
ARTIGO

Escrutinando o Instrumento: em questão, as provas de Matemática do Sistema de Avaliação da Educação Básica

Scrutinizing the Instrument: in focus the Mathematics tests of the Basic Education Assessment System

Elisabete Zardo **Búrigo***

 ORCID iD 0000-0003-1532-7586

Resumo

O artigo toma como objeto de investigação o modelo de provas de Matemática aplicadas pelo Sistema de Avaliação da Educação Básica, desde 2001, como instrumentos de aferição de aprendizagens e da qualidade do ensino na Educação Básica. Ancorado em análise de documentação produzida pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, o estudo identifica uma notável estabilidade dos traços principais desse modelo. Considerando o percurso das normatizações e em diálogo com a produção acadêmica sobre a temática, o texto discute algumas possíveis explicações para essa estabilidade. Entre os seus elementos explicativos, está o entendimento hegemônico que identifica qualidade de ensino com a aprendizagem de determinados conteúdos e que considera plausível aferir aprendizagens por provas, bem como medir e comparar proficiências.

Palavras-chave: Avaliação em Larga Escala. Qualidade do Ensino. Educação Matemática. Educação Básica.

Abstract

The article investigates the model of math tests applied by the Basic Education Assessment System since 2001 as instruments for measuring learning and the quality of teaching in Basic Education. Based on an analysis of documentation produced by the Anísio Teixeira National Institute for Educational Studies and Research, the study identifies a notable stability in the main features of this model. Considering the course of the standards and in dialogue with academic production on the subject, the text discusses some possible explanations for this stability. Among the explanatory elements for this stability is the hegemonic understanding that identifies teaching quality with learning certain content and that considers it plausible to measure learning through tests, as well as measuring and comparing proficiencies.

Keywords: Large-scale assessment. Educational standards. Mathematics Education. Basic Education.

1 Introdução

* Doutora em Educação pela Universidade de São Paulo (USP). Professora titular do Instituto de Matemática e Estatística da Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil. E-mail: elisabete.burigo@ufrgs.br.

Constituído nos anos 1990, o Sistema de Avaliação da Educação Básica (Saeb) tem se consolidado nos últimos 30 anos como subsídio e referência para as políticas que regulam o ensino fundamental e o ensino médio, no Brasil. Ainda que atravessadas por diferentes tipos de questionamentos e modulações, as avaliações em larga escala implementadas pelo Saeb preservam notável estabilidade na formatação de seu principal componente: as provas padronizadas consideradas nos cálculos do Índice de Desenvolvimento da Educação Básica (Ideb), instituído pelo Decreto n. 6.094 (Brasil, 2007). Desde 2001, testes de Língua Portuguesa e de Matemática são aplicados a estudantes do final de cada ciclo do ensino fundamental e médio, com a finalidade da aferição de aprendizagens (Inep, 2002; 2023).

Este texto interroga essa estabilidade, enfocando as provas de Matemática. Examina traços desse modelo e discute possíveis explicações para o seu acatamento, pontuado por críticas eventuais e marginais, no âmbito da literatura que trata do tema.

Inicialmente, o texto apresenta o modelo de provas adotado pelo Saeb desde 2001, centrado em provas de Língua Portuguesa e de Matemática, e a formatação das provas de Matemática, bem como o percurso de políticas e decisões institucionais pelas quais esse modelo foi reafirmado, pelo menos, até 2023. Essa apresentação está ancorada em documentação do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), produzida de 2001 a 2023.

A segunda parte do texto, ancorada em análise documental e resultados de pesquisas anteriores, trata da constituição de um entendimento hegemônico sobre as provas, incluindo um predominante acatamento do modelo na literatura, como um componente dessa estabilidade. Essa constituição contempla tendências progressivas: à identificação entre qualidade de ensino e a aprendizagem de determinados conteúdos, à aceitação de que provas de múltipla escolha podem aferir aprendizagens e comparar desempenhos, à validação dos procedimentos pelos quais as provas são elaboradas e os escores são calculados.

Considerando que o foco anunciado das provas de Matemática é a resolução de problemas por parte dos estudantes (Inep, 2002), o texto também examina deslizamentos produzidos em torno da noção, cotejando sua mobilização em documentos do Inep e nas provas com discussões transcorridas no campo acadêmico da Educação Matemática.

