

Avaliação de larga escala em leitura: considerações sobre construto e método de avaliar no SAEB e no PISA

Adriana de Oliveira Barbosa
Universidade Estadual de Campinas
Instituto de Estudos e Pesquisas Educacionais Anísio Teixeira
adriarbarbosa19@gmail.com

Resumo

O presente trabalho investiga o construto de leitura de duas avaliações externas de larga escala presentes no cenário educacional brasileiro: o Sistema de Avaliação da Educação Básica (SAEB) e o Programa Internacional de Avaliação de Estudantes (PISA), com foco em seus métodos de avaliar (ou nos formatos dos testes). O objetivo é contribuir com subsídios para melhor entendimento da leitura no contexto da avaliação, considerando que ambas avaliações têm potencial de influenciar não só as decisões políticas, como também as práticas pedagógicas. A análise tem aporte teórico da Linguística Aplicada, na área de avaliação em contexto de línguas, tem caráter qualitativo e baseia-se na metodologia de análise documental. Evidenciaram-se as diferenças nos construtos das duas avaliações, que ficam acentuadas pelos formatos dos testes: restrito a itens objetivos de múltipla escolha no SAEB e variado no PISA. Nesse sentido, remarca-se que o teste internacional tem maior proximidade a situações e demandas de leitura mais condizentes com práticas reais, o que não se verifica na avaliação nacional.

Palavras-chave: leitura; avaliação em larga escala; SAEB; PISA; análise de construto; métodos de avaliar

Abstract

The present work investigates the reading construct of two large scale external assessments present in the Brazilian educational context The Brazilian System of Assessment for Basic Education (SAEB) and the Programme for International Student Assessment (PISA), focusing on its testing methods (or tests formats). The objective is to contribute to a better understanding of reading in the assessment context, considering that both assessments have the potential to influence not only the political decisions but also the pedagogical practices. The theoretic basis comes from Applied Linguistics, in the Language Assessment field. It has qualitative

approach and based on the document analysis method. The differences in the constructs of the two assessment were emphasized, which are reinforced by the test formats: restricted to multiple choice items in SAEB and varied in PISA. In this sense, it should be noted that the international assessment presents greater proximity to reading situations and demands that are more consistent with real practices, which is not the case in the national assessment.

Keywords: Reading; Large scale assessment; SAEB; PISA, construct analysis; test methods

1 Introdução

As avaliações externas em larga escala de sistemas educacionais têm feito parte do cenário brasileiro há algumas décadas, gerando consequências na esfera das políticas educacionais. Duas dessas avaliações são o Sistema de Avaliação da Educação Básica (doravante SAEB), em nível nacional, e o Programa Internacional de Avaliação de Estudantes (doravante PISA), em nível internacional¹. Os testes em Leitura e Matemática (além de Ciências, no caso do PISA) medem, periodicamente, habilidades dos estudantes, aos quais são vinculados, também, dados socioeconômicos, com intuito de contextualizar essas medidas, para traçar diagnósticos sobre o rendimento dos estudantes ao longo do processo de escolarização da educação básica. Ambas as avaliações foram incorporadas à política educacional brasileira, respaldando ideias por vezes generalizadoras sobre a baixa qualidade da educação brasileira e figurando, inclusive, como metas no Plano Nacional de Educação vigente (PNE 2014-2024)².

¹ O SAEB é uma avaliação federal, em conjunto com estados e municípios, de caráter censitário para a rede pública e amostral para a rede privada, sob responsabilidade do Instituto Educacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), órgão vinculado ao Ministério da Educação do Brasil (MEC). O PISA é uma avaliação internacional comparativa, de caráter amostral, aplicada em mais de 60 países, desenvolvido pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE).

² A Lei nº 13.005, de 25 de junho de 2014, aprova o Plano Nacional de Educação – PNE e estabelece metas e estratégias de melhoria do fluxo e da aprendizagem, de modo a impactar na meta nacional do Índice de Desenvolvimento da Educação Básica (IDEB), a partir do SAEB, bem como nas médias dos resultados nacionais no PISA, até 2021 em ambos os casos.

No que diz respeito aos testes na área de Linguagem, ambas avaliações³ têm como foco a leitura, envolvem público similar (especialmente se considerarmos o 9º ano do ensino fundamental)⁴, organizam-se em torno de questões de caráter cognitivo, padronizadas e parametrizadas pela mesma teoria estatística, a Teoria de Resposta ao Item (TRI), além de apresentarem seus resultados em escalas de níveis de proficiência. Entretanto, naturalmente apresentam diferenças em termos de construto e impacto educacional. Os métodos de avaliar, ou formato dos testes, é uma delas.

Discussão anterior de Bonamino, Coscarelli e Franco (2002), referente às edições iniciais das duas avaliações, analisou a concepção de letramento a elas subjacente, ressaltando o caráter de letramento individual do SAEB e social do PISA, nos termos de Magda Soares (1999), mostrando tanto a presença de maior variedade de gêneros textuais no PISA quanto a exigência de análise crítica da linguagem, ausente no SAEB. Naquela discussão, os autores mencionam o aspecto do formato das questões dos testes como fator interveniente. (BONAMINO; COSCARELLI; FRANCO, 2002, p. 110).

Essas diferenças parecem se acentuar ao longo dos anos, considerando que o SAEB continua a se realizar nos mesmos moldes desde 2001, enquanto o PISA é revisto a cada edição trienal de avaliação. No caso de leitura, as edições em que essa área esteve em foco foram: 2000; 2009 e 2018. O foco em determinada área significa que ela ocupa a maior parte da avaliação e a investigação contextual daquela edição, com atualização de seus referenciais e de itens, preenchendo aproximadamente metade do tempo do teste. As demais áreas avaliadas ocupam a outra metade, apenas com itens já aplicados anteriormente (BRASIL, 2019, p. 22). Além disso, o PISA vem sendo aplicado em formato eletrônico desde 2015 no Brasil, em uma

³ Embora os termos “avaliação” e “teste” se situem dentro do mesmo contexto geral e muitas vezes sejam usados indistintamente, neste texto, avaliação será usada na referência mais ampla de projeto, envolvendo diversos instrumentos, e teste se referirá ao instrumento formal definido pelo construto de leitura.

⁴ O SAEB é abarca toda a educação básica, tendo como foco os anos finais dos ciclos de formação: 5º ano, 9º ano do ensino fundamental e 3ª série do ensino médio; já o PISA tem como foco o recorte por idade, 15 anos, o que define o público avaliado no Brasil, dado o nosso contexto, entre os alunos do 7º ano do ensino fundamental até a 3ª série do ensino médio.

transição do formato impresso, e, desde 2018, produzido completamente para o meio eletrônico.

Em relatório produzido pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) sobre resultados do PISA 2015, uma análise comparativa entre PISA e SAEB – baseada em seus marcos referenciais – aponta maior abrangência da avaliação de leitura na primeira que na segunda, embora se considere que ambas sejam voltadas para a leitura para além de decodificação, abrangendo diferentes gêneros textuais, a situação de comunicação e espaços de circulação social. Isso se daria especialmente na ausência de foco em posicionamento crítico no SAEB. (OCDE, 2016, p. 253-254).

Nesse contexto, a Linguística Aplicada, especialmente a área de avaliação no contexto de línguas, fornece importantes subsídios para ampliar as discussões em torno dessas avaliações, dada a relevância que têm adquirido nos contextos educacionais ao redor do mundo. Um dos aspectos proeminentes gira em torno do entendimento dos construtos medidos nos testes aplicados em sua relação com bases epistemológicas e metodológicas. Neste artigo, objetiva-se esse olhar sobre os construtos de leitura das duas avaliações, em especial para os formatos dos testes, que os operacionalizam.

2 Avaliação de leitura em larga escala: perspectivas teóricas e metodológicas

A leitura é um dos pilares do ensino, basilar para o desenvolvimento de outras capacidades e conhecimentos, para a formação discursiva do leitor e sua atuação no mundo letrado. Por isso ela compõe a maioria dos programas de avaliação em larga escala na área de linguagem. No âmbito da avaliação, o foco maior tem se dado na compreensão de leitura, que envolve, em definição geral, a concepção de várias habilidades de componentes linguísticos, fontes de conhecimento e competências (*abilities*) mais gerais, cujo uso é variável de acordo com os contextos e propósitos (GRABE; JIANG, 2013, p. 4).

Muito se tem acumulado em termos de pesquisas sobre leitura, em consonância com o próprio desenvolvimento da ciência e das próprias práticas sociais de leitura. Grabe (2009) nos apresenta uma síntese de seis objetivos principais de leitura ligados ao contexto educacional atualmente. São eles:

*Objetivos acadêmicos para leitura*⁵:

1. *Leitura para procurar informação (scanning e skimming)*
2. *Leitura para entendimento rápido (skimming)*
3. *Leitura para aprender*
4. *Leitura para integrar informação*
5. *Leitura para avaliar, criticar e usar informação*
6. *Leitura para compreensão geral (em muitos casos, ler por interesse ou por diversão) (GRABE, 2009, p. 8, minha tradução).*

Esses objetivos revelam as inúmeras demandas e interesses envolvidos atualmente na leitura e trazem implicações tanto para o ensino quanto para a avaliação.

Uma das características centrais envolvidas em testes padronizados refere-se ao fato de se constituírem como avaliação de *produto*, que diz respeito ao resultado de um *processo* que é individual, interno e variado a cada leitura, implicado na dinâmica entre leitor e texto. Enquanto decodifica, o leitor também reflete sobre o que lê, se posiciona sobre o que lê, de acordo com seus propósitos particulares. A externalização de processos requer o acesso às estratégias utilizadas pelo leitor. No caso da avaliação do produto, acessa-se, por meio de testagem, habilidades, capacidades internalizadas (ALDERSON, 2000, p. 3-4). Nos testes padronizados, assume-se que determinados produtos são aceitáveis e outros não, com base em construtos e critérios definidos. Isso é também verdade, em certa medida, para a avaliação em geral.

No contexto linguístico, está presente a visão sobre a natureza da língua/linguagem, que direciona um caráter mais cognitivo ou mais social e interacional ao teste (MCNAMARA, 2000, p. 10). Portanto, toda avaliação e todo teste carregam consigo uma limitação intrínseca a seu escopo, sendo sempre circunscrito a contextos específicos. Não há uma ideia monolítica e acabada do que seja a leitura ou a compreensão leitora. As avaliações em larga escala são, em geral, compostas por instrumentos formais, a exemplo

⁵ *Academic purposes for reading*

1. Reading to search for information (scanning and skimming)
2. Reading for quick understanding (skimming)
3. Reading to learn
4. Reading to integrate information
5. Reading to evaluate, critique, and use information
6. Reading for general comprehension (in many cases, reading for interest or reading to entertain)

dos testes de caráter objetivo e questionários que compõem SAEB e PISA. Esses testes são aplicados como medidas indiretas de habilidades e capacidades internas dos indivíduos, por isso, não se deve esquecer que interessam mais pelo que permitem construir no âmbito de inferências e generalizações sobre o que os indivíduos podem ou sabem fazer para além dos testes do que sobre o que eles sabem (ALDERSON, 2000, p. 117).

Além disso, testes envolvem questões concernentes à sua própria natureza relacionadas a validade, confiabilidade, praticidade, custo, consequências, entre outros (GRABE, 2009, p. 357). Scaramucci (1997, p. 81-82) ressalta, por exemplo, o contexto histórico em que os testes de múltipla escolha, ou de itens isolados (*discrete-point tests*), surgiram, ligados a teorias de bases estruturalista e psicométrica, dentro de uma visão atomística de linguagem, em que se privilegiava a confiabilidade, dada pelo caráter de objetividade da correção dos itens, em sobreposição à própria validade da medida. A essa orientação contrapôs-se uma visão mais funcionalista da linguagem, cuja ênfase era dada na comunicação, com foco em combinação de habilidades e em testes baseados em tarefas mais próximas de situações reais de comunicação, com destaque na validade. Douglas (2009, p. 71) refere-se a essas abordagens, respectivamente, como de itens isolados (*discrete-point*) e integrativa (*integrative*). Scaramucci (1997) e Douglas (2009) ressaltam que esses conceitos, que se apresentam como aparentemente dicotômicos, devem, no entanto, ser vistos dentro de um contínuo, como dois polos em uma gradação.

Haladyna e Rodriguez (2013) enxergam nas práticas de testagem atuais uma junção dessas duas direções. Temos, assim, variadas abordagens de testagem desses construtos e a associação de métodos de avaliar tradicionais, como os testes de itens de múltipla escolha, a práticas contemporâneas de testagem, ao mesmo tempo em que se desenvolvem métodos inovadores, como a integração de habilidades leitura e escrita, por exemplo, em tarefas que simulam situações discursivas reais.⁶

⁶ No Brasil, uma avaliação exemplar nessa direção é o exame Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras), avaliação oficial desenvolvida há duas décadas, voltada a avaliar a proficiência em português como língua estrangeira. Para maiores informações, consultar: <<http://portal.inep.gov.br/acoes-internacionais/celpe-bras>.> Acesso em 19 mar. 2020.

Alguns outros conceitos da área de avaliação são centrais para a discussão aqui empreendida, como o de *construto*, que é teoricamente definido em um teste, quer explícita ou implicitamente, e operacionalizado em suas questões e na interpretação dos resultados produzidos. Trata-se, portanto, de um conceito psicológico, que, no caso da leitura, diz respeito a que conhecimentos ou habilidades o indivíduo testado possui (MCNAMARA, 2000, p. 13), definidos a partir da concepção de leitura (e de linguagem) que baseia o teste, podendo ser mais cognitiva ou mais social, por exemplo.

Segundo Grabe e Jiang (2013, p. 4), há inúmeras maneiras de definir construto na avaliação de leitura, e, em geral, três direções norteiam essas definições: i) processos cognitivos que sustentam a compreensão; ii) objetivo de leitura; e iii) tarefas representativas. A dimensão dos objetivos de leitura tem sido considerada como a de mais fácil entendimento e mais transparente para os usuários e outros envolvidos nas avaliações, além de incorporarem tarefas e habilidades associadas. Leitura para aprender, leitura para avaliar, leitura para compreensão geral, e leitura para integrar são alguns exemplos desses objetivos. Essa organização incorporaria um rol de tarefas básicas e processos cognitivos subjacentes. Não se pode ignorar, porém, que em situação de testagem, a leitura tem como objetivo primeiro a resposta (potencialmente esperada) ao teste (ALDERSON, 2000; GRABE, 2009, entre outros).

Outro aspecto importante que envolve a avaliação de leitura se refere a propósito da avaliação. Essa definição direciona o tipo de teste a ser construído. Testes padronizados em larga escala (como Saeb e PISA) situam-se em contextos de avaliação de rendimento (*achievement*), que procura olhar para trás para medir o resultado de um processo, no caso, educacional (SCARAMUCCI, 1997, 2011). Ao mesmo tempo, ambas servem como base para diagnóstico da situação educacional de forma a, não raro, projetar usos e aprendizagens futuros. Para isso, seus resultados têm por princípio metodológico a noção de proficiência em leitura como domínio de habilidades.

3 Os construtos de leitura de Saeb e PISA em suas orientações de base

Tanto a avaliação nacional quanto a internacional têm como foco a leitura, mas cada uma define e operacionaliza esse construto por métodos (ou

formatos) variados, por isso é importante olhar para seus documentos orientadores e suas bases teóricas.

No SAEB não se encontra, em seu documento de base, uma definição de leitura, mas os pressupostos que nortearam a avaliação. Adotou-se a ideia de integração entre conteúdos (curriculares) e competências cognitivas. A noção de competência assumida vem de Perrenoud (1993), para o qual competência seria a “capacidade de agir eficazmente em um determinado tipo de situação, apoiando-se em conhecimentos, mas sem se limitar a eles” (PERRENOUD, 1993⁷, apud. INEP, 2001, p. 11). Competências cognitivas são definidas como modalidades estruturais da inteligência, e habilidades a consequência dessas competências, de caráter instrumental (INEP, 2001, p. 11)

A especificação do teste e melhor definição do construto encontram-se em uma Matriz de Referência, organizada por tópicos mais gerais e habilidades especificadas em descritores. Construído a partir de leitura de textos, advoga-se também um caráter discursivo ao teste:

Os testes de Língua Portuguesa têm como foco a leitura e seu objetivo é verificar se os alunos são capazes de apreender o texto como construção de conhecimento em diferentes níveis de compreensão, análise e interpretação. A alternativa por esse foco parte da proposição de que ser competente no uso da língua significa saber interagir, por meio de textos, nas mais diferentes situações de comunicação. É uma atividade complexa, que exige do leitor demonstrar habilidades, como reconhecer, identificar, agrupar, associar, relacionar, generalizar, abstrair, comparar, deduzir, inferir, hierarquizar. (BRASIL, 2018, p. 26).

No entanto, essas situações de comunicação referidas não são especificadas em contextos de uso. A matriz é composta por 21 descritores para o 9º ano do ensino fundamental e igualmente para a 3ª série do ensino médio, e orienta a formulação do teste, composto por itens de múltipla escolha. O caráter cognitivo é fortemente marcado nessas delimitações, ainda que intenções discursivas sejam trazidas, especialmente na definição de alguns descritores que envolvem menção a gêneros textuais ligados, por

⁷ PERRENOUD, P. Práticas pedagógicas, profissão docente e formação. Lisboa: Dom Quixote, 1993.

exemplo, ao Descritor 5 (D5), que propõe interpretação de textos com auxílio de material gráfico diverso (propagandas, quadrinhos, fotos, etc.); reconhecimento de efeitos de sentido decorrentes de pontuação ou outras notações (D17), escolha vocabular (D18), recursos ortográficos e/ou morfossintáticos (D19).

De toda forma, há a afirmação do rompimento com uma tradição conteudista do ensino, de abordagens descontextualizadas. A própria centralidade dada ao texto na avaliação parece reveladora dessa intenção e mostra uma adequação ao construto definido. Isso é afirmado explicitamente: “Na perspectiva assumida, do texto como objeto de estudo, os descritores têm como referência algumas das competências discursivas dos sujeitos tidas como essenciais nas situações de leitura.”. (BRASIL, 2001, p. 18).

Como forma de ilustrar a matriz de referência de leitura do SAEB, por limitação de espaço, apresentamos apenas um dos tópicos, que abarca cinco descritores⁸:

Tópico I – Procedimentos de leitura

D1 – Localizar informações explícitas em um texto.

D3 – Inferir o sentido de uma palavra ou expressão.

D4 – Inferir uma informação implícita em um texto.

D6 – Identificar o tema de um texto.

D14 – Distinguir um fato da opinião relativa a esse fato.

No PISA, tem-se, na área de Linguagem, igualmente uma avaliação de compreensão leitora, cuja perspectiva pauta-se assumidamente em uma dimensão social de leitura, e que, como dissemos, tem-se revisado e atualizado entre as edições da avaliação. Dessa forma, o PISA assume ser uma avaliação não apenas de leitura – mas de letramento em leitura, assim definido: “Letramento em leitura é a capacidade de compreender, utilizar, refletir e se envolver com textos escritos, com a função de alcançar uma

⁸ A Matriz completa, tal qual publicada, encontra-se em: <<http://portal.inep.gov.br/web/guest/educacao-basica/saeb/matrizes-e-escalas>>. Acesso em: 25 maio 2019.

meta, desenvolver seu conhecimento e potencial, e participar da sociedade.”. (OCDE, 2013)⁹.

O termo “compreender” liga-se à já propalada noção de compreensão leitora, e “utilizar” diz respeito a ideias de aplicação e função, do fazer algo com o que se lê. Já o termo “refletir sobre” é adicionado para agregar o caráter interativo da leitura, que implica reflexão para além do texto, os conhecimentos que o leitor traz a partir de suas experiências. Finalmente, o termo envolver-se com textos (ou engajar-se) implica a motivação para ler, tanto de ordem comportamental quanto afetiva, o senso de controle sobre o processo, levando-se em consideração sua dimensão social, em práticas frequentes e diversas com os textos escritos. Nesse sentido, o PISA ressalta a organização em torno de objetivos de leitura, conforme discutidos por Grabe (2009).

Ressalta-se a opção pelo letramento em leitura por sua natureza variada de situações e propósitos. Essa capacidade envolveria um amplo espectro de competências cognitivas. No caso do PISA, voltados tanto para usos acadêmicos presentes e futuros, para o campo do trabalho, para a participação na vida adulta, nas esferas pessoal, comunitária e econômica (OCDE, 2013, p. 62).

Para abarcar essa gama de elementos na avaliação, o PISA estabelece um rol de categorias envolvidas no teste: *situação*, que inclui os contextos ou finalidades amplas a que se aplica a leitura; *texto*, que diz respeito aos materiais lidos, que envolvem tanto formatos – a) textos verbais (chamados contínuos), b) imagéticos (chamados não contínuos), c) mistos (ou múltiplos), quanto tipos textuais –; e *aspecto*, que se refere à abordagem cognitiva que determina como os leitores se envolvem com o texto (BRASIL, 2016, p. 19). Esses elementos são sintetizados no Quadro 1.

⁹ O termo original em inglês é “*Reading literacy*”. Documento disponível no site do Inep:

<http://download.inep.gov.br/acoes_internacionais/pisa/marcos_referenciais/2013/matriz_avaliacao_leitura.pdf>. Acesso em: 25 maio. 2019.

Quadro 1. Características do domínio avaliado

Situação	Formato de texto	Tipos de texto	Aspectos
Pessoal Pública Educativa Ocupacional	Contínuo Não contínuo Múltiplo Combinados	Descrição Narração Exposição Argumentação Instrução Interação	Localizar e recuperar informação Integrar e interpretar Refletir e analisar

Fonte: Adaptado de BRASIL (2016, p. 19)

Os *aspectos* constituem os eixos do teste: localizar e recuperar informação; integrar e interpretar; refletir e analisar. Um quarto eixo, chamado de *complexo*, foi definido para agregar as atividades consideradas complexas no meio digital.

Uma aproximação entre as matrizes de referência produzida no relatório elaborado pela OCDE sobre os resultados do Brasil no PISA 2015 mostra que as habilidades do SAEB, mais detalhadas na matriz de referência, referem-se, principalmente, ao eixo cognitivo de *integrar e interpretar* (20 habilidades), encontrando-se apenas uma correspondência no eixo do *localizar e recuperar informação* e uma no eixo *refletir e analisar*, voltado no PISA, justamente, ao posicionamento do leitor frente ao que lê.

Na sequência, investigamos as implicações dessas definições para os métodos de avaliação e a forma como os construtos definidos se operacionalizam.

4 Os métodos de avaliar a leitura: formatos dos testes

Os métodos de avaliar operacionalizam as intenções presentes na elaboração de um teste. Esses métodos, ou formatos, são determinados pelas concepções e funções que compõem o construto, mas também por questões pertinentes ao próprio universo da avaliação em larga escala, tais como questões de validade, confiabilidade, praticidade, que orientam fortemente os processos de testagem em larga escala, conforme já apontado.

O teste do SAEB é composto por itens independentes de múltipla escolha, de quatro alternativas, sendo apenas uma correta, baseados em textos variados. Os itens são respaldados estatisticamente pela Teoria de

Resposta ao Item (TRI), que os insere em uma mesma escala, permitindo a calibração de escores de dificuldade. Para isso, o processo de validação do teste inclui a pré-testagem dos itens com público semelhante ao que será avaliado, de forma que o desenho do teste seja equilibrado em termos de distribuição. Um caderno de questões típico do SAEB 9º ano é composto por 26 questões de leitura, respondidas por um estudante. O conjunto do teste, a cada edição, é composto por 96 itens, distribuídos por um contingente de cadernos dentro da metodologia denominada “Blocos Incompletos Balanceados (BIB)” (BRASIL, 2018, p. 38-39), que cobre toda a matriz de referência. Isso significa que o conjunto do teste é respondido por uma turma de estudantes avaliados, e não individualmente por um mesmo estudante.

Um item típico de leitura do SAEB 9º ano é exemplificado a seguir no Quadro 2, relativa ao tópico de *procedimentos de leitura*, habilidade de *distinguir fato de opinião*:

Quadro 2. Item do SAEB 9º ano

No mundo dos sinais

Sob o sol de fogo, os mandacarus se erguem, cheios de espinhos.
Mulungus e aroeiras expõem seus galhos queimados e retorcidos,
sem folhas, sem flores, sem frutos.
Sinais de seca brava, terrível!
Clareia o dia. O boiadeiro toca o berrante, chamando os
companheiros e o gado.
Toque de saída. Toque de estrada.
Lá vão eles, deixando no estradão as marcas de sua passagem.
TV Cultura, *Jornal do Telecurso*.

A opinião do autor em relação ao fato comentado está em

- (A) “os mandacarus se erguem”
- (B) “aroeiras expõem seus galhos”
- (C) “**Sinais de seca brava, terrível!!**”
- (D) “Toque de saída. Toque de entrada”.

Fonte: Inep¹⁰

¹⁰ Disponível em: <<http://portal.inep.gov.br/web/guest/educacao-basica/saeb/instrumentos-de-avaliacao>>. Acesso em: 30 nov. 2018. Os itens

O item apresenta um texto narrativo curto e requer que o leitor diferencie, entre os fatos narrados na exposição de um cenário, uma opinião – expressa na adjetivação da seca narrada, resumida pelo vocábulo *terrível*. Isso pode não ser trivial para os leitores de 9º ano, uma vez que a habilidade de diferenciar fato e opinião geralmente é medida a partir de textos do gênero argumentativo.

Como exposto anteriormente, os itens do teste são determinados pela matriz de referência, que, no caso do SAEB, é atendida em sua integralidade, cada habilidade direcionando um item. Um dos princípios da TRI é a independência dos itens, ou seja, um item não pode condicionar a resposta a outro item. Conforme afirma Klein (2015, p. 189), essa teoria muda o foco para o item, não para o teste como um todo. Trata-se, portanto, de uma abordagem de item isolado (*discrete-point*), em que a intenção é medir um ponto de cada vez, em contraponto a abordagens analíticas, que procuram por um foco mais global, com um olhar para o todo não como soma das partes. Não há, no entanto, consenso sobre qual dessas abordagens seria mais apropriada para a testagem de leitura, dada a complexidade do fenômeno. (ALDERSON, 2000, p. 207).

Várias são as vantagens apresentadas para o uso de testes de múltipla escolha, sobretudo considerando o contexto de larga escala: o tempo de aplicação; a formulação de gabaritos; o uso de métodos computacionais para contabilidade dos resultados; e a aplicação de metodologias estatísticas para a interpretação dos dados, como é o caso da TRI. O controle de qualidade sobre os itens é outra vantagem, pois a subjetividade intrínseca aos elaboradores passa por um controle estatístico, o que confere um novo filtro de qualidade aos itens e à validade do teste.

Porém, algumas desvantagens e questionamentos sobre a pertinência desse tipo de questão para leitura também existem. O próprio processo de elaboração requer um longo período e a atuação de especialistas nas várias etapas envolvidas. Ademais, a própria associação entre produto e processo torna-se complexa, isto é, a inferência feita a partir de uma resposta correta em relação ao processo para se chegar até ela requer uma análise de

utilizados no SAEB, assim como no PISA, não são públicos, pois são reutilizados em edições subsequentes. Apenas um rol de itens é disponibilizado de tempos em tempos, quando não são mais passíveis de uso nos testes. Os itens públicos datam das primeiras edições das avaliações, da década de 2000.

especialistas. As escalas de proficiência dos resultados são fruto desse tipo de análise.

Outro ponto bastante relevante para a discussão aqui empreendida diz respeito ao questionamento sobre se é de fato a habilidade de compreensão leitora que está em jogo nesse tipo de item ou outro tipo de habilidade, mais ligada a raciocínio e a resolução de problemas, o que a diferencia em larga medida da compreensão leitora fora dos contextos de avaliação (RUPP; FERNE; CHOI 2006). Assim, o objetivo principal de leitura em um teste de múltipla escolha seria responder corretamente às questões, pois o leitor faria um ajuste de seu propósito devido ao contexto artificial. (NEVO, 1989, apud RUPP; FERNE; CHOI, 2006).

Não se pode negar também que a permanência de um único formato de teste, de base tradicional, na avaliação, dificulta o acompanhamento das inovações no campo da ciência, ao mesmo tempo em que pode respaldar práticas de estreitamento de currículos no ensino, o que é altamente indesejável em um contexto educacional.

No caso do PISA, o método de avaliar é assim definido: os itens do teste são uma mistura de itens de múltipla escolha e questões de resposta aberta. Os itens são organizados em grupos em torno de um ou mais textos autênticos. Esses textos e itens são elaborados em língua inglesa e posteriormente traduzidos para a língua dos países participantes. Com isso, há de se considerar a presença de viés cultural na avaliação, o que pode ser um fator de desmotivação ou mesmo interveniente para o público brasileiro.

Um exemplo de um conjunto de itens (unidade) típicos do PISA é apresentado no quadro 3. Ao texto vinculam-se quatro questões, das quais selecionamos duas para discussão¹¹.

¹¹ Disponível em:

<http://download.inep.gov.br/download/internacional/pisa/Itens_Liberados_Leitura.pdf>. Acesso em: 20 mar. 2020.

Quadro 3. Texto Chocolate

TEXTO 1
CHOCOLATE DE MAL GOSTO
<i>De Arnold Jago</i>
Você sabia que em 1996 os Australianos gastaram com chocolate quase a mesma quantia que o Governo Australiano gastou em ajuda internacional aos países pobres? Será que há algo errado com o Estabelecimento de nossas prioridades? O que você vai fazer a este respeito? Sim, você.
Arnold Jago, Mildura
Fonte: <i>The Age</i> , Terça-feira, 1º de Abril de 1997
A carta da página anterior apareceu em um jornal australiano, em 1997. Baseie-se nela para responder às perguntas abaixo.

Fonte: Inep

Quadro 4. Exemplo de Item Pisa – resposta construída

CHOCOLATE - QUESTÃO 3
Que tipo de resposta ou ação você acha que Arnold Jago gostaria que sua carta provocasse?

Fonte: Inep

O item 3, do tipo resposta construída curta, insere-se no eixo *refletir e analisar* e solicita do leitor uma interpretação sobre a intencionalidade da mensagem trazida pelo autor da carta e uma extrapolação da mensagem para uma possível reação da sociedade em que o leitor está

inserido. Esse modelo de item abre a possibilidade de diferentes respostas, porém também precisa prever as possibilidades de correção.

Nesse tipo de item, prevê-se a atuação de especialistas corretores e um processo de codificação das respostas. No caso desse item específico, a chave de correção está dada no Quadro 5.

Quadro 5. Chave de correção da questão 3

CHOCOLATE – CORREÇÃO 3

Codifique as respostas seguintes

a: O Governo / os indivíduos deveriam gastar mais em ajuda (externa).

Eles gostariam que as pessoas dessem mais dinheiro à ajuda externa.

Eles gostariam que se fizessem doação de dinheiro para caridade.

As pessoas deveriam gastar menos em chocolate e mais com os pobres.

b: O Governo / os indivíduos deveriam mudar suas prioridades ou percepções.

Mudar nossas prioridades.

Ele gostaria que as pessoas tivessem consciência de como nossos recursos são gastos.

c: Gastar menos em chocolate / ser menos ganancioso.

Não comprar mais chocolate.

Parar de comprar porcarias.

d: Concordância geral com o tom ou o propósito da carta.

Concordo com ele.

Sinto-me culpado / envergonhado.

Nota 1: a ou b.

Nota 0: Outras, inclusive c, d e respostas impróprias ou irrelevantes.

Gastar mais em caridade na Austrália.

Ele gostaria que o governo fosse demitido.

Ele gostaria que as pessoas dissessem, “Eu doarei todo meu dinheiro à caridade.”

Nada.

Eu não concordo com Arnold Jago.

Fonte: Inep

Nesse caso, as respostas consideradas corretas previram tanto que o leitor apontasse uma reação da sociedade, sem se incluir nela (exemplos de respostas a), quanto em respostas coletivas em que o leitor se incluísse no coletivo (exemplos de respostas b). Porém, a mera subjetivação do eu, ou a menção exclusiva a um dos elementos de comparação do texto (exemplo de respostas c), a mera concordância ou discordância ou respostas destoantes (exemplos de resposta d e demais) foram consideradas incorretas. Não houve, nesse caso, pontuações intermediárias, uma possibilidade que esse formato de item permite. O exemplo de resposta “Gastar mais em caridade na Austrália”, embora restrinja localmente a ação, poderia, no nosso entender, estar nessa categoria intermediária. Isso reflete o maior grau de subjetividade desse formato de item, o que implica questões para a confiabilidade da avaliação.

Quadro 6. Exemplo de Item PISA – múltipla escolha

CHOCOLATE - QUESTÃO 4

Neste tipo de texto usa-se, frequentemente, *fatos e opiniões* para apresentar seu argumento.

Qual destas frases da carta, se é que há uma, contém um fato?

A “Você sabia que em 1996 os Australianos gastaram com chocolate quase a mesma quantidade que o Governo australiano gastou em ajuda externa aos pobres?”

B “Será que há algo errado com nossas prioridades?”

C “O que você pretende fazer sobre isso?”

D “Sim, você.”

E Não há fatos na carta.

Fonte: Inep

Esse item também aborda a habilidade de distinguir fato de opinião, dentro do eixo *refletir*. Nesse caso, o gabarito diz respeito ao fato (letra A). Diferentemente do SAEB, essa questão de múltipla escolha apresenta cinco alternativas, sendo a última a possibilidade de não haver resposta possível. Esse elemento traz um grau maior de dificuldade para o item.

Observa-se que os itens do PISA são organizados nos testes de modo a contemplar vários aspectos de compreensão em torno do texto que encabeça o conjunto. Além dos dois exemplos trazidos, os outros dois itens referentes ao texto (ambos de múltipla escolha) exploravam a inferência sobre a intenção que o autor da carta tinha ao escrevê-la, o sentimento de culpa, além de relacionar as ideias trazidas em comparação (chocolate e ajuda internacional), para novamente solicitar do leitor a identificação de tema subjacente ao texto.

Embora preserve um caráter controlado da leitura, visto que as questões de resposta construída são questões de resposta curta e centradas na compreensão definida pelos elaboradores, essa variação de formatos permite maior liberdade na organização do teste e possibilita uma simulação de leitura mais aproximada com as práticas cotidianas dos estudantes, com a leitura no “mundo real”, sobretudo se comparada ao método exclusivo de múltipla escolha de itens isolados. Porém, conforme discutido, o método de respostas construídas, à semelhança dos métodos de avaliação de produção escrita guiados por tarefas, requer o controle rigoroso do processo de correção das respostas, com critérios claros e envolvimento de especialistas, além de ser um processo mais longo e dispendioso para avaliações em larga escala.

A partir da análise aqui trazida, é possível advogar por um método de avaliar? Considerando que avaliações são também práticas sociais, existem no mundo concreto e são também instrumentos políticos e, nesse caso, educacionais, é preciso potencializá-las de forma a que tragam benefícios e possam gerar impactos positivos aos processos educacionais. Nesse sentido, acreditamos que a avaliação educacional brasileira pode se beneficiar da experiência com o PISA e buscar ampliar o construto da avaliação e novos formatos de itens para seus testes de leitura, uma vez que a presença de variação nos formatos de itens em um mesmo teste seguramente favorece medir habilidades que avançam para além das possibilidades do modelo de múltipla escolha, especialmente na formação de leitores. Essa ampliação no método de mensuração de leitura permitiria,

também, fazer inferências mais acuradas a respeito da proficiência dos estudantes brasileiros.

Ainda nesse contexto, há de se considerar questões de ordem prática e econômica, visto que a avaliação brasileira tem, hoje, caráter censitário, atingindo milhões de alunos a cada edição. Porém, considerando que os resultados do SAEB não dizem respeito individualmente aos alunos, mas às escolas, que são a unidade de avaliação, o recorte amostral do público não traria prejuízos aos resultados pedagógicos da avaliação. Além disso, outro desafio não distante é o desenvolvimento de testes em formato eletrônico, que pode ampliar as possibilidades da avaliação.

O momento para tal desafio mostra-se nesse contexto de mudança que impactará a avaliação, a partir da recente aprovação da Base Nacional Comum Curricular (BNCC), que determina e orienta a revisão das matrizes de referências do SAEB.

5 Conclusão

Este artigo refletiu sobre a definição dos construtos de leitura e métodos de avaliar nos testes de duas importantes avaliações presentes no contexto educacional brasileiro, por meio da análise dos construtos definidos nos documentos de base e de exemplos de itens.

A análise mostrou que o PISA apresenta definições e formato explicitamente associadas a uma visão de leitura funcional, que permite a abordagem de habilidades que vão além da compreensão de elementos textuais presentes no SAEB, mais próximas de leituras da vida cotidiana, no caso da primeira.

Evidentemente apenas a existência de um teste bem concebido não garante sua eficácia, pois há inúmeros outros fatores envolvidos no processo de avaliação em larga escala, tais como motivação, vieses culturais e usos dos resultados dessas avaliações, fatores que aqui não abordamos, mas que são também parte dessas práticas. Por fim, sabemos quantas demandas de leitura nos são postas na atualidade e acreditamos que as avaliações, embora sempre limitadas, podem fornecer subsídios efetivos para o ensino, e por isso devem ser potencializadas.

Referências

- ALDERSON, J. C. *Assessing reading*. Cambridge: Cambridge University Press, 2000.
- BONAMINO, A.; COSCARELLI, C.; FRANCO, C. Avaliação e letramento: concepções de aluno letrado subjacentes ao SAEB e ao PISA. *Educação e Sociedade*, Campinas, v. 23, n. 81, p. 91-113, 2002.
- BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais. *Relatório Brasil no PISA 2018*. Versão Preliminar. Brasília, 2019
- BRASIL. *Relatório SAEB (ANEB e ANRESC) 2005-2015: panorama da década*. Brasília, 2018.
- BRASIL. *Matriz de avaliação de leitura PISA 2012*. Disponível em: <http://download.inep.gov.br/acoes_internacionais/pisa/marcos_referenciais/2013/matriz_avaliacao_leitura.pdf>. Acesso em: 25 maio. 2019.
- BRASIL. *SAEB 2001: novas perspectivas*. Brasília: INEP, 2001.
- DOUGLAS, D. *Understanding language testing*. London; New York, Routledge, 2010.
- GRABE, W. *Reading in a second language: moving from theory to practice*. Cambridge: Cambridge University Press, 2009.
- GRABE, W.; JIANG, X. Assessing reading. KUNNAN, A. J. (Ed.) *The companion to language assessment*. John Wiley & Sons, 2014.
- HALADYNA, THOMAS M.; RODRIGUEZ, MICHAEL. C. *Developing and validating test items*. New York; London: Routledge, 2013.
- KLEIN, R. Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB). In: BROOKE, N.; ALVES, M. T. G.; OLIVEIRA, L. K. M. (Orgs.) *A avaliação da Educação Básica: a experiência brasileira*. Belo Horizonte: Fino Traço, 2015. p. 189-192.
- MCNAMARA, T. *Language testing*. Oxford: Oxford University Press, 2000.
- OCDE. *Brasil no PISA 2015: análises e reflexões sobre o desempenho dos estudantes brasileiros*. São Paulo: Santilhana, 2016.

OCDE. *PISA 2012 Assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. OECD Publishing, 2013. Disponível em: <<http://dx.doi.org/10.1787/9789264190511-en>>. Acesso em: 14 jul. 2017.

RUPP, A. A., FERNE, T.; CHOI, H. How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. *Language Testing*, v. 23, n. 4, p. 441-74, 2006.

SCARAMUCCI, M. V. R. Validade e consequências sociais das avaliações em contextos de ensino de línguas. *Lingvarvm Arena*, v. 2, p. 103-120 2011.

SCARAMUCCI, M. V. R. Proficiência em LE: considerações terminológicas e conceituais. *Trabalhos em Linguística Aplicada*, v. 36, p. 11-22, jul./dez. 2000.

SCARAMUCCI, M. V. R. Avaliação de rendimento no ensino de português língua estrangeira. In: ALMEIDA FILHO, J. C. (Org.). *Parâmetros atuais para o ensino de português língua estrangeira*. Campinas, SP: Pontes, 1997. p.75-88.

Recebido em: 01/08/2019

Aceito em: 17/03/2020

Title: Large scale assessment in reading: considerations on construct and testing method in SAEB and PISA