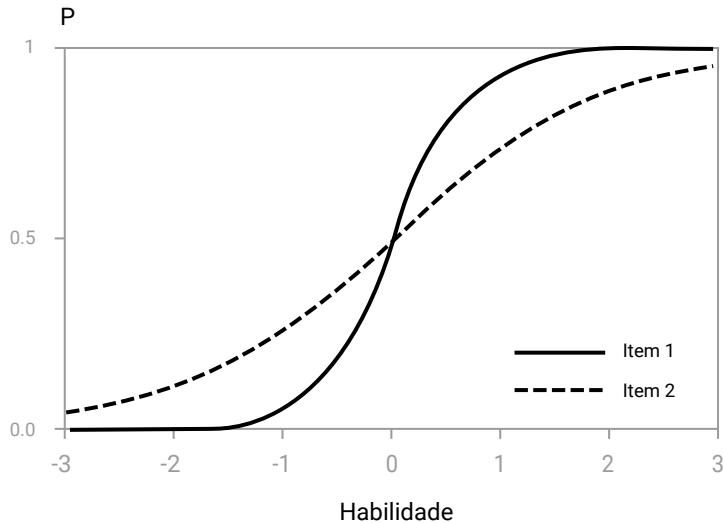


É possível observar, no Gráfico 14, que o item descrito pela curva 1 tem dificuldade igual a -1; o item descrito pela curva 2, dificuldade igual a 0; e o item descrito pela curva 3, dificuldade igual a 1. Nesses três pontos da escala da proficiência, as respectivas probabilidades de acerto correspondem a 50% (ou 0,5).

3.5.6 Modelo de dois parâmetros

No modelo de um parâmetro, ou de Rasch, dissemos que os itens, em geral, diferem-se quanto à dificuldade, porém são idênticos quanto à discriminação. Contudo, uma possibilidade adicional seria introduzir uma variabilidade para o parâmetro da inclinação. Isso é o que precisamente faz o chamado modelo de dois parâmetros da TRI, sobre o qual fornecemos um exemplo por meio do Gráfico 15 a seguir.

Gráfico 15: Exemplo de itens em modelo de dois parâmetros



Nesse exemplo, é possível verificar que os dois itens têm o mesmo parâmetro de dificuldade, visto que a probabilidade de 50% de acerto corresponde ao mesmo valor (nesse caso, 0), na escala horizontal de proficiência ou habilidade.

Contudo, a inclinação máxima do item 1 é maior do que a inclinação máxima do item 2, ou seja, o parâmetro de discriminação do primeiro item é superior ao do segundo. Comumente, o parâmetro de inclinação na TRI é chamado de a .

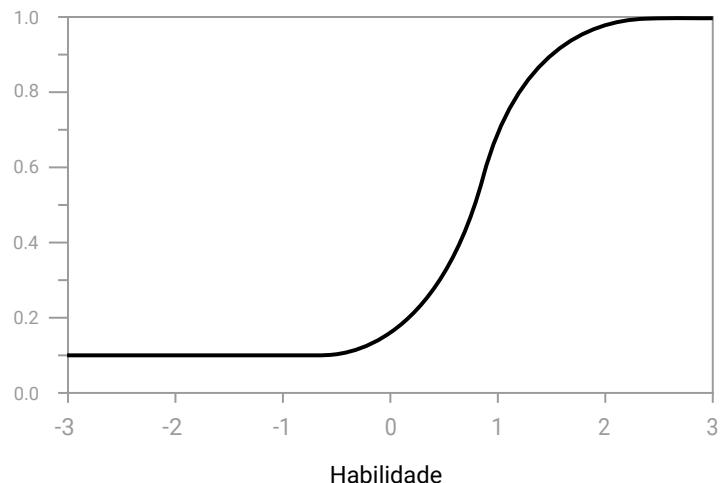
3.5.7 Modelo de três parâmetros

O modelo de três parâmetros da TRI é aquele no qual, além dos parâmetros de dificuldade b e de discriminação a , existe um terceiro parâmetro, c , para melhor modelar o chamado acerto casual.

Esse tipo de acerto é muito comum, por exemplo, em questões de múltipla escolha, quando o examinado pode acertar um determinado item por sorte, mesmo eventualmente possuindo proficiência inferior à necessária para responder a ele corretamente.

O Gráfico 16 a seguir mostra a curva de um item obtida segundo um modelo de três parâmetros.

Gráfico 16: Exemplo de item em modelo de três parâmetros



Conforme podemos ver, a característica que mais chama a atenção nesse terceiro modelo, em relação aos anteriores, é o fato de que a sua assíntota inferior, à esquerda, ao invés de ser igual a zero, corresponde a um valor maior do que isso – nesse exemplo, a algo em torno de 0,1. Isto é, por menor que seja a proficiência ou habilidade de um indivíduo, ou quanto mais para a esquerda caminharmos no gráfico, a probabilidade de acerto nunca será, nesse exemplo, menor do que 10%. Logo, $c=0,1$, valor correspondente a 10% na forma de proporção.

O valor de c , naturalmente, pode variar e, no caso dos itens de múltipla escolha, essa variação costuma depender do número de opções de resposta do item. Por exemplo, em itens de cinco opções, o parâmetro c poderia ser 0,2 (ou 20%), valor correspondente à probabilidade de se acertar casualmente uma opção em cinco possíveis. Caso o item tenha apenas quatro opções, o parâmetro c poderia subir para 0,25 (ou 25%), e assim por diante.

Entretanto, embora o valor de c costume variar conforme o número de opções, ele não depende apenas desse fator. Isso porque, em um item de múltipla escolha, também pode haver uma influência decorrente das outras opções. Por exemplo, se houver uma opção claramente falsa, ou esdrúxula, ela tenderá a ser evitada, mesmo pelos respondentes que sabem pouco sobre o que está sendo avaliado.

Algo nesse sentido acaba favorecendo quem “chuta” a questão, visto que o número de opções disponíveis diminui, o que aumenta a chance do acerto casual. Dessa forma, dependendo das características gerais de um item de múltipla escolha, o seu parâmetro c pode assumir valores ainda mais elevados, como 0,3 ou mais.

Em outras situações diferentes da reportada, pode ocorrer o contrário. É o caso das opções capciosas, popularmente conhecidas como “pegadinhas”, as quais podem levar os respondentes ao erro, principalmente aqueles com proficiência ou habilidade relativamente baixa, o que os impede de perceber a armadilha. Em casos assim, o parâmetro c tende a ser menor, pois o item passa a ser ainda menos acertado pelos respondentes de menor proficiência ou habilidade.

Antes, afirmamos que, quando pensamos na discriminação de um item, quanto maior for a inclinação da reta tangente à curva do item, maior é o seu poder discriminante, o que, por sua vez, aumenta a qualidade do item. Entretanto, devido a isso, o modelo de três parâmetros costuma apresentar um problema considerável: a curva já começa, à esquerda, dentro de um valor ou patamar relativamente elevado, e não mais do zero, como ocorria com os modelos de um e dois parâmetros. Com isso, a curva tem menos “teto” para subir.

Portanto, no modelo de três parâmetros, a curva de subida tende a ser mais suave, produzindo tangentes menos inclinadas, o que, naturalmente, resulta em itens com menor poder de discriminação.

Outra observação sobre esse modelo diz respeito ao parâmetro de dificuldade não mais corresponder à probabilidade de acerto igual a 50%, mas sim a probabilidades maiores do que isso, precisamente devido à diminuição de teto que já destacamos.

3.5.8 Afinal, qual tipo de modelo empregar?

O tipo de modelo da TRI a ser empregado na correção de um teste depende de vários fatores, não existindo uma regra única. Mesmo itens de múltipla escolha podem ser corrigidos segundo os modelos de um ou dois parâmetros, desconsiderando, assim, o acerto casual.

Geralmente, a escolha por um modelo específico é resultado de um equilíbrio, ou compromisso entre pontos favoráveis e desfavoráveis para a avaliação, que costumam estar presentes em todos os modelos estatísticos. Em um extremo, o modelo de um parâmetro tem a vantagem de ser mais simples. Além disso, como ele depende ape-

nas da dificuldade b , os parâmetros assim calculados tendem a ser mais robustos. Mas, ao mesmo tempo, essa aparente simplicidade tem um preço: o modelo de Rasch não considera, por exemplo, o fato bastante comum de que itens diferentes costumam ter diferentes capacidades máximas de discriminação.

No outro extremo, encontra-se o modelo de três parâmetros, que também possui vantagens e desvantagens. Uma de suas principais vantagens é ser, costumeiramente, mais realista no caso dos testes de múltipla escolha, os quais, por sua vez, são a forma dominante nas avaliações em larga escala. Por outro lado, o modelo de três parâmetros tem o problema, que já ressaltamos, de que a capacidade de discriminação dos itens tende a ser menor. Além disso, como o número de parâmetros a serem estimados cresce, enquanto o número de testes não, esses parâmetros assim obtidos tendem a ser mais instáveis, necessitando de uma maior quantidade de testes para serem estimados com segurança.

3.5.9 Pressupostos da TRI

Os itens tratados pela TRI devem satisfazer a dois pressupostos básicos. Vamos elencar cada um:

Unidimensionalidade

O primeiro pressuposto sinaliza a importância de um item avaliar uma única dimensão, e não duas ou mais simultaneamente. Já apontamos isso nas diretrizes do Capítulo 2, mas é importante reforçarmos aqui.

Por dimensão, entendemos como um determinado construto-alvo, que deve ser definido por especialistas – no caso educacional, por professores e pesquisadores de um determinado componente curricular a ser avaliado.

Naturalmente, nunca é possível obter itens puramente unidimensionais. Entretanto, há certas maneiras de se aumentar a unidimensionalidade, como deixar os enunciados dos itens mais simples e diretos.

Se uma avaliação de matemática tiver enunciados na forma de textos muito longos e/ou rebuscados, estará sendo testado não apenas o conhecimento matemático, mas também a capacidade de o estudante ler e entender o significado das instruções em português. E, nesse ponto, haveria uma quebra da unidimensionalidade.

Uma técnica estatística comum para se aferir a dimensionalidade dos itens é a chamada análise fatorial. Por meio dela, é possível analisar o comportamento de diversos indivíduos em relação a determinados itens, de modo a observar se esses itens estão ou não se comportando como parte de um grupo associado a uma dimensão em particular. Desse modo, é possível selecionar apenas os que estejam suficientemente vinculados a uma dada habilidade dominante.

Independência local

O segundo pressuposto diz respeito à independência local do item. Para satisfazer a esse pressuposto, um item deve funcionar dentro de um teste como se fosse um compartimento estanque, isto é, a resposta que se dá à tarefa requisitada não deve ser influenciada por eventuais respostas dadas a outros itens, e vice-versa. Em outras palavras, a resolução do item não pode influenciar o desempenho nos demais itens que compõem o teste.

3.6 Critérios de ancoragem de itens na construção de escalas de proficiência

Em avaliação educacional, mais especificamente na parte da psicometria a ela aplicada, o termo *ancoragem de itens* costuma se referir a dois fenômenos distintos, sendo que ambos são bastante relevantes para esse processo.

Um desses sentidos deriva da expressão em inglês *anchoring items*, ou itens-âncora, que são utilizados basicamente nos processos de equalização (*equating*, em inglês), com o propósito de colocá-los numa mesma escala de proficiência.

Esse procedimento surge do fato de que, naturalmente, é recomendável que os resultados obtidos por um sistema de avaliação sejam expressos em uma única escala de proficiência dentro de um determinado componente curricular, como matemática ou língua portuguesa. Desse modo, é garantida comparação dos resultados obtidos a partir de avaliações aplicadas em anos e locais distintos, mesmo se tratando de testes compostos de itens diferentes.

Para que isso seja possível, é necessário utilizar uma série de procedimentos, tanto segundo a TCT, quanto segundo a TRI. No caso da TRI, uma forma relativamente comum de se proceder à equalização é se valer precisamente desses chamados itens-âncora. Trata-se de itens que preferencialmente possuem um bom ajuste aos dados reais, ou

seja, suas respectivas curvas características aproximam-se bastante das distribuições reais dos respondentes aos itens. Além disso, esses itens devem também possuir valores relativamente elevados de discriminação, isto é, uma correlação alta ou ao menos moderada do item com o teste. Desse modo, estudantes que alcançam um melhor desempenho no teste como um todo tendem a acertar mais o item, enquanto aqueles com os menores escores gerais tendem a errá-lo.

Os itens-âncora devem ser escolhidos de tal forma que consigam cobrir uma faixa relativamente ampla, não só do currículo contemplado na avaliação em questão, mas também da sua respectiva escala de proficiência. Nesse sentido, alguns procedimentos bastante empregados, para se colocar os resultados de uma nova avaliação B na mesma escala de uma avaliação A já aplicada, seriam os seguintes:

- 1) *Estimar os parâmetros dos itens aplicados na avaliação A, segundo algum procedimento-padrão.* Esses parâmetros, quanto ao seu número, naturalmente variam de acordo com o modelo da TRI utilizado. Por exemplo, caso se utilize o modelo de Rasch, haverá somente um parâmetro a ser estimado para cada item, o da sua respectiva dificuldade, comumente chamado de b . Caso se adote o modelo de dois parâmetros, além da dificuldade, também haverá o parâmetro de discriminação do item, e caso se empregue o terceiro modelo, igualmente haverá o parâmetro c do acerto casual.
- 2) *Selecionar os itens-âncora da avaliação A, a serem utilizados também na Avaliação B.* Para isso, é preciso levar em conta uma série de critérios e características dos itens, conforme ressaltado anteriormente. Os itens assim selecionados deverão então ser também utilizados nas provas da avaliação B.
- 3) *Ao se avaliarem os resultados das provas da avaliação B, fixar os parâmetros dos itens-âncora de tal modo que eles conservem os mesmos valores que tinham na avaliação A.* Isso forçará o programa a colocar os demais itens de B, não mais comuns com o A, na mesma escala dessa última avaliação.

O procedimento de equalização – seja segundo o que foi esboçado anteriormente, seja conforme outro critério qualquer e adequado – permite que os resultados expressos a partir de testes compostos por itens diferentes, e aplicados em momentos e lugares diferentes, sejam colocados numa mesma escala. Isso, por sua vez, é de fundamental importância para o acompanhamento temporal dos resultados, bem como para a comparabilidade entre eles.

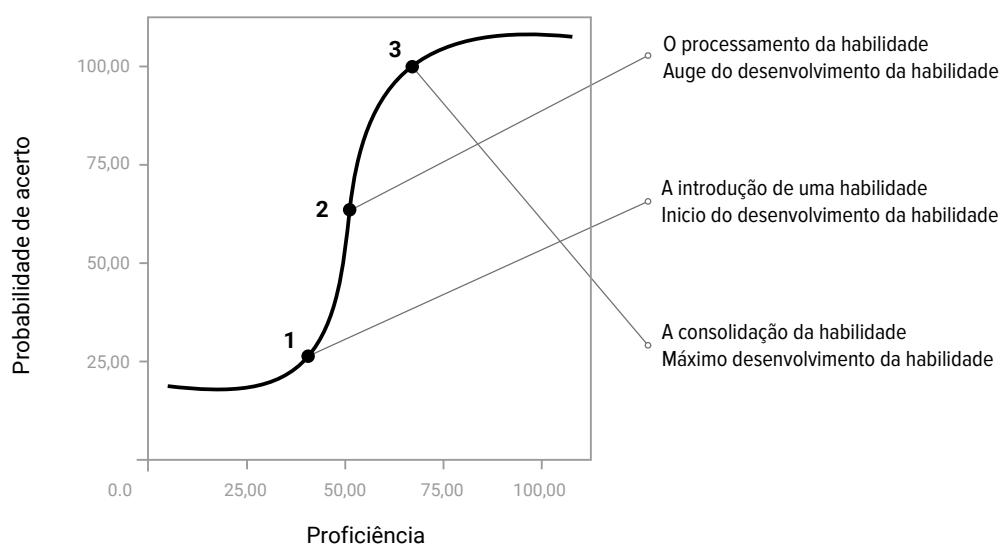
Uma segunda forma de se referir ao termo *item-âncora* diz respeito à sua respectiva utilização para dar sentido às escalas de proficiência utilizadas. Por exemplo, no caso do Saeb, para cada disciplina avaliada no ensino básico, existe uma matriz de referência, com a relação das habilidades a serem avaliadas nos testes. Mas também existe uma escala de proficiência para cada disciplina. Portanto, o desafio é criar uma associação entre a matriz e a escala, de tal modo que, para o estudante cuja proficiência estimada pelos testes esteja situada num determinado ponto da escala, seja possível afirmar o que ele já domina e o que ainda não domina do currículo.

Nesse sentido, os itens-âncora funcionam como *padrões* pelos quais, analisando o seu respectivo comportamento estatístico e também o componente curricular ao qual se refere, é possível descobrir, com base na proporção de estudantes que errou ou acertou o item, o que ele está de fato mensurando, isto é, que relação existe entre o desempenho do estudante e o presumido nível de proficiência que ele possui.

Uma das formas de se obter isso é por meio da consideração da CCI. Conforme já detalhamos, ela representa a probabilidade de acerto de um item – mensurada num eixo vertical, com escala valendo de 0 a 1 – para a proficiência theta (θ) do aluno.

O Gráfico 17, a seguir, apresenta uma CCI, bem como alguns de seus pontos de particular interesse, para fins dessa interpretação.

Gráfico 17: Fase de desenvolvimento das habilidades



A respeito desse exemplo, podemos observar que o ponto 1, que aparece em destaque mais à esquerda, é aquele no qual a habilidade começa a ser introduzida aos estudantes. Isso também quer dizer que o percentual de acerto do item, nesse ponto da habilidade, é ainda muito baixo. No entanto, também se percebe que, a partir desse ponto, a CCI passa a se tornar mais inclinada enquanto se caminha mais para a direita, indicando um aumento da capacidade de discriminação do item a partir desse ponto.

O ponto 2 – que aqui aparece próximo ao centro da CCI – corresponde ao lugar da curva onde a sua inclinação é máxima. Isso quer dizer que é exatamente nesse ponto que uma dada variação da proficiência – medida no eixo x – possibilita um maior aumento na probabilidade de acerto do item – no eixo y. Contudo, nesse ponto, a probabilidade de acerto ainda se encontra em torno de 50%.

Por fim, o ponto 3, mais à direita, indica onde a habilidade já pode ser considerada consolidada, visto que, para um estudante com esse nível de proficiência, a probabilidade de acerto é bastante alta, sendo, nesse caso, bem próxima de 100%.

Ao escolhermos os itens-âncora, precisamos nos atentar para a relação existente, por exemplo, entre o seu respectivo ponto de consolidação e a habilidade a ele correspondente. Dessa forma, é possível situar os diferentes itens-âncora nos diferentes pontos da escala de proficiência, ao mesmo tempo que também se promove uma interpretação pedagógica desses itens, com base nos seus respectivos descritores, habilidades, unidades temáticas etc.

Dissemos anteriormente que é importante que esses itens-âncora possuam uma boa capacidade de discriminação. Em termos visuais, como no gráfico da CCI apresentando, isso quer dizer que o item deve – em algum ponto próximo ao centro de sua respectiva CCI – possuir uma elevada inclinação.

O trecho a seguir mostra qual foi o critério para a ancoragem dos itens em um importante sistema de avaliação norte-americano, denominado National Assessment of Educational Progress (NAEP). O NAEP avalia os anos finais de determinados ciclos do ensino básico dos Estados Unidos, à semelhança do que o Saeb, no Brasil, também faria décadas depois, ao privilegiar a avaliação do 5º e do 9º anos do ensino fundamental e do 3º ano do ensino médio.

A ancoragem é um meio utilizado pelo NAEP para a construção da escala de proficiência, que se realiza por meio de um procedimento simples e direto, baseado apenas no aumento da proporção de alunos que respondem corretamente ao item entre pontos-âncora adjacentes.

Esse processo permite verificar o que a maioria dos alunos em um ponto-âncora sabe e pode fazer, e o que a maioria dos alunos situados nos níveis imediatamente mais baixos não sabe e não pode fazer. O NAEP utiliza dois métodos para decidir que itens serão selecionados para a descrição dos pontos-âncora.

O primeiro método é denominado método direto, porque o que os alunos sabem e conseguem fazer relaciona-se diretamente a um ponto-âncora. Como poucos alunos têm (ou nenhum aluno tem) proficiência exatamente no ponto-âncora na escala do NAEP, as proficiências são agrupadas em um pequeno intervalo acima e abaixo do ponto-âncora, de modo a haver pelo menos 100 alunos no intervalo representativo do ponto em questão. Esse método se fundamenta em dois critérios:

- i) 80% dos alunos em um ponto-âncora respondem corretamente ao item, e, no ponto imediatamente inferior, menos de 50% dão a resposta correta ao item.
- ii) 65% dos alunos no ponto mais alto respondem corretamente ao item, e menos de 50% dos alunos, no nível imediatamente mais baixo, são capazes de responder corretamente ao item, sendo que a diferença de acerto entre esses níveis é de, no mínimo, de 30 pontos percentuais.

Para os escores que estão no ponto-âncora mínimo (o primeiro da escala), ou próximo dele, o critério para a seleção do item é: (i) 80% ou (ii) 65% de alunos respondendo corretamente o item. O primeiro critério é menos aplicável e pode não produzir itens suficientes para a interpretação; por isso, o segundo critério é o mais utilizado.

O segundo método é chamado de método refinado, porque se enquadra no modelo logístico de três parâmetros para a escalonagem de escores e usa as Curvas Características dos Itens (CCIs) para aproximar, nos pontos de ancoragem selecionados, a porcentagem de acerto em cada item. Os parâmetros do item tornam-se facilmente disponíveis, se a escala tiver sido produzida segundo os métodos da TRI. Esse método utiliza o mesmo critério de seleção de itens, exceto pelo fato de que os tamanhos da amostra nos pontos-âncora não estão disponíveis, e, portanto, o tamanho mínimo da amostra não é utilizado. Ambos os métodos produzem resultados semelhantes. (OLIVEIRA, 2008, p. 50-51).

Conforme se disse, o exemplo do NAEP é apenas uma possibilidade de procedimento quanto aos critérios de ancoragem dos itens. Dependendo do tipo de sistema de avaliação utilizado, ou mesmo dentro de um único sistema, também é possível empregar outros procedimentos. No entanto, independentemente de qual procedimento se utilize, o propósito geral acaba sendo o mesmo, ou parecido: utilizar os itens – ou um conjunto deles – para, com base no desempenho observado dos estudantes, construir uma interpretação da escala de proficiência que se coadune com a matriz de referência utilizada.

3.7 Padrões de desempenho estudantil para nortear as interpretações pedagógicas das escalas de proficiência

Conforme vimos até aqui, para cada componente curricular avaliado, é possível criar uma escala de proficiência, no nível intervalar. Isso significa que, nessa escala, o zero é arbitrário, ao mesmo tempo que se supõe que os intervalos entre duas proficiências diferentes – ou entre os parâmetros de dificuldade de dois itens – seja mensurável.

Sendo assim, com base na descrição e análise dos itens-âncora, pode-se “mapear” a aprendizagem curricular dos estudantes. Ou seja, a determinados pontos da escala de proficiência, é possível associar determinados itens-âncora, cuja interpretação indica que tipo de habilidade se encontra ou não consolidada pelos alunos naquela parte específica da escala.

Portanto, também é possível criar intervalos nessa mesma escala que definem determinados níveis de habilidade ou proficiência, segundo uma interpretação a ser feita por especialistas em cada disciplina, eventualmente auxiliados tecnicamente pelos estatísticos que produziram os parâmetros dos itens e as proficiências dos alunos nos testes. Com base nisso, pode-se, por exemplo, para um determinado ano escolar, dividir o possível desempenho dos estudantes em níveis. Isso tem, por sua vez, o potencial de auxiliar na interpretação dos resultados e, consequentemente, em eventuais ações e esforços em prol da melhoria do ensino e aprendizado nas escolas.

Nesse sentido, para um dado ano e disciplina, é possível dividir a escala de proficiência por meio da delimitação de diferentes níveis de desempenho, sem interseção entre si e em ordem ascendente quanto às proficiências envolvidas. Tal divisão possui um caráter cumulativo, pelo qual as habilidades de um dado nível devem estar sendo dominadas não só pelos estudantes cuja proficiência se situa nesse mesmo nível, mas também pelos alunos situados nos níveis mais acima.

Um ponto particularmente útil, no uso desses níveis, é que eles permitem uma interpretação mais panorâmica da escala de proficiência. Ou seja, os níveis permitem que se observe a escala de proficiência levando em conta determinados agrupamentos de habilidades associadas a diferentes etapas do ensino. E tal visão, por sua vez, é mais informativa do que, por exemplo, aquela que se baseia apenas na pontuação pura e simples obtida pelos estudantes, sem se atentar para o que esses números significam.

Imaginemos que uma escola obteve uma média de 261 pontos em matemática no 9º ano do ensino fundamental, em uma avaliação do Saeb. E que, na edição seguinte do teste, o resultado foi de 258 pontos. Será que, entre essas duas avaliações, essa escola regrediu? Ou será que podemos dizer que, em linhas gerais, ela está no mesmo lugar? Dito ainda de outro modo, estamos preocupados em saber se uma queda de 3 pontos na escala do Saeb é, de fato, um decréscimo significativo de desempenho, ou se está mais para uma flutuação aleatória. Embora, certamente, não se trate de algo simples a ser resolvido, esse problema pode ser razoavelmente bem atacado quando se leva em conta a dispersão dos resultados – ou seja, as variações de desempenho observadas entre os estudantes, e também entre suas respectivas escolas. A partir da medida dessa variabilidade, comparada com as diferenças reais observadas, pode-se, então, associar a isso uma função probabilística que nos permite decidir se as diferenças observadas são ou não significativas do ponto de vista estatístico.

Nesse sentido, um exemplo de níveis de desempenho, em relação ao Saeb, definido pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), é o seguinte:

Tabela 6: Padrões de desempenho do Saeb

Nível de aprendizado	5º ano		9º ano	
	Língua Portuguesa	Matemática	Língua Portuguesa	Matemática
Insuficiente	Até 150	Até 175	Até 200	Até 225
Básico	De 150 a 200	De 175 a 225	De 200 a 275	De 225 a 300
Proficiente	De 200 a 250	De 225 a 275	De 275 a 325	De 300 a 350
Avançado	Acima de 250	Acima de 275	Acima de 325	Acima de 350

Como se vê, tanto para língua portuguesa quanto para matemática, são apresentados quatro níveis de proficiência, chamados de insuficiente, básico, proficiente e avançado.

Observamos que, para cada disciplina, a escala é diferente, bem como também são diferentes, em geral, os limites entre os níveis de aprendizado assim estabelecidos. Não obstante, dentro de cada disciplina, a escala é a mesma. No presente caso, uma preocupação que também se nota na tabela é a de criar níveis de aprendizado específicos para o 5º e o 9º anos do ensino fundamental, que correspondem aos términos, respectivamente, dos ciclos inicial e final desse mesmo segmento. Da mesma forma, também é possível criar níveis para o 3º ano do ensino médio, que também é avaliado pelo Saeb.

É importante salientar a necessidade de se possuir um objetivo e um contexto específico para se propor uma classificação em níveis. Mas, ainda que, quanto a isso, possa haver diferenças entre sistemas de avaliação, o objetivo geral dessa classificação é o de facilitar a interpretação das escalas e dos resultados, além de contribuir para que a apropriação desses resultados venha a possibilitar um aperfeiçoamento da realidade educacional.

4. O que você aprendeu até aqui?

Neste capítulo, mostramos que os testes de avaliações em larga escala possuem diferentes propósitos, os quais estão relacionados às consequências dos seus resultados para os atores educacionais envolvidos – estudantes, professores e gestores –, que podem ser baixas ou altas, e à forma como esses dados serão utilizados. Uma divisão bastante conhecida é a de avaliação somativa e formativa: a primeira, geralmente aplicada após o fim de um ciclo escolar, tem como propósito verificar se as expectativas de aprendizagem foram cumpridas, com uma análise mais centrada no nível das redes; enquanto a segunda se dá ao longo do ano letivo, de forma a se realizar um acompanhamento da aprendizagem, com foco maior no uso dos resultados por professores.

O resultado das avaliações, por sua vez, podem ser calculados com base na TCT, que se vale, basicamente, de procedimentos estatísticos mais convencionais, como os percentuais de acerto dos estudantes, e na TRI, que desenvolve modelos probabilísticos de maior sofisticação, com base em outros critérios, como o nível de dificuldade dos itens. A TCT é mais empregada em avaliações internas, que são aplicadas, realizadas e corrigidas pelos próprios professores, enquanto a TRI se enquadra, com maior frequência, no campo da avaliação externa em larga escala, conduzida por redes e sistemas educacionais.

Com isso, entendemos que há três parâmetros possíveis para se calcular a proficiência de um estudante, com base na TRI. O modelo de um parâmetro, ou Rasch, cujo critério é o nível de dificuldade do item; o modelo de dois parâmetros, que, além do nível de dificuldade, leva em conta o parâmetro de discriminação; e o modelo de três parâmetros, que, além desses dois mencionados, utiliza também o critério de acerto ao acaso. Cada um desses parâmetros possui suas vantagens e limitações, de modo que a escolha em relação a qual utilizar deve estar associada ao objetivo da avaliação.

Compreendemos também o que é o índice de discriminação de um item. Ao contrário do que o nome possa sugerir, um item capaz de discriminar representa algo positivo, pois tende a ser acertado por aqueles que dominam a habilidade avaliada, e a ser errado por quem ainda não a desenvolveu. O índice de discriminação é, também, algo fundamental para o uso de itens-âncora, que são itens utilizados em mais de uma avaliação, de modo a ser possível realizar comparações entre os resultados de cada uma, com o objetivo de se acompanhar possíveis evoluções ou retrocessos.

Todas essas questões devem ser compreendidas e levadas em conta quando discutimos a fidedignidade das avaliações. Isso porque a fidedignidade de um teste relaciona-se à sua capacidade de produzir resultados consistentes em relação àquilo que foi avaliado, de modo que depende dos critérios utilizados para sua aplicação, correção e cálculo dos resultados de desempenho. Embora não existam critérios perfeitos para a aplicação de um teste, cada critério adotado depositará um grau de fidedignidade à medida produzida ao final desse processo.

Por fim, neste capítulo, abordamos os padrões de desempenho, que são intervalos na escala de proficiência que associam os estudantes a determinados níveis de aprendizagem, de acordo com a etapa e o componente curricular avaliados. Nas avaliações em larga escala, costumamos encontrar quatro padrões de desempenho, que vão desde o nível mais baixo de proficiência – aquém das expectativas do estudante em relação à etapa de ensino em questão –, até o mais alto – em consonância com as expectativas.

Isso permite uma compreensão mais clara do que exatamente significam certos resultados de desempenho, em termos de desenvolvimento escolar. Pois, para cada padrão de desempenho que possa vir a ser alcançado pelo estudante, deve haver uma descrição sobre o conjunto de habilidades e competências que ele já desenvolveu e que, a depender, ainda precisa desenvolver.

5. Referências

- HAMBLETON, R. et al. *Fundamentals of item response theory*. Newbury Park: SAGE Publications, 1991.
- HOGAN, T. *Introdução à prática de testes psicológicos*. Rio de Janeiro: LTC Editora, 2006.
- KORETZ, D. *Measuring up: what educational testing really tells us*. Cambridge: Harvard University Press, 2008.
- LEVINE, D. M.; STEPHAN, D. F.; SZABAT, K. A. *Estatística – teoria e aplicações: usando Microsoft Excel em português*. Rio de Janeiro: LTC Editora, 2000.
- OLIVEIRA, L. K. M. *Três investigações sobre escalas de proficiência e suas interpretações*. 2008. Tese (Doutorado em Educação) – Instituto de Educação, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.
- SZABÓ, G. *Applying item response theory in language test item bank building*. Frankfurt: Peter Lang, 2008.

6. Testando o que você aprendeu

- 1) Em um teste de geografia composto de sete itens, todos de mesmo peso, a proficiência foi calculada com base no número de acertos. Esse teste foi aplicado a 200 estudantes, com os resultados apresentados no quadro a seguir:

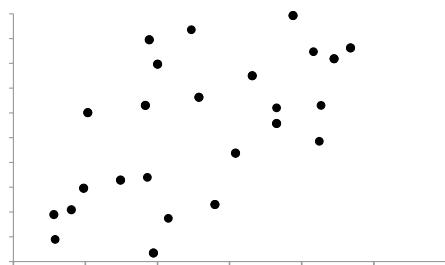
PROFICIÊNCIA	N	%	%_ACUMULADO
0	10	5	
1	20	10	
2	20	10	
3	40	20	
4	50	25	
5	20	10	
6	20	10	
7	20	10	

Nesse quadro, a coluna “N” apresenta o número de estudantes por quantidade de acerto e a coluna “%” apresenta o percentual que cada classe de resultados da proficiência obtida representa em relação à totalidade de alunos.

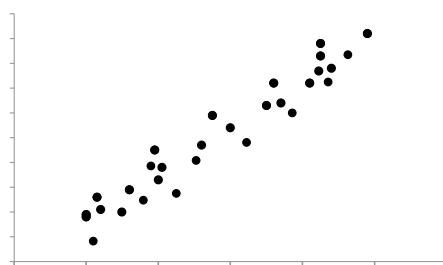
- 1.1) Preenche a coluna “%_acumulado”, de acordo com os dados e informações apresentados.
- 1.2) A qual percentil corresponde a proficiência 2?
- 1.3) Quantos por cento dos estudantes tiraram, no máximo, nota 2?
- 1.4) Qual é a mediana desse grupo?
- 1.5) Quantos por cento dos estudantes tiraram, no máximo, nota 5?
- 1.6) A qual nota corresponde o 7º percentil?
- 2) Nos dois diagramas de dispersão a seguir, a escala vertical é a mesma, correspondente à proficiência em geografia mensurada em um conjunto de estudantes, por meio de testes psicométricamente consolidados e aplicados ao longo de vários anos. Dessa forma, cada ponto nos gráficos representa um aluno. Por

outro lado, existem dois novos testes de geografia – o A e o B – ainda em fase de estudos pela equipe de avaliação responsável pela sua elaboração, de modo que o eixo horizontal corresponde aos resultados do teste A no diagrama da esquerda, e ao teste B no da direita.

Teste A



Teste B



Com base nos resultados apresentados nesses diagramas, responda ao que se pede:

- 2.1) Que sinal tem a correlação r de Pearson em cada um desses diagramas: negativo, positivo ou aproximadamente zero?
- 2.2) Se você tivesse que escolher um desses testes para aplicar aos seus estudantes, qual teste você escolheria? Justifique sua resposta.

- 3) **Os casos a seguir correspondem a algumas situações hipotéticas envolvendo avaliações de desempenho em larga escala, sendo que, em todas elas, ocorrem certos problemas que têm o potencial de comprometer a qualidade da avaliação. Para cada caso identifique o tipo de problema existente, descreva-o sucintamente e proponha uma forma de solucioná-lo ou, pelo menos, amenizá-lo.**

Caso 1. No 3º ano do ensino médio, na Escola A, o estudo dos poliedros costuma ser feito em março, enquanto, na Escola B, esse mesmo tópico costuma ser tratado em novembro. Porém, ambas as escolas farão uma avaliação externa de matemática em outubro.

Caso 2. Uma secretaria municipal de educação aplicou um teste de redação para o ensino médio, a um total aproximado de 900 estudantes. Para corrigi-lo, foi formada uma equipe de 10 professores de língua portuguesa. As provas, então, foram divididas em 10 conjuntos de aproximadamente 90 redações cada um, e enviadas para os professores. A cada um, solicitou-se que corrigissem os testes segundo as práticas já seguidas em suas respectivas escolas, deixando claro que as notas dadas por eles seriam as definitivas, isto é, cada redação passaria apenas por um único corretor.

- 4) **Suponha que 4 itens de língua portuguesa tenham sido pré-testados em um grupo de 200 estudantes, com os resultados divulgados no quadro a seguir.**

Item	Acertos		
	Total	Grupo 1	Grupo 2
1	68	11	57
2	145	66	79
3	28	9	19
4	106	22	84

Nesse quadro, a primeira coluna apresenta o número ou a identificação do item; a segunda coluna, o número total de acertos, levando em conta os 200 estudantes; a terceira, o número de acertos no grupo de 100 estudantes com os menores resultados no teste como um todo, aqui chamados de alunos do Grupo 1; e, finalmente, na quarta coluna, o número de acertos no grupo de 100 estudantes com os maiores resultados no teste, aqui chamados de Grupo 2.

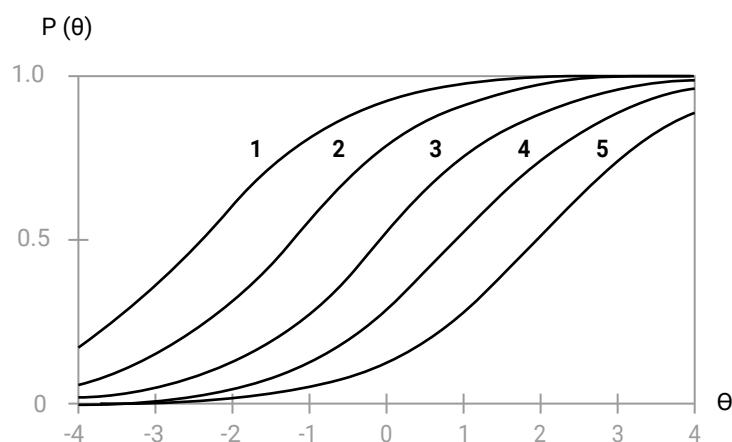
- 4.1) Calcule, para cada item, o seu índice – geral ou total – de dificuldade, segundo a Teoria Clássica dos Testes (TCT).
- 4.2) Caso você tivesse que tirar dois itens desse grupo para compor um teste considerado fácil, quais itens você escolheria, com base nos resultados obtidos na sua resposta à questão 4.1?

- 4.3) Caso você tivesse que tirar dois itens desse grupo para compor uma prova considerada difícil, que itens você escolheria, com base nos resultados obtidos na sua resposta à questão 4.1?
- 4.4) Calcule os índices de discriminação para cada item, segundo o método da Teoria Clássica dos Testes (TCT).
- 4.5) Caso você tivesse que tirar dois itens desse grupo para compor um teste que tivesse o maior grau possível de discriminação, quais itens você escolheria, com base nos resultados obtidos na sua resposta à questão 4.4?
- 4.6) Com base nos resultados obtidos, é possível estabelecer algum tipo de associação, ainda que genérica, entre o grau de dificuldade dos itens e o seu respectivo grau de discriminação? Justifique sua resposta.
- 5) **Considere o quadro a seguir, que apresenta uma amostra dos resultados de um teste de matemática. Cada aluno avaliado se encontra em uma linha, na qual o acerto em um item corresponde a 1, e o erro, a 0. Responde às questões 5.1 e 5.2, levando em conta que NÃO é preciso resolver nenhum dos problemas matemáticos do quadro!**

Item (nº)	1	2	3	4	5
Item	$a = 4 + 7$	$a = 3 + 2 \times 5$	$a^2 - a = 0$	$a = \log(1000 \times 100) - \log 1$	$a = d(x^2 + 9x - 4)/dx; x = 3$
Pedro	1	0	0	0	0
Paula	1	1	1	0	0
Francisco	1	1	1	1	1

- 5.1) Considere que a escala de dificuldade dos itens também aparece na primeira linha, ou seja, suponha que o primeiro item tem um parâmetro de dificuldade 1 em uma escala proficiência; que o item 2 tem dificuldade 2; o item 3, dificuldade 3; e assim por diante. Dessa forma, usando o procedimento cumulativo de Guttman, qual é o valor da proficiência de cada um dos estudantes avaliados?
- 5.2) Suponha que um quarto estudante, Maurício, que não aparece no quadro acima, respondeu, em uma prova à parte, às questões 2 e 3 desse mesmo teste, acertando a segunda e errando a terceira. Mesmo sem ter respondido às demais questões (1, 4 e 5), seria possível prever o desempenho de Maurício em cada uma delas, caso tivesse tentado fazê-las? Explique e estime a proficiência de Maurício.

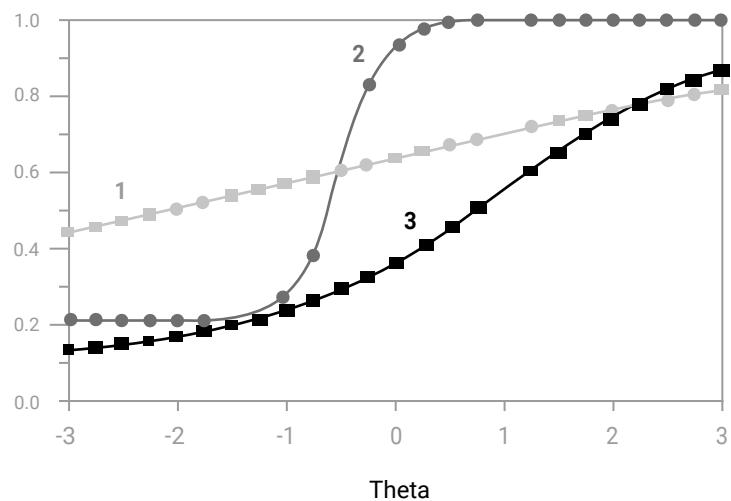
- 6) Na figura a seguir, estão representados quatro itens de matemática, com a habilidade sendo mensurada no eixo horizontal, e a probabilidade de acerto, no eixo vertical. As curvas desses itens têm uma assíntota inferior igual a zero. Isso quer dizer que, caso fosse possível visualizar as curvas nas suas respectivas extremidades esquerdas, elas se aproximariam do eixo horizontal na direção dos menores valores de θ , porém sem jamais tocarem esse eixo.



Com base na representação desses quatro itens, responda às questões 6.1 a 6.5.

- 6.1) Que tipo de modelo da Teoria da Resposta ao Item (TRI) está sendo empregado para analisar esses itens?
- 6.2) Qual é o item mais fácil?
- 6.3) Qual é o item mais difícil?
- 6.4) Qual é o item mais discriminante?
- 6.5) Quanto vale, aproximadamente, o parâmetro de dificuldade de cada item?
- 6.6) Qual item tem a maior capacidade de discriminação para $\theta=3$?

- 7) O gráfico a seguir mostra as curvas características (CCIs) de três itens de física, sendo que o eixo horizontal representa a proficiência nessa disciplina, caracterizada pela letra Theta (θ), e o eixo vertical, a probabilidade de acerto.



Com base nessas CCIs, responda às questões 7.1 a 7.5.

- 7.1) A que tipo de modelo parece pertencer o item 2?
- 7.2) Qual é o item mais discriminante quando $\theta = -2,5$?
- 7.3) Qual é o item mais discriminante quando $\theta = -0,5$?
- 7.4) Qual é o item mais discriminante quando $\theta = 2$?
- 7.5) Imagine que, em uma turma, houvesse uma divisão em três grupos, de acordo com o desempenho dos estudante em física: alto, médio e baixo. Supondo que você tivesse que aplicar apenas um dos itens desse gráfico a cada um desses grupos de alunos, a qual grupo você destinaria cada item?

GABARITOS E RESOLUÇÕES DAS ATIVIDADES

Atividades do Capítulo 1

1) Gabarito: M, O, O, M, O, M, O, O, M, O.

2) Para corrigir essa atividade é necessário observar a estrutura do objetivo elaborado, devendo conter, necessariamente: O **verbo no infinitivo**, que deverá indicar o processo cognitivo – isto é, memorizar, compreender, aplicar, analisar etc. – que é empregado no objetivo; e o **objeto**, que precisará indicar o tipo de conhecimento que é o foco desse processamento mental.

3) Gabarito:

Habilidade	Processo cognitivo					
	Lembrar	Entender	Aplicar	Analisar	Avaliar	Criar
Identificar teses/opiniões/ posicionamentos explícitos e argumentos em textos		X				
Analisar o uso de figuras de linguagem como estratégia argumentativa					X	
Avaliar diferentes graus de parcialidade em textos jornalísticos						X
Identificar elementos constitutivos de gêneros de divulgação científica	X					
Analisar as variedades linguísticas em textos				X		
Producir texto em língua portuguesa, de acordo com o gênero textual e o tema demandados						X
Utilizar marcas de parcialidade em textos jornalísticos			X			
Avaliar a adequação das variedades linguísticas em contextos de uso					X	

Atividades do Capítulo 2

Atividades de conhecimento geral sobre elaboração de itens

1) Gabarito: C

2) Gabarito: C

3) Gabarito: B

4) Gabarito: D

Atividades sobre itens da etapa de alfabetização

5.1) Gabarito C

5.2) O item avalia a habilidade de *ler palavras formadas por sílabas canônicas*, relacionada à apropriação do sistema de escrita alfabética. Os estudantes que marcaram a alternativa C – VACINA –, que é o GABARITO, identificaram a resposta correta entre as palavras, cujas primeiras sílabas se aproximam pelo mesmo ponto de articulação – labiodentais –, mas se afastam em relação ao modo de articulação, pois o fonema /f/ é surdo e o /v/, sonoro. A presença das mesmas vogais como centro de sílaba em todas as palavras contribui para que os estudantes tenham que ler todas as palavras até o final para acertar o item.

6) Gabarito: D

7.1) O gabarito é a alternativa A.

7.2) Reconhecer o assunto de um texto.

7.3) Sim. Trata-se de um texto de nível 1, que é adequado a estudantes em fase de alfabetização, isto é: texto curto formado por períodos de, no máximo 2 linhas, na ordem direta, com poucas intercalações e que apresenta temática próxima ao universo das crianças.

7.4)

RESOLUÇÃO JUSTIFICADA	
ALTERNATIVA A	Distrator. Os estudantes que assinalaram esta alternativa provavelmente se apoiaram na ideia de sujeira causada pela queda dos pelos dos cães. É um tópico que contribui para a compreensão do assunto, mas que não corresponde ao que foi solicitado no enunciado do item.
ALTERNATIVA B	Gabarito. Os estudantes que assinalaram esta alternativa demonstraram ter desenvolvido a habilidade avaliada e reconheceram corretamente o assunto, identificando pistas, ao longo do texto, que reforçam o tópico central da escovação diária dos pelos dos cães: o título do texto que aponta para o tema do cuidado diário e é complementado pela primeira frase do texto que assinala que cuidado é esse – a escovação.
ALTERNATIVA C	Distrator. Os estudantes que escolheram a alternativa como resposta provavelmente se fixaram na informação sobre a ideia de cuidado e associado ao tema da saúde, concluindo que o foco do texto poderia ser o cuidado com a saúde dos cães.
ALTERNATIVA D	Distrator. Os estudantes que marcaram esta alternativa provavelmente se apoiaram nas informações referentes ao fato de cães soltarem pelos e sujarem as casas, concluindo que o foco do texto é a sujeira causada por eles.

8.1)

Sugestão de gabarito e distratores:

- A. DEPOIS DE COZINHAR POR 30 MINUTOS.
- B. DEPOIS QUE A ÁGUA ESTIVER QUENTE.
- C. JUNTO COM O MOLHO PREFERIDO.
- D. JUNTO COM O SAL E A MANTEIGA.

8.2)

RESOLUÇÃO JUSTIFICADA – 3º ano do ensino fundamental	
ALTERNATIVA A	Distrator. Os estudantes que marcaram esta alternativa provavelmente se apoiam no fato de que esse é o terceiro passo da receita, na sequência do que se refere ao fubá, desconsiderando a informação solicitada no enunciado do item.
ALTERNATIVA B	Gabarito. Os estudantes que marcaram esta alternativa desenvolveram a habilidade requerida e localizaram, no segundo passo do modo de fazer, que é para colocar o fubá depois que a água estiver quente.
ALTERNATIVA C	Distrator. Os estudantes que marcaram esta alternativa provavelmente se guiaram pelo fato de ser a última informação apresentada no modo de fazer, desconsiderando a complementação do procedimento solicitado no enunciado do item.
ALTERNATIVA D	Distrator. Os estudantes que marcaram esta alternativa provavelmente se fixaram no fato de ser a primeira informação apresentada no modo de fazer, desconsiderando a informação solicitada no enunciado do item.

8.3) A habilidade de localização de informações explícitas no texto diz respeito à recuperação de informações que se encontram na superfície textual, isto é, informações claramente expressas linguisticamente, que não demandem a realização de inferências, ainda que bastante simples. A localização pode envolver procedimentos de repetição, paráfrase ou referenciamento por anáfora direta. Essa é, portanto, uma habilidade muito elementar de leitura que, em situações reais de leitura e relacionadas às práticas da vida cotidiana, é mobilizada, em geral, de forma bastante pragmática. Sem o desenvolvimento dessa habilidade o estudante não conseguirá concluir a sua alfabetização inicial e terá grandes dificuldades de progredir em seu percurso escolar.

9.1)

Veja a figura abaixo.



Disponível em: <https://br.pinterest.com/pin/365706432221475205/>. Acesso em: 10 ago. 2022.

Qual é o nome dessa figura?

- A) ENFERMEIRA
- B) ENFILEIRADA
- C) INFESTADA
- D) INVERNEIRA

9.2) O item avalia a habilidade de ler palavras formadas por sílabas não canônicas. Essa é uma habilidade que diz respeito à apropriação, pelo estudante, das convenções que organizam o sistema de escrita. Embora seja um componente pequeno do grande desafio de dominar a leitura com compreensão de textos que circulam em diferentes campos de atuação, os processos envolvidos no reconhecimento das diferentes realizações sonoras de uma letra a partir do lugar que ela ocupa na sílaba e da sílaba na palavra é fundamental para que o sistema de escrita se desenvolva. No caso deste item, a realização da tarefa requer que o estudante identifique a palavra que nomeia a figura utilizada como suporte. Para tanto, é necessário decodificar as sílabas e armazená-las na memória de trabalho para que sejam retomadas globalmente e associadas à figura, atribuindo, assim, sentido à combinação de letras na ordem em que aparecem na palavra.

Os estudantes que marcaram a alternativa A – Enfermeira –, o gabarito, identificaram corretamente a palavra que nomeia a figura apresentada entre outras que apresentam centros silábicos próximos, consoantes com semelhanças em pontos e modos de articulação e marca de oralidade na primeira sílaba.

Atividades sobre itens de língua portuguesa

10.1) Gabarito: D

10.2) Este item avalia a habilidade de reconhecer diferentes formas de tratar uma informação na comparação de textos que tratam do mesmo tema. Essa é uma habilidade que integra o currículo para os anos finais do ensino fundamental. Seu desenvolvimento requer que os estudantes já tenham desenvolvido outras habilidades de leitura, como a habilidade de identificar o assunto de um texto. No caso deste item, os dois textos oferecidos como suporte pertencem a gêneros diferentes, mas que podem circular no mesmo campo de atuação. No caso do primeiro texto, trata-se de uma notícia de jornal, que circula no campo jornalístico-midiático. O assunto do texto, a dificuldade de se locomover no trânsito em época de feriado, é apresentado já no título da notícia. O segundo suporte, que também pertence ao campo jornalístico-midiático, assim como ao artístico-literário, é uma tirinha de dois quadros, em que o segundo traz a imagem de um engarrafamento, descrita no primeiro texto, com um comentário irônico do personagem no segundo quadro, contrapondo-se ao que fora falado no primeiro quadro. O item solicita que os estudantes identifiquem qual é a informação comum aos dois textos.

Os estudantes que marcaram a alternativa B, o gabarito, demonstraram ter desenvolvido a habilidade avaliada pelo item. Aqueles que marcaram qualquer uma das outras opções de resposta optaram por tópicos que são tratados em um ou outro texto, mas não conseguiram estabelecer relações entre eles, seja porque tiveram dificuldade de reconhecer o assunto de um ou dos dois textos, seja porque não conseguiram realizar a comparação entre eles.

11) Gabarito: D

12.1) Gabarito: A

12.2) Este item avalia a habilidade de identificar elementos da narrativa, mais especificamente, o seu conflito gerador.

12.3) O suporte escolhido para o item é adequado para estudantes que já se encontram na segunda etapa do ensino fundamental, podendo ser no início ou no final dela, a depender do desenvolvimento da turma. Isso porque o texto apresenta um fragmento de um diálogo entre dois amigos, cuja linguagem é coloquial e marcada por períodos curtos, sintaxe simples e tema familiar aos estudantes avaliados. Por sua vez, tendo em vista que apresenta algumas gírias, é importante verificar se essa linguagem já condiz com o estágio de desenvolvimento dos estudantes.

12.4)

RESOLUÇÃO JUSTIFICADA	
ALTERNATIVA A	Gabarito. Os estudantes que marcaram esta alternativa identificaram corretamente que a atitude de Marcos de convidar Zeca para o cinema é o que gera os demais acontecimentos do texto, isto é, aceitar o convite, pegar o ônibus, chegar ao cinema atrasado, assistir ao filme e depois sair.
ALTERNATIVA B	Distrator. Os estudantes que marcaram esta alternativa provavelmente se concentraram no foco da ação das duas personagens, que era ir ao cinema.
ALTERNATIVA C	Distrator. Os estudantes que marcaram esta alternativa provavelmente se concentraram no acontecimento que atrapalhou o andamento das ações das duas personagens, que foi o ônibus ter se atrasado, considerando que esse poderia ter sido o evento mais importante.
ALTERNATIVA D	Distrator. Os estudantes que marcaram esta alternativa provavelmente se concentraram no momento em que Zeca aceita o convite de Marcos, pois, se Zeca rejeitasse, os demais eventos não teriam ocorrido. Porém, para que fosse aceito, foi necessário antes que Marcos fizesse o convite.
ALTERNATIVA E	Distrator. Os estudantes que marcaram esta alternativa provavelmente se concentraram no foco da ação das duas personagens, que era ir ao cinema, considerando que o momento em que chegaram lá representou o evento mais importante da narrativa.

13.1) Sugestão de alternativas:

- A) defender uma opinião.
- B) indicar um endereço.
- C) noticiar uma polêmica.
- D) transmitir um ensinamento.

13.2)

RESOLUÇÃO JUSTIFICADA	
ALTERNATIVA A	Distrator. Os estudantes que marcaram esta alternativa provavelmente se concentraram nas opiniões das pessoas entrevistadas na notícia, concluindo que o objetivo do texto era defender as posições delas.
ALTERNATIVA B	Distrator. Os estudantes que marcaram esta alternativa provavelmente se concentraram na informação que diz respeito à mudança de endereço da feira, concluindo que o objetivo do texto era comunicar essa mudança aos leitores.
ALTERNATIVA C	Gabarito. Os estudantes que marcaram esta alternativa identificaram corretamente que o objetivo do texto é noticiar um fato que provocou opiniões divergentes entre as pessoas, isto é, uma polêmica, o que é indicado por meio das posições contrárias emitidas pelas pessoas entrevistadas, bem como de forma bastante explícita já no título: "Novo endereço da Feira Hippie divide opiniões".
ALTERNATIVA D	Distrator. Os estudantes que marcaram esta alternativa provavelmente se concentraram nas opiniões das pessoas entrevistadas e considerou que elas buscavam passar algum ensinamento a respeito do episódio noticiado.

13.3) Este item avalia a habilidade de identificar a finalidade de textos de diferentes gêneros. Trata-se de uma habilidade que integra o currículo dos anos iniciais e também dos anos finais do ensino fundamental, pois apresenta diferentes níveis de complexidade, a depender do gênero textual cuja finalidade se pretende que os estudantes identifiquem. Este item tem como suporte uma notícia, gênero que circula no campo de atuação jornalístico-midiático e que, nesse exemplo, apresenta posicionamentos dos entrevistados sobre um fato. Para identificarem a finalidade desse texto, os estudantes devem compreender a função social de uma notícia, que é informar, neste caso, de for-

ma imparcial, sobre um evento que gerou opiniões diferentes entre as pessoas. Ao compreender o propósito comunicativo de diferentes textos, os estudantes se tornam mais aptos e interpretar o seu conteúdo e posicionar-se de forma ativa e crítica sobre eles.

14.1) Sugestão de item:

De acordo com esse texto, o tomate tinha má reputação, porque

- A) acreditava-se que era venenoso.
- B) era desprezado pelos europeus.
- C) era usado em molho com pimenta.
- D) parecia-se com uma cereja.

14.2) O item avalia a habilidade de estabelecer relação de causa e consequência entre partes e elementos do texto, que integra o currículo desde os anos iniciais do ensino fundamental. Seu desenvolvimento permite ao leitor produzir sentido para os textos que lê, estabelecendo a coesão sequencial entre seus elementos constitutivos. Leitores que ainda não desenvolveram essa habilidade, de modo geral, têm dificuldade de responder a perguntas sobre o porquê de fatos e/ou acontecimentos relatados ou narrados. O texto que dá suporte ao item, do campo jornalístico-midiático, é o fragmento de uma nota jornalística que apresenta uma curiosidade histórica sobre a origem dos tomates. A relação de causa/consequência que o item solicita se encontra no meio do texto, no final do segundo parágrafo. O fato de, no comando do item, haver a mesma expressão presente no texto sobre o tomate ter má reputação facilita a tarefa, visto que a consequência se encontra logo depois, isto é, os colonizadores considerarem o tomate venenoso.

Os estudantes que marcaram a alternativa A, o gabarito, realizaram, portanto, o percurso cognitivo descrito e estabeleceram, de forma adequada, a relação de causa/consequência existente nesse parágrafo do texto.

Atividades sobre itens de matemática

15.1) Esse item avalia a habilidade de o estudante utilizar a operação de subtração entre números naturais, em um contexto que envolve a comparação da diferença de quantidade de unidades que um indivíduo possui em relação a outro. O uso das operações entre números naturais pode assumir diferentes significados, que fica definido pelo contexto proposto pela situação-problema.

15.2) O gabarito deste item é a alternativa C.

15.3) O estudante deve observar as quantidades de figurinhas que Ana e Paula possuem, que são 255 e 98, respectivamente, e perceber que, para determinar a quantidade de figurinhas que Ana tem a mais do que Paula, é necessário efetuar a diferença entre essas quantidades, fazendo $255 - 98 = 157$. A maneira que o estudante procede para obter o valor 157 pode se dar por meio de um algoritmo convencional ou de estratégias pessoais.

16.1) Sugestão de comando: Quantos graus Celsius a mais a temperatura ambiente do centro de pesquisa estava superior à temperatura registrada no interior da câmara fria nesse instante?

16.2) Gabarito: 39°C .

16.3) Sugestão de alternativas:

- A) -39°C
- B) -3°C
- C) 3°C
- D) 39°C

16.4)

RESOLUÇÃO JUSTIFICADA	
ALTERNATIVA A	Distrator. O estudante provavelmente procurou efetuar a diferença entre as temperaturas, porém subtraiu a temperatura ambiente da temperatura da câmara fria, fazendo: $(-18^{\circ}\text{C}) - (21^{\circ}\text{C}) = -39^{\circ}\text{C}$
ALTERNATIVA B	Distratir. O estudante provavelmente procurou efetuar a diferença entre as temperaturas, porém, equivocou-se com o sinal da temperatura da câmara fria, subtraindo a temperatura da câmara fria da temperatura ambiente, ao fazer: $18^{\circ}\text{C} - 21^{\circ}\text{C} = -3^{\circ}\text{C}$
ALTERNATIVA C	Distrator. O estudante provavelmente procurou efetuar a diferença entre as temperaturas, porém ignorou o sinal da temperatura da câmara fria, que é -18°C , e efetuou a subtração entre a temperatura ambiente e a temperatura da câmara fria, obtendo: $21^{\circ}\text{C} - 18^{\circ}\text{C} = 3^{\circ}\text{C}$
ALTERNATIVA D	Gabarito. O estudante efetuou corretamente a diferença entre as temperaturas, subtraindo a temperatura da câmara fria da temperatura ambiente, considerando corretamente os sinais e obtendo: $21^{\circ}\text{C} - (-18^{\circ}\text{C}) = 21^{\circ}\text{C} + 18^{\circ}\text{C} = 39^{\circ}\text{C}$

17.1) Interpretar dados dispostos em um gráfico de linhas.

17.2) Sugestão de alternativas:

- A) 31
- B) 37
- C) 60
- D) 69

17.3)

RESOLUÇÃO JUSTIFICADA	
ALTERNATIVA A	<p>Distrator. O estudante provavelmente considera corretamente apenas as quantidades de medalhas olímpicas ganhas pelo Brasil nas olimpíadas realizadas nos anos 2000, 2004 e 2008, somando todas as medalhas de ouro, prata e bronze. Porém, registra somente uma vez a quantidade de 6 medalhas ganhas no ano 2000, não percebendo que naquele ano foram ganhas 6 medalhas de prata e 6 medalhas de bronze. Sendo assim, calcula:</p> $6 + 0 + 5 + 3 + 2 + 8 + 4 + 3 = 31$
ALTERNATIVA B	<p>Gabarito. O estudante considera corretamente apenas as quantidades de medalhas olímpicas ganhas pelo Brasil nas olimpíadas realizadas nos anos 2000, 2004 e 2008, somando todas as medalhas de ouro, prata e bronze. Sendo assim, calcula:</p> $6 + 6 + 0 + 5 + 3 + 2 + 8 + 4 + 3 = 37$
ALTERNATIVA C	<p>Distrator. O estudante provavelmente considera as quantidades de medalhas olímpicas ganhas pelo Brasil em todas as olimpíadas registradas no gráfico, ou seja, nos anos 1996, 2000, 2004, 2008 e 2012, somando todas as medalhas de ouro, prata e bronze. Porém, leva em conta em todas as situações em que duas linhas do gráfico se encontram, registrando somente uma vez a quantidade de medalhas, o que ocorre em 1996 e em 2000. Sendo assim, calcula:</p> $9 + 3 + 6 + 0 + 5 + 3 + 2 + 8 + 4 + 3 + 9 + 5 + 3 = 60$
ALTERNATIVA D	<p>Distrator. O estudante provavelmente considera as quantidades de medalhas olímpicas ganhas pelo Brasil em todas as olimpíadas registradas no gráfico, ou seja, nos anos 1996, 2000, 2004, 2008 e 2012, somando todas as medalhas de ouro, prata e bronze. Sendo assim, calcula:</p> $9 + 3 + 3 + 6 + 6 + 0 + 5 + 3 + 2 + 8 + 4 + 3 + 9 + 5 + 3 = 69$

18.1) Comando: A tabela que representa os dados informados nesse gráfico é

Alternativas de resposta:

A)	Atribuição	Porcentagem
Treinamentos		8%
Tarefa da Diretoria		11%
Reunião		26%
Projeto externo		20%
Projeto interno		25%
Outros		10%

B)	Atribuição	Porcentagem
Treinamentos		8%
Tarefa da Diretoria		11%
Reunião		26%
Projeto externo		25%
Projeto interno		20%
Outros		10%

Atribuição	Porcentagem
Treinamentos	10%
Tarefa da Diretoria	25%
Reunião	20%
Projeto externo	26%
Projeto interno	11%
Outros	8%

Atribuição	Porcentagem
Treinamentos	10%
Tarefa da Diretoria	25%
Reunião	20%
Projeto externo	11%
Projeto interno	26%
Outros	8%

O Gabarito é a alternativa C.

18.2) Se um funcionário trabalha 40 horas por semana nessa empresa, quantas horas de sua jornada semanal de trabalho são dedicadas a reuniões e treinamentos?

Gabarito: 12 horas (= 30% de 40 horas).

Atividades do Capítulo 3

Atividades de conhecimentos gerais sobre itens de resposta construída

1) Gabarito: C

2) Gabarito: B

3) Gabarito: D

Atividades sobre itens de resposta construída de alfabetização

4.1) Escrever palavra formada por sílabas canônicas.

4.2) Sim. Pois a avaliação em fase inicial do processo de alfabetização deve trazer informações sobre como os estudantes registram graficamente o processo de trabalho cognitivo, possibilitando que seja observado como diferenciam as unidades que formam a palavra – sílabas/letras. Sendo assim, a palavra é adequada, pois permite a realização desse processo avaliativo.

4.3) Trata-se de uma palavra dissílaba, familiar, paroxítona, formada por sílabas no padrão consoante/vogal, ou seja, canônicas.

4.4) Sim. A imagem representa um copo, que não apresenta denominações regionais diferentes e possibilita que sejam avaliados os aspectos anteriormente descritos. Caso fosse uma palavra ditada, poderia incorrer na questão da pronúncia “guiada”, com o aplicador criando um ambiente artificial no que se refere à pronúncia da letra O da sílaba final.

4.5) Alternativa B. A criança que fez esse registro reproduziu corretamente as sílabas da palavra canônica – consoante/vogal, representando graficamente o som ouvido. Porém, cometeu desvio em relação à última letra.

4.6) O objetivo das intervenções didáticas seria o de promover a passagem da criança do nível alfabético ao ortográfico, o que poderia acontecer por meio da exploração da palavra em tela e de outras em que ocorre a mesma relação entre som ouvido/representação gráfica. Isso poderia se dar por meio de jogos como o da memória, dominó e alfabeto móvel. Em relação esse último, a identificação de palavras com essas realizações sonoras em textos de diferentes gêneros também contribuiria para que os estudantes vivessem experiências nas quais poderiam pensar sobre diferenças e semelhanças na grafia das palavras.

5.1)

ASPECTO 1 – PLAUSIBILIDADE COM A IMAGEM	
A	O PALHAÇO DORME NA BOLA. O PALHAÇO DORME EM CIMA DA BOLA. O PALHAÇO DESCANSA NA BOLA. O PALHAÇO DORMINHOCA. O PALHAÇO SONOLENTO.
B	O PALHAÇO ESTÁ ROLANDO NA BOLA. O PALHAÇO BRINCALHÃO.
C	A FLOR ROLOU DO CHAPÉU. MENINO JOGANDO BOLA. VOANDO NO ESPAÇO.

5.2) Não seria possível, pois não haveria na frase elementos que possibilitassem reconhecer a plausibilidade da resposta construída a partir da leitura da imagem. A própria observação da chave de correção sobre a letra anterior – D – evidencia essa impossibilidade de cruzamento das informações sobre o processo de construção do registro escrito, a saber: “a frase que não contiver verbo será avaliada, automaticamente, com o conceito D nesse aspecto, independentemente da avaliação atribuída no Aspecto 1”.

Atividades sobre itens de resposta construída de língua portuguesa

6.1) Ambos os textos se aproximam por abordarem temática relacionada ao uso e à adequação de um tipo de linguagem e por utilizarem o recurso do humor. Contudo, o primeiro texto pertence ao gênero charge, situando-se, conforme a BNCC, no campo de atuação artístico-literário. Ao fazer uso de uma linguagem mista, verbal e não-verbal, denominada multissemiótica, o primeiro texto relaciona-se a um acontecimento recente e específico, sendo, portanto, de caráter efêmero, precisando de uma contextualização para a sua real compreensão.

Por sua vez, o segundo texto pode ser considerado um conto – mais comumente chamado de “causo” –, inserindo-se, conforme a BNCC, no campo da vida cotidiana. Esse gênero textual narrativo não faz uso de imagens e caracteriza-se por apresentar fatos reais ou fictícios, os quais, nesse exemplo, são contados a fim de produzir efeitos de humor, aproximando-se da anedota. No caso específico desse texto, há uma referência a uma personalidade e ao fato de que se trataria de uma situação real, entretanto, não se faz necessário conhecer o personagem mencionado, tampouco o contexto histórico da narrativa, para que o texto seja compreendido e provoque efeitos de riso.

6.2) Todos os itens, do A ao C, são de resposta construída subjetiva. Isso porque todos eles exigem do respondente a capacidade de realizar um raciocínio crítico, articulando informações, para poder se expressar por meio da produção da escrita.

6.3)

- A) Reconhecer o gênero de textos diversos.
- B) Identificar as marcas linguísticas que evidenciam o locutor e o interlocutor de um texto.
- C) Estabelecer relação causa/consequência entre partes e elementos do texto.

6.4) Visto que responder ao item pressupõe a integração das informações constantes no texto, o estudante precisa ser capaz de: ler o texto do início até o fim; ler com cuidado o comando para a resposta; compreender que o termo “por que” remete à busca pela causa da pergunta feita pelo ladrão; e entender e operar com o conceito de causa e consequência.

6.5) Há várias possibilidades de se corrigir o item A. Apresentamos, a seguir, um exemplo de chave de correção que pode ser utilizada, a qual se divide entre a capacidade do estudante de produzir o conteúdo esperado e de escrever um texto de acordo com critérios coesivos e ortográficos. Neste caso, o item vale 2 pontos.

Chave de correção – Item A

Escrita	
Nível I – 0 pontos	O estudante produz um texto sem articulação entre suas partes, configurando uma lista de palavras e/ou frases soltas, de modo que o conteúdo não é compreensível. Ao alcançar este nível, o estudante zera toda a atividade.
Nível II – 0,5 pontos	O estudante produz um texto compreensível, porém ainda com problemas de articulação entre as partes, bem como desvios ortográficos.
Nível III – 1 ponto	O estudante produz um texto compreensível, com pouco ou nenhum desvio ortográfico e/ou problema de articulação entre as partes.
Conteúdo	
Nível I – 0 pontos	O estudante afirma que os textos pertencem ao mesmo gênero textual, não conseguindo diferenciar o formato e a proposta comunicativa de cada um. Ao alcançar este nível, o estudante zera toda a atividade.
Nível II – 0,5 pontos	O estudante reconhece que os dois textos pertencem a gêneros distintos, podendo chegar até a compreender um dos dois, mas ainda não estabelece uma diferenciação clara entre cada um.
Nível III – 1 ponto	O estudante reconhece que os dois textos pertencem a gêneros distintos, tendo em vista que o primeiro é uma charge e o segundo, um conto, ou “causó”, ou anedota. Uma primeira característica que diferencia ambos os textos é pode ser destacada pelo estudante é o fato de a charge fazer uso de uma linguagem mista, verbal e não-verbal, enquanto o conto se caracteriza por apresentar uma narrativa puramente verbal. Uma segunda diferença diz respeito à linguagem verbal empregada, dado que a charge em questão recorre ao chamado “internetês”, uma linguagem improvisada produzida geralmente por jovens no meio virtual; por sua vez, esse conto em específico faz uso de um vocabulário culto, podendo até ser considerado excessivamente formal.

Atividades sobre itens de resposta construída de matemática

7.1) Item proposto:

Efetue a divisão de 216 por 18, registrando todo o seu raciocínio.

7.2) Espera-se que os estudantes consigam efetuar a divisão, utilizando algum algoritmo ou estratégias pessoais, conforme exemplificado a seguir.

Solução 1: Crédito integral

$$\begin{array}{r}
 21'6 \\
 - 18 \\
 \hline
 36 \\
 - 36 \\
 \hline
 0
 \end{array}
 \quad \boxed{18}$$

Solução 2: Crédito integral

$$\begin{array}{r}
 18+18+18+18+18+18+18+18+18+18 = 180 \\
 \hline
 10 \qquad \qquad \qquad +18 \\
 \hline
 198 \\
 +18 \\
 \hline
 216
 \end{array}$$

Essa resposta sugere que o estudante comprehende o significado da divisão e consegue implementar estratégias pessoais de cálculo coerentes com esse significado. No entanto, ele ainda não utiliza a multiplicação para otimizar suas totalizações parciais. Embora a estratégia seja correta e permita encontrar o valor desejado, ela se revela pouco eficiente por ser pouco generalizável. Com isso, pode-se conceder crédito integral para situações em que a estratégia pessoal permite alcançar o resultado desejado. Entretanto, é importante revisitar o algoritmo da divisão com o estudante e analisar com a eficiência do método por ele empregado e o oferecido pelo algoritmo, apresentando-lhe outras operações.

Solução 3: Crédito integral

$$\begin{array}{r}
 10 \times 18 = 180 \\
 1 \times 18 = 18 \\
 \hline
 1 \times 18 = 18 \\
 \hline
 12 \times 18 = 216
 \end{array}$$

Essa resposta tem estrutura similar à anterior, sendo-lhe superior por fazer uso da multiplicação como forma de agilizar os procedimentos. Valem as mesmas observações feitas para a Solução 2.

Solução 4: Crédito parcial

$$\begin{array}{r}
 21'6 \quad | \quad 18 \\
 - 18 \quad \quad \quad 102 \\
 \hline
 3 \\
 - 0 \\
 \hline
 36 \\
 - 36 \\
 \hline
 0
 \end{array}$$

Essa resposta sugere que o estudante conhece o algoritmo tradicional, pois busca implementá-lo. No entanto, não consegue aplicá-lo corretamente, cometendo erro em uma das etapas do algoritmo, ao dividir o resto da primeira etapa pelo divisor, obtendo o algarismo zero no dividendo. É importante, nesse caso, trabalhar com o significado de cada passo do algoritmo, sempre em conexão com as propriedades do sistema de numeração e recomendar que o estudante sempre confira a divisão efetuada. Nesse caso, se ele efetuasse o produto entre 18 e 102, perceberia que o resultado obtido estaria incorreto. Merece crédito parcial toda resolução que, ao fazer uso do algoritmo, comete um único erro, em alguma de etapa de sua execução.

Solução 5: Nenhum crédito

Toda resposta que se ampara em algoritmos totalmente incorretos ou respostas sem nenhum registro do raciocínio adotado pelo estudante.

8) Item e considerações propostos:

Marcelo efetuou corretamente a divisão de 216 por 18.

O quociente encontrado por Marcelo foi

- A) 12
- B) 13
- C) 100
- D) 102

Em um item de resposta selecionada, quando um estudante o acerta, não é possível saber se ele já domina o algoritmo da divisão e o emprega eficientemente, ou se ele se vale de estratégias pessoais, nem quais seriam essas estratégias pessoais por ele empregadas. Dessa forma, o diagnóstico que itens de resposta selecionada podem fornecer sobre o estudante é menos preciso do que o diagnóstico que um item de resposta construída é capaz de oferecer. Num item de resposta construída, o percurso feito pelo estudante, ao ser registrado, nos permite observar melhor o estágio de desenvolvimento que ele se encontra em relação aos saberes mobilizados na sua solução, qual a natureza dos tipos de erros ainda cometidos, que processos ainda não foram dominados, quais conceitos ainda não foram consolidados etc.

9.1) Como se deseja conhecer os preços desses tipos de lápis e de caneta, vamos representá-los por: x o preço do lápis e y o preço da caneta em reais.

A frase “Ana comprou dois lápis e uma caneta e pagou R\$ 12,00” pode ser interpretada por:

$$2x + y = 12$$

Já a frase “Beatriz comprou três lápis e duas canetas, idênticas às compradas por Ana, e pagou R\$ 21,00” pode ser interpretada por:

$$3x + 2y = 21$$

Para se obter os valores de x e de y que satisfaçam as duas equações, torna-se necessário resolver o seguinte sistema de equações lineares:

$$\begin{cases} 2x + y = 12 \\ 3x + 2y = 21 \end{cases}$$

A partir desse ponto, deve-se empregar algum dos métodos conhecidos para resolver o sistema. Nessa etapa de escolaridade, os métodos mais prováveis de serem empregados pelos estudantes são: método da substituição e método da adição.

Resolveremos pelo método da substituição, isolando y na primeira equação e substituindo-o na segunda equação:

$$2x + y = 12 \Rightarrow y = 12 - 2x$$

$$3x + 2y = 21 \Rightarrow 3x + 2(12 - 2x) = 21 \Rightarrow 3x + 24 - 4x = 21 \Rightarrow -x = -3 \Rightarrow x = 3$$

Substituindo esse valor de x na equação: $y = 12 - 2x$, obtém-se o valor de y :

$$y = 12 - 2(3) = 12 - 6 = 6$$

Dessa forma, pode-se concluir que, nessa papelaria, o preço de um lápis é R\$ 3,00 e o de uma caneta é R\$ 6,00.

9.2) Esse item avalia a seguinte habilidade: utilizar sistema de equações lineares, com duas equações e duas incógnitas, na resolução de problema.

9.3) Proposta de chave de correção:

CHAVE DE CORREÇÃO	
Crédito integral	Resoluções construídas a partir de uma correta modelagem da situação proposta e do emprego de algum dos métodos resolutivos de sistema de equações lineares que alcancem a resposta esperada.
Crédito parcial	Efetuar uma modelagem algébrica da situação proposta, alcançando o sistema, mas errando em alguma etapa de sua resolução.
Crédito parcial	Efetuar uma modelagem algébrica da situação proposta, alcançando o sistema, mas não encaminhando sua resolução.
Crédito parcial	Apresentar os valores corretos para os preços e verificar aritimeticamente que esses valores apresentados atendem à situação proposta.
Crédito parcial	Obter, por tentativa e erro, os valores corretos para os preços do lápis e da caneta, com registro dessas tentativas.
Nenhum crédito	Outras situações incorretas e sem significados importantes ou em branco.

Atividades do Capítulo 4

1.1)

PROFI CIÊNCIA	N	%	%_ACUMULADO
0	10	5	5
1	20	10	15
2	20	10	25
3	40	20	45
4	50	25	70
5	20	10	80
6	20	10	90
7	20	10	100

1.2) 25º percentil.

1.3) 25%

1.4) 4

1.5) 80%

1.6) 5

2.1) Sinal positivo em ambos os gráficos.

2.2) O teste B seria o escolhido, pois sua correlação com o construto (medido no eixo vertical) é maior, conforme é possível verificar pela menor dispersão dos dados (pontos) em torno da reta que sintetiza a associação entre as duas variáveis em questão.

3) Caso 1: problema de fidedignidade instrucional. O teste tenderá a desfavorecer os estudantes da escola B, pois não terão tempo de aprender parte das habilidades que estão sendo avaliadas. Para corrigir esse problema, uma possibilidade seria fazer com que as escolas seguissem um mesmo cronograma, no que diz respeito à apresentação dos diferentes conteúdos das áreas de conhecimento avaliadas. Outra possibilidade seria aplicar o teste no início de dezembro, quando os estudantes de ambas as escolas terão tido tempo de serem apresentados a todo o conteúdo.

Caso 2: problema de fidedignidade de correção. Alguns estudantes poderão ser prejudicados ao terem a sua redação corrigida por um professor ou uma professora mais exigente do que a média, e vice-versa. Para resolver esse problema, uma possibilidade seria promover uma chave de correção geral e treinamento aos corretores. Além disso, poder-se-ia fazer com que cada redação fosse corrigida por dois ou mais professores, de modo a se tirar uma média dos resultados.

4.1)

Item	Dificuldade (proporção de acertos)
1	0,34
2	0,725
3	0,14
4	0,53

Os resultados acima também poderiam ser expressos em termos de percentuais, basta multiplicar os valores da última coluna por 100.

4.2) Os itens 2 e 4, pois têm os maiores percentuais de acerto.

4.3) Os itens 1 e 3, pois têm os menores percentuais de acerto.

4.4)

Item	Proporção de acerto		
	Grupo 1	Grupo 2	Diferença
1	0,11	0,57	0,46
2	0,66	0,79	0,13
3	0,09	0,19	0,1
4	0,22	0,84	0,62

Os resultados acima também poderiam ser expressos em termos de percentuais, basta multiplicar os valores por 100.

4.5) Os itens 1 e 4, pois têm os maiores coeficientes de discriminação: 0,46 e 0,62, respectivamente.

4.6) A capacidade de discriminação tende a ser menor nos extremos de dificuldade, como no item 2, que é o mais fácil, e no item 3, que é o mais difícil. Por outro lado, a discriminação tende a se elevar nos itens de dificuldade média, que são os itens 1 e 4.

5.1) Pedro: 1; Paula: 3; e Francisco: 5

5.2) Maurício provavelmente teria acertado a questão 1 e errado a 4 e a 5, visto que os itens estão dispostos em uma escala cumulativa de dificuldade. Segundo essa escala, e com base nas informações fornecidas, Maurício teve a sua respectiva proficiência localizada no ponto 2.

6.1) Modelo TRI de 1 parâmetro, porque todos os itens têm assíntotas em $P=0$ e $P=1$ e também têm aparentemente o mesmo parâmetro de discriminação, isto é, inclinações iguais.

6.2) O item 1, porque se encontra mais à esquerda no gráfico. Ou seja, a probabilidade P de acerto de 50% corresponde a uma proficiência mais baixa que as demais, na escala horizontal.

6.3) O item 5, porque se encontra mais à direita no gráfico. Ou seja, a probabilidade P de acerto de 50% corresponde a uma proficiência mais alta que as demais, na escala horizontal.

6.4) Todos têm a mesma discriminação, conforme é possível verificar por suas respectivas formas ou CCIs, que são idênticas, excetuando-se pelo deslocamento horizontal.

6.5) Aproximadamente -2,5; -1,5; 0; 1 e 2, para os itens de 1 a 5, respectivamente. Esses são os valores das proficiências para as quais $P=50\%$ em cada item.

6.6) O item f, que possui uma maior inclinação para esse valor da proficiência.

7.1) Modelo TRI de 3 parâmetros, pois nitidamente possui uma assíntota inferior maior que zero.

7.2) O item 1, pois sua tangente neste ponto é a mais inclinada de todas.

7.3) O item 2, pois sua tangente neste ponto é a mais inclinada de todas.

7.4) O item 3, pois sua tangente neste ponto é a mais inclinada de todas.

7.5) Seria mais conveniente aplicar o item 1 ao grupo de desempenho baixo, porque esse item possui maior capacidade de discriminação na faixa inferior de proficiência. Já o item 2 poderia ser aplicado ao grupo de desempenho médio, e o item 3, ao grupo de desempenho alto, por motivos análogos.

SOBRE OS AUTORES

Carlos Palacios

Doutor e mestre em Letras Vernáculas pela Universidade Federal do Rio de Janeiro (UFRJ) e bacharel em Comunicação Social pela Universidade Federal de Juiz de Fora (UFJF). Atualmente, é pesquisador do Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF).

Girlene Ribeiro de Jesus

Professora da Faculdade de Educação da Universidade de Brasília (FE/UnB), membra do corpo docente do Programa de Pós-graduação em Educação e coordenadora do Mestrado Profissional. Possui doutorado e mestrado pela mesma instituição, na área de Instrumentos e Medidas Psicológicas e Educacionais. Atualmente, é uma das conselheiras da Associação Brasileira de Avaliação Educacional (ABAVE) e faz parte da comissão assessora de psicometria do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

Glauker Amorim

Professor do Departamento de Matemática da Universidade Federal de Juiz de Fora (UFJF) e pesquisador do CAEd/UFJF, com doutorado e mestrado em Matemática pela Universidade Federal do Rio de Janeiro (UFRJ).

Hilda Micarello

Professora associada da Faculdade de Educação da UFJF, onde atua no curso de Licenciatura em Pedagogia, no Programa de Pós-Graduação em Educação e no Programa de Pós-graduação Profissional em Gestão e Avaliação da Educação Pública (PPGP). Tem pós-doutorado em Educação pela PUC-Rio e realiza pesquisas nos campos da linguagem, infâncias e formação de professores, atuando em projetos e pesquisas relacionados à avaliação em larga escala e à gestão da educação pública desenvolvidos pelo CAEd/UFJF.

Joaquim José Soares Neto

Professor titular da UnB, com graduação e mestrado na mesma instituição, e doutorado na Aarhus University (Dinamarca) e pós-doutorado no California Institute of Technology (Caltech, EUA). Foi Presidente do INEP e relator no Conselho Nacional de Educação da Base Nacional Comum Curricular (BNCC). Atualmente, é Conselheiro e Presidente da Câmara de Educação Superior do Conselho Nacional de Educação (CNE).

Josiane Toledo

Doutora em Educação pela PUC-Rio, com período sanduíche na Université Paris-Nord – Paris 13, mestre em Letras pelo Centro de Ensino Superior de Juiz de Fora e licenciada em Letras – Português e Francês pela UFJF. Atualmente, é Supervisora de Apoio à entrega de resultados de avaliações educacionais no CAEd/UFJF e membra do comitê diretor e grupo de trabalho de conteúdo FLIP+ e-assessment Community, como representante do CAEd. Desde 2004, atua como analista e especialista em avaliação educacional, com foco no desenvolvimento de instrumentos avaliativos e ações para devolutivas e desenvolvimento profissional de professores e gestores para uso pedagógico dos resultados de avaliações.

Lina Kátia Mesquita de Oliveira

Doutora em Educação pela PUC-Rio, com mestrado em Educação pela UFJF e graduação em Matemática pela Faculdade de Filosofia, Ciências e Letras de Belo Horizonte. Atualmente é Diretora Executiva da Fundação CAEd, Vice-diretora da ABAVE e membra do Conselho Estadual de Educação de Minas Gerais. Dedica-se principalmente à implementação de sistemas de avaliação da educação básica, com foco em escalas de proficiência e educação matemática.

Luis Antônio Fajardo

Doutor e mestre em Educação e graduado em Física pela UFJF e especialista em métodos quantitativos de pesquisa social pela Universidade de Michigan (EUA). Atualmente, é analista sênior de dados educacionais do CAEd/UFJF e professor adjunto do Programa de Pós-graduação Profissional em Gestão e Avaliação da Educação Pública (PPGP) na mesma instituição. Sua área de atuação concentra-se em indicadores educacionais e em modelagem linear hierárquica aplicada aos resultados das avaliações.

Manuel Palacios Cunha Melo

Doutor em Sociologia pelo Instituto Universitário de Pesquisas do Rio de Janeiro (IUPERJ), é professor da Faculdade de Educação da UFJF e coordenador geral do CAEd, onde lidera a implementação de sistemas de avaliação e gestão em diversas redes públicas de ensino estaduais e municipais. Foi também Secretário de Educação Básica do Ministério da Educação.

Rosângela Veiga

Doutora em Educação pela UFJF e professora EBTT do Colégio de Aplicação João XXIII/UFJF. Atualmente, é vice-líder do grupo de pesquisa Linguagens, Infâncias e Educação (FACEd/UFJF) e pesquisadora do CAEd/UFJF. Desenvolve pesquisas sobre formação do leitor, linguagens e infâncias; avaliação em larga escala e alfabetização; e formação de professores e gestores.

Wellington Silva

Doutor em Educação pela PUC-Rio, mestre em Educação pela UFJF e graduado em Engenharia Elétrica pela mesma instituição. Atualmente, é coordenador de medidas educacionais do CAEd/UFJF, atuando, desde 2002, como psicometrista em análises de bases de dados com ênfase em medidas educacionais pela Teoria da Resposta ao Item (TRI) e modelagens multiníveis.



Investir na avaliação educacional é acreditar que o futuro de nossas crianças e de nossos jovens não está relegado à condição social e à sorte de cada um. Afinal, não se avalia o desenvolvimento de um estudante senão com o objetivo principal de identificar os obstáculos que ainda existem em seu caminho, de modo a se criar condições para que consiga vencê-los. Isso implica, por sua vez, que os profissionais e as instituições envolvidas confiem na escola como um espaço capaz de promover a superação desses desafios.

Lina Kátia Mesquita de Oliveira