2 Matemática na Prova Brasil – estabilidade em meio às mudanças

A institucionalização do Sistema de Avaliação da Educação Básica (Saeb) nos anos 1990, segundo Bonamino e Sousa (2012), nucleou uma primeira geração de avaliações em larga

escala no âmbito da Educação Básica, que se caracterizou pela predominância da finalidade de diagnóstico da qualidade da educação ofertada no país.

De acordo com cronologia do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), até 2001 foram realizados seis ciclos de provas compostas de testes cognitivos aplicadas a grupos amostrais de estudantes do ensino fundamental e médio (Inep, 2023). Ao longo dessas aplicações, houve variações em torno dos objetivos anunciados, das metodologias de elaboração e aplicação das provas, dos grupos de estudantes e das disciplinas avaliadas, ao mesmo tempo em que se fortaleceu o papel do Inep na normatização, coordenação e execução de seus procedimentos (Bonamino; Franco, 1999; Castro, 2016). A partir de 2001, sob a direção do Inep, estabilizou-se um modelo de avaliação em larga escala no país: provas de Língua Portuguesa e de Matemática aplicadas a estudantes ao final de cada ciclo – nas quarta e oitava séries (posteriormente quinto e nono anos) do ensino fundamental e no terceiro ano do ensino médio –, organizadas segundo Matrizes de Referência construídas pelo próprio Inep (2002) e declaradamente alinhadas com os pressupostos da chamada Teoria de Resposta ao Item (TRI) (Inep, 2023).

As Matrizes de Referência do Saeb, em 2001, foram organizadas segundo a proposição da aferição de competências cognitivas, adotada desde 1997 (Pestana *et al.*, 1999; Inep, 2002). Com referência a Pierre Perrenoud, as competências cognitivas são descritas pelo Inep como “as diferentes modalidades estruturais da inteligência que compreendem determinadas operações que o sujeito utiliza para estabelecer relações com e entre os objetos físicos, conceitos, situações, fenômenos e pessoas” (Inep, 2002, p. 11). Os descritores que compõem cada uma das Matrizes são “concebidos e formulados como uma *associação entre conteúdos curriculares e operações mentais* desenvolvidas pelos alunos, que se traduzem em certas competências e habilidades” (Inep, 2002, p. 12, grifo nosso). Outro trecho do mesmo documento afirma que o modelo prioriza a “*avaliação de conteúdos* na perspectiva das competências e habilidades neles implícitas” (Inep, 2002, p. 12, grifo nosso).

A elaboração das Matrizes pressupôs, portanto, a seleção de conteúdos curriculares. Mas, como foram selecionados conteúdos, em um contexto de autonomia dos estabelecimentos de ensino na construção de seus projetos político-pedagógicos, reconhecida pela Lei de Diretrizes e Bases da Educação Nacional, Lei n. 9.394 (Brasil, 1996) e reiterada pelas Diretrizes Curriculares Nacionais para o Ensino Fundamental e para o Ensino Médio, expressas nas resoluções CNE/CEB n. 2 e n. 3 (Brasil, 1997, 1998)? O documento do Inep (2002) apresenta duas versões ligeiramente distintas.

Segundo uma primeira versão, a seleção de conteúdos foi realizada no âmbito do próprio

Inep, a partir das matrizes anteriores, e então revisada “por especialistas nas áreas de Currículo, Psicologia do Conhecimento, Língua Portuguesa e Matemática” e validada por “especialistas das diferentes secretarias de educação de todos os estados brasileiros e por amostra representativa de professores regentes de turmas de Língua Portuguesa e Matemática” (Inep, 2002, p. 7). A menção a educadores matemáticos reconhecidos e ativos na Sociedade Brasileira de Educação Matemática (Sbem), como Marcelo Câmara, Paulo Figueiredo e Regina Buriasco, sugere que teria havido algum tipo de chancela da comunidade às matrizes e ao próprio Saeb. Entretanto, Carvalho (2001) aponta que nem o processo de elaboração dos PCNs nem as matrizes do Saeb contaram com a participação dos professores; essa participação ficou restrita a poucos convidados, nominados como especialistas.

A segunda versão diz que equipes de ensino e professores de diferentes unidades da Federação foram solicitados a verificar “a compatibilidade entre as matrizes então vigentes e o currículo proposto pelos sistemas estaduais para as disciplinas de Língua Portuguesa e Matemática” (Inep, 2002, p. 15). Ao tratar da Matriz de Matemática, o texto afirma que os conhecimentos e competências matemáticas indicadas nos descritores “estão presentes, de forma consensual, nos currículos das unidades da Federação e nas Diretrizes Curriculares Nacionais” (Inep, 2002, p. 24). Observamos o pressuposto implícito de uma certa homogeneidade curricular no âmbito dos sistemas estaduais e entre eles, a despeito da autonomia acima referida.

Mas, afinal, a prova enfatiza a avaliação de conteúdos ou habilidades? Alguns trechos do documento enfatizam a avaliação de habilidades: “A relação entre o *nível estrutural das competências já adquiridas e traduzidas em habilidades* possibilitará ao aluno responder ao item” (Inep, 2002, p. 15, grifo nosso). Outros trechos sugerem que a organização da prova prioriza os conteúdos:

No processo de elaboração dos itens do Saeb, é levado em conta que, em última análise, podemos dizer que um aluno *aprendeu uma noção* quando ele é capaz de utilizá-la para resolver um problema diferente daquele que deu origem à construção da noção, quando questiona as resoluções efetuadas e as respostas encontradas (Inep, 2002, p. 24, grifo nosso).

A seção dedicada à Matriz de Matemática propugna um ensino de matemática norteado pela resolução de problemas e não limitado a “um processo que tenha como finalidade a simples memorização de regras e técnicas” (Inep, 2002, p. 23). Em acordo com essa orientação, a organização da prova buscaria apresentar “problemas significativos”, isto é, “situações que permitam ‘recontextualizar’ os conhecimentos que foram, por esse aluno, ‘descontextualizados’ por ocasião de seu processo de aprendizagem” (Inep, 2002, p. 24). A organização da prova

contempla, também, a avaliação do “domínio de determinadas técnicas” (Inep, 2002, p. 24). As Matrizes de Referência de Matemática para a quarta série do Ensino Fundamental, para a oitava série do Ensino Fundamental e para o terceiro ano do Ensino Médio contemplam, respectivamente, 27, 38 e 35 descritores. Esses descritores estão organizados nos temas “Espaço e forma”, “Grandezas e medidas”, “Números e operações/álgebra e funções” e “Tratamento da informação”, reproduzindo desse modo a lógica dos Parâmetros Curriculares Nacionais de Matemática para o Ensino Fundamental. A organização da prova prevê que temas e descritores prioritários sejam abordados em maior número de itens; a distribuição de prioridades varia conforme o ciclo avaliado.

O modelo de provas adotado em 2001, segundo o Inep, pretendia cobrir “um amplo espectro curricular” a cada aplicação (Inep, 2002, p. 31). Treze blocos com 13 itens cada foram organizados, e com diferentes agrupamentos de blocos foram compostos 26 cadernos diferentes; desse modo, embora cada aluno respondesse apenas 39 itens, a prova aferia o desempenho em 169 itens diferentes. Conforme o próprio Inep, o objetivo era o de produzir “inferências sobre o sistema educacional brasileiro e não sobre os conhecimentos individuais de cada aluno” (Inep, 2002, p. 31).

Seguindo a lógica da Teoria de Resposta ao Item (TRI), adotada desde 1995, a aferição de proficiências é baseada na estimação do grau de dificuldade dos itens que cada aluno é capaz de responder, considerando que os itens podem ser escalonados segundo “o seu grau empírico de dificuldade, ou seja, o grau de dificuldade observado de acordo com o comportamento do item quando apresentado a participantes de diversos níveis de proficiência” (Inep, 2018, p. 45). Os graus de proficiência são agrupados em níveis.

O modelo de 2001 foi replicado em 2003. Essa estabilização foi um dos suportes para a instituição da Prova Brasil, em 2005, por iniciativa do governo de Luiz Inácio Lula da Silva. Ao invés de abolir as avaliações em larga escala baseadas em provas de proficiência, como a campanha eleitoral do PT em 2002 havia sinalizado, naquele ano o Ministério da Educação criou, pela Portaria Ministerial n. 931, a Avaliação Nacional do Rendimento Escolar (Anresc) - conhecida como Prova Brasil -, de caráter censitário nas escolas públicas de ensino básico (Castro, 2016). A avaliação amostral foi mantida nas edições seguintes no âmbito da Avaliação Nacional da Educação Básica (Aneb).

O caráter censitário e o formato estabilizado da Anresc possibilitaram a criação do Ideb – indicador que considera os resultados das provas, corrigidos segundo as taxas de fluxo escolar – como carro-chefe de um novo modelo de negociação entre entes federados, consubstanciado no Plano de Metas Compromisso Todos pela Educação, estabelecido pelo Decreto n. 6.094

(Brasil, 2007). A instituição do Ideb foi, desse modo, um marco na transição para o que Bonamino e Sousa (2012) denominam uma segunda geração de políticas de avaliação, que agregam à finalidade diagnóstica a noção da responsabilização – de escolas, municípios, estados – pelo confronto entre metas firmadas e resultados alcançados.

Desde 2005, cada escola recebe um Boletim Escolar que traz, além de várias outras informações, a média dos resultados para os alunos de cada série (ou ano escolar) avaliada para cada uma das duas provas aplicadas, bem como o percentual de alunos classificados em cada um dos níveis de proficiência. Os dados agregados por escola são cotejados com aqueles relativos ao respectivo município, estado e país (Inep, 2018).

Desde a instituição da Prova Brasil e do Ideb várias normatizações foram produzidas, alterando o universo de escolas e turmas participantes das provas e os critérios para divulgação dos resultados das provas e sua utilização no cálculo do Ideb. O Saeb foi gradativamente complexificado, com a criação da Avaliação Nacional da Alfabetização (ANA) em 2013, e a aplicação experimental de provas de Ciências da Natureza e Ciências Humanas, conforme matrizes propostas pelo Inep (2018). Não houve, contudo, alterações no formato das provas consideradas para o cálculo do Ideb, exceto pelos números de blocos de questões e de número de itens a serem respondidos (Inep, 2018).

Em 2012, o próprio Inep, a partir de estudos e de críticas ao modelo das competências e habilidades, deu início à elaboração de novas matrizes, orientadas para a “construção de itens que possam medir as apropriações pelo estudante do conhecimento socioculturalmente produzido, a partir das experiências de aprendizagem escolares” (Inep, 2014, p. 82). Entretanto, as matrizes aplicadas na elaboração da Prova Brasil não foram modificadas nas edições seguintes.

Por meio de sua Resolução CNE/CP n. 2 (Brasil, 2017) o Conselho Nacional de Educação estabeleceu prazo de um ano, a partir da publicação da Base Nacional Comum Curricular (BNCC), para que as Matrizes de Referência das avaliações em larga escala fossem alinhadas à Base. Em 2018, o Inep divulgou uma versão inicial de novas matrizes alinhadas com a BNCC, abrangendo as áreas de Ciências Humanas e Ciências da Natureza, e desse modo contemplando um arco mais amplo de componentes curriculares abrangidos pela BNCC (Inep, 2019).

Para as avaliações de Língua Portuguesa e de Matemática também foram elaboradas novas matrizes, com o propósito anunciado de alinhamento com as competências específicas indicadas na BNCC para cada componente. No caso da Matemática, o documento aponta que o constructo a ser avaliado pelas provas seria “o Letramento Matemático, conceituado como a

compreensão e aplicação de conceitos e procedimentos matemáticos na resolução de problemas nos campos de Números, Álgebra, Geometria, Grandezas e Medidas e Probabilidade e Estatística, bem como na argumentação acerca da resolução de problemas” (Inep, 2019, p. 84). Para uma adequada avaliação desse constructo, seria necessário incluir, além dos usuais itens de resposta objetiva, “itens de resposta construída” (Inep, 2019, p. 57), com alterações importantes no formato da prova, no processo de correção e no cálculo dos resultados.

Entretanto o Inep advertia que, para preservar a comparabilidade das versões do Saeb, e considerando as metas do Plano Nacional de Educação – Lei n. 13.005 (Brasil, 2014), as provas consideradas para o cálculo do Ideb em 2021 manteriam o mesmo formato e, portanto, as Matrizes de Referência de Língua Portuguesa e de Matemática adotadas desde 2001 (Inep, 2019). O mesmo formato e as mesmas matrizes foram mantidos em 2023 (Inep, 2023).

Portanto, ao longo de vinte e três anos – de 2001 até 2023, inclusive - as provas de Matemática e de Língua Portuguesa aplicadas ao final de cada ciclo (e consideradas para o cálculo do Ideb desde 2007) mantiveram notável estabilidade. Nesse mesmo período, houve alternância de governos com diferentes orientações para as políticas educacionais, foram produzidas novas Diretrizes Curriculares, foi debatido e aprovado um novo Plano Nacional de Educação e, a partir dele, foi estabelecida nova Base Nacional Comum Curricular, mandatória para as escolas e redes de ensino desde o início do ano letivo de 2020.

3 Traços de um entendimento hegemônico

Estudos sobre as avaliações em larga escala, no país e no mundo, explicam os processos de legitimação dessas avaliações pela combinação de tendências internacionais com dinâmicas locais (Addey *et al.*, 2017).

No caso brasileiro, Lima e Gandin (2019) apontam os anos 2000 como tempos de consolidação de um entendimento hegemônico sobre a necessidade das avaliações, em que visões alternativas são relegadas a “vozes periféricas, com pouca visibilidade” e em que se produz a “sensação de não haver alternativas viáveis” (Lima; Gandin, 2019, p. 24). A criação da Anresc, em 2005, e do Ideb, em 2007, são, para os autores, evidências de que a influência dos organismos internacionais sobre a necessidade das avaliações em larga escala já fora “apropriada como um discurso do nosso próprio país e, portanto, não há mais a necessidade de um convencimento externo constante” (Lima; Gandin, 2019, p. 18). Um componente dessa apropriação teria sido a valorização da transparência, segundo o pressuposto de que a avaliação permitiria à sociedade conhecer a qualidade do ensino em cada escola do país.

Traços desse entendimento hegemônico podem ser identificados no Documento Final da Conferência Nacional de Educação (Conae) (Brasil, 2010a), precedendo a tramitação do novo Plano Nacional de Educação no Congresso Nacional. Repercutindo debates realizados nas conferências municipais e estaduais, o texto da Conae expressa uma visão crítica às políticas de avaliação centradas em provas, reconhecendo a qualidade da educação como “fenômeno também complexo e abrangente, de múltiplas dimensões” (Brasil, 2010a, p. 47), e propugnando novos marcos para os processos avaliativos, segundo uma “visão ampla de avaliação, que se contrapõe à centralidade a ela conferida e que resulta em controle e competição institucional” (Brasil, 2010, p. 53). Entretanto, as avaliações do chamado “rendimento escolar” são aceitas no documento como componentes desses processos mais complexos; como afirmam Lima e Gandin (2019), o “uso de avaliações em larga escala é tomado como um dado, não sendo cogitado pensar que elas possam não existir” (Lima; Gandin, 2019, p. 3).

Durante a tramitação do Projeto de Lei n. 8.035/2010 (Brasil, 2010b), que propunha o novo Plano Nacional de Educação (PNE), também foram registradas críticas às avaliações em larga escala e proposições de exclusão do Ideb como referência para a aferição da qualidade do ensino. Deputados do Partido dos Trabalhadores (PT), do Partido Democrático Trabalhista (PDT) e do Partido Socialismo e Liberdade (PSOL) propuseram suprimir referências ao Ideb na Meta 7 do PNE (Emendas n. 123 e 982/2011). Entretanto, todos os questionamentos foram refutados pelo Deputado Ângelo Vanhoni, relator do Projeto na Câmara dos Deputados, com a justificativa de que o Ideb “é hoje um índice relevante” (Brasil, 2011, p. 108). Com a aprovação do substitutivo do relator, o Ideb, instituído por decreto presidencial em 2007, foi cancelado pelo Congresso Nacional. Já no texto da Lei n. 13.005 (Brasil, 2014) foi incluída pelo Congresso a previsão da constituição de um Sistema Nacional de Avaliação da Educação Básica (Sinaeb), mais amplo e complexo que o Saeb. As menções ao Ideb e ao *rendimento escolar* foram mantidas, embora deslocadas do caput para os parágrafos do Artigo 11 da Lei. A construção do Sinaeb foi objeto de estudos desenvolvidos sob a coordenação da Diretoria de Avaliação da Educação Básica (Daeb) do Inep (Santos; Horta Neto; Junqueira, 2017). Entretanto a Portaria MEC n. 369/2016, que instituía o Sinaeb, foi revogada pelo Ministro Mendonça Filho, durante a presidência interina de Michel Temer (Almeida; Xabregas, 2017).

Numa direção diversa do entendimento hegemônico, encontramos interrogações sobre a qualidade e acuidade das provas do Saeb.

Schwartzman (2013) aponta que, apesar da profusão de dados divulgados pelo Inep,

Há pouca ou nenhuma informação sobre as formas como as escalas foram construídas, a confiabilidade dos testes, a qualidade dos itens e dos bancos de itens e outros aspectos técnicos de que os psicometristas e estatísticos externos precisariam para

avaliar a qualidade e a relevância dos resultados finais (Schwartzman, 2013, p. 283, tradução nossa).

Os questionamentos sobre a qualidade técnica das provas, inicialmente concebidas como amostrais e diagnósticas, são adensados mediante considerações sobre o uso dos escores para o estabelecimento de metas e aferição de resultados. Schwartzman (2013, p. 281) menciona a “lei de Goodhart”: quando um indicador social ou econômico se torna uma referência para condução de políticas sociais ou econômicas – como ocorre com o uso do Ideb –, ele perde o conteúdo informativo que o qualificaria para desempenhar esse papel. O autor aponta, ainda, a inconsistência no uso de testes que são de baixo impacto na vida escolar dos estudantes para definir políticas de alto impacto – como a comparação de resultados ou a alocação de recursos – sobre as escolas e redes. Todavia, não descarta a contribuição das provas para as avaliações diagnósticas sobre os sistemas de ensino, endossando parcialmente, portanto, sua validação.

Ponderações sobre a qualidade técnica do Saeb, ainda que vagas, também estão presentes em documento do próprio Inep:

Desde o final de 2008, servidores das quatro áreas de conhecimento [...] registram suas considerações acerca dos problemas identificados nas fundamentações e nos instrumentos dos referidos exames e avaliações, ora apresentando algumas possibilidades de ajustes em curto prazo, ora enfatizando a urgência de se promover discussão ampla, tanto interna quanto externa ao Inep, a fim de rever exames e avaliações da educação básica, para garantir coerência e atualização teórico-metodológica de seus instrumentos (Inep, 2014, p. 7).

O mesmo documento constata desarticulação entre currículo e avaliação, bem como discrepâncias entre descritores componentes das matrizes e itens das provas. Apresenta ponderações sobre os tamanhos das provas, tipos de questão, sua adequação à faixa etária e aos respondentes: “as matrizes de referência do Saeb/Prova Brasil não são interdisciplinares, e os itens, em geral, não são contextualizados” (Inep, 2014, p. 52).

A literatura que trata das avaliações em larga escala no Brasil registra um amplo espectro de posicionamentos em relação ao tema. Entretanto, as sistematizações do debate indicam que questionamentos às provas enquanto instrumentos capazes de medir aprendizagens, como os acima referidos, são rarefeitos.

Bauer, Alabarse e Oliveira (2015, p. 1369) apresentam uma discussão das “principais críticas às avaliações em larga escala presentes na literatura nacional e internacional”. Associando as avaliações à emergência das políticas de controle dos processos educativos por resultados verificáveis, os autores explicam a validação das provas como “chave para se verificar a qualidade do ensino, entendida apenas como elevação dos níveis de aprendizagem dos alunos” (Bauer; Alabarse; Oliveira, 2015, p. 1370). Ao tratarem das críticas às avaliações, os autores mencionam, sobretudo, as discussões sobre seus propósitos e efeitos. Um breve

registro trata de questionamentos que “colocam em dúvida o que são capazes de medir e da própria medição, em termos de aprendizagem dos alunos, o cerne daquilo para o que foram projetadas e com a precisão anunciada de seus resultados” (Bauer; Alabarse; Oliveira, 2015, p. 1373). Também é breve o registro sobre ponderações relativas à “seleção dos conteúdos e habilidades que embasam essas avaliações, bem como a definição de níveis ou padrões de rendimento” (Bauer; Alabarse; Oliveira, 2015, p. 1374).

Ribeiro e Sousa (2023, p. 3) apresentam um “continuum argumentativo com diferentes gradações de apoio e de recusa às avaliações em larga escala”. O espectro identificado no texto varia desde o apoio incondicional até a recusa categórica das avaliações. Entre os defensores do apoio incondicional são mencionados atores participantes da constituição e do fortalecimento do Saeb, como Maria Helena Guimarães Castro, presidente do Inep de 1995 a 2002. Na categoria nomeada como “recusa propositiva”, Ribeiro e Sousa (2023) situam autores que questionam as limitações e os usos do Ideb, apontando o estreitamento curricular e outros efeitos deletérios das políticas de responsabilização. Já a “recusa categórica” é identificada com a denúncia dos interesses e intenções que orientam as políticas de avaliação em larga escala, e com a proposição da autoavaliação da escola como alternativa. Nesta sistematização apresentada pelos autores, não são mencionados questionamentos à possibilidade de se aferir aprendizagens pelas provas padronizadas; apenas a reivindicação de “mais discussão pública acerca da validade política, teórica e técnica das avaliações” (Ribeiro; Sousa, 2023, p. 7).

Miguel (2016) registra uma tendência predominante à validação das provas, também no âmbito da comunidade de educadores matemáticos:

Todos se alarmam e se chocam com os resultados desastrosos de desempenho em matemática de nossos estudantes de todos os níveis da escola básica. Quase ninguém se interroga, porém, acerca dos jogos normativamente regrados de linguagem que orientam não só a produção das matrizes de avaliação do desempenho, com base nas quais se elaboram as questões das provas, mas que orientam também, os próprios pressupostos relativos ao conhecimento, à matemática e à aprendizagem, com base nos quais essas próprias matrizes são construídas (Miguel, 2016, p. 350).

4 Sentidos de qualidade e centralidade curricular

O afunilamento curricular – ou a seleção de determinados componentes do currículo como indispensáveis, ou mais relevantes que outros – é uma das condições para a aceitação de que a qualidade do ensino seja aferida por testes de Língua Portuguesa e Matemática, em um estreitamento das Diretrizes Curriculares vigentes e dos projetos político-pedagógicos adotados nas escolas.

Matheus e Lopes (2014) argumentam que as disputas e deslocamentos sobre os

significados atribuídos à qualidade do ensino, com seu decorrente esvaziamento de significados, compõem um dos mecanismos de constituição da hegemonia da política centralizada de currículo. É importante lembrar que a expressão *padrão de qualidade* já estava presente no Artigo 206 da Constituição Federal de 1988 (Brasil, 1988). A Lei de Diretrizes e Bases da Educação Nacional, Lei n. 9.394 especificou em seu Artigo 4º, inciso IX, que se tratava de estabelecer “a variedade e quantidade mínimas, por aluno, de insumos indispensáveis ao desenvolvimento do processo de ensino-aprendizagem” (Brasil, 1996, s.p.). Portanto, o sentido atribuído à qualidade do ensino, nos termos da Lei n. 9.394, está relacionado à garantia de condições de oferta consideradas adequadas. Esse sentido foi sendo esvaziado pela ausência de regulamentação e consequente ineficácia do inciso: a primeira tentativa de sua regulamentação pelo CNE, consubstanciada no Parecer n. 8/2010 da sua Câmara de Educação Básica (CEB), foi engavetada pelo Ministério da Educação (MEC) e retomada apenas em 2019.

Nos anos 1990, o discurso da qualidade total, reiterado por organismos internacionais que avalizavam aportes de recursos, como o Banco Mundial, fez circular o sentido da qualidade de ensino como mensurável, aferível pelos testes padronizados (Matheus; Lopes, 2014).

A produção dos Parâmetros Curriculares Nacionais para o Ensino Fundamental (PCNs), iniciada em 1994, seguiu a lógica da padronização curricular assumida, a partir das pressões internacionais, no Plano Decenal Educação para Todos (Moreira, 1996). Em acordo com os preceitos da Lei de Diretrizes e Bases da Educação Nacional (LDBEN), o CNE, em seu Parecer CNE/CEB n. 03/97, caracterizou os Parâmetros Curriculares Nacionais (PCNs) como “proposição pedagógica, sem caráter obrigatório” (Brasil, 1997, p. 6). Entretanto o detalhamento das orientações didáticas ali contidas, o diálogo entre essas orientações e discussões correntes entre os professores, a ampla circulação do documento nas escolas e sua adoção por parte do Programa Nacional do Livro Didático favoreceram que fossem, desde sua publicação em 1997 e 1998, tratados e percebidos como currículo nacional de fato, mesmo entre as educadoras e os educadores matemáticos (Búrigo, 2021). Os PCNs também fizeram circular a noção de *competências*, mobilizada em reformas curriculares de outros países e prestigiada pelos organismos internacionais. A aceitação das Matrizes de Referência do Saeb foi, desse modo, preparada pela disseminação e progressiva validação dos PCNs.

A circunscrição das provas do Saeb às áreas de Língua Portuguesa e Matemática, adotada desde 2001, não foi justificada nos documentos divulgados pelo Inep (2002, 2006). Um exame do *Relatório Nacional Saeb 2003* permite, entretanto, identificar que os resultados das provas foram tomados como sintomas de um quadro considerado grave: “As análises das médias de proficiência, nos diversos estratos amostrais do Saeb, permitiram constatar a falta de

eficiência nos sistemas de ensino no Brasil” (Inep, 2006, p. 137). Nas avaliações de Matemática para a oitava série do Ensino Fundamental, apenas 3,3% dos alunos que compunham a amostra em 2003 foram considerados como em estágio adequado de construção de competências. O desastre revelado – sob o pressuposto de que os instrumentos de medição eram adequados e precisos – foi tomado então como indicador da urgência da melhoria da qualidade de ensino nas áreas medidas pelos testes, portanto na Língua Portuguesa e na Matemática. Percebemos que as avaliações em larga escala engendraram sua própria validação, suscitando novos esforços e novas comparações nas áreas do conhecimento já avaliadas.

O discurso da qualidade social, como apontam Matheus e Lopes (2014), foi formulado como resistência e em antagonismo ao da qualidade total, defendendo, no sentido contrário à mercantilização e aos interesses do empresariado, o direito à educação como conquista das camadas populares. A diversidade curricular – reiterada na Lei n. 9.394/1996 e em diversos documentos curriculares – seria uma decorrência do protagonismo desses setores. Entretanto, ao longo dos anos 2000, os discursos sobre a qualidade social foram incorporando elementos do discurso da qualidade total – em particular, a identificação entre o exercício do direito à educação e a aprendizagem de conteúdos curriculares: “Estar na escola e alcançar níveis instrucionais comuns a todos os alunos, evidenciados por exames nacionais e internacionais, são naturalizados como expressão da qualidade do currículo e, portanto, da educação” (Matheus; Lopes, 2014, p. 341).

A centralidade curricular foi reforçada pela adesão do Brasil ao *Programme for International Student Assessment* (Pisa). Mesmo sem ser membro da Organização para Cooperação e Desenvolvimento Econômico (OCDE), o país participa do exame desde sua primeira aplicação, no ano 2000.

De acordo com o Inep, a meta estabelecida em 2007 de alcançar Ideb 6,0 em 2021 foi estipulada com a intenção de se alcançar o desempenho médio dos países membros da OCDE no Pisa (Fernandes, 2007). O *Relatório Nacional Saeb 2003* já enfatizara e detalhara resultados da aplicação do Pisa no Brasil, entre 2000 e 2003 – por exemplo, o avanço de 300 para 350 pontos no conteúdo “Espaço e forma” da prova de Matemática (Inep, 2006). Essa referência do Ideb ao Pisa foi formalizada na estratégia 7.11 do já mencionado Plano Nacional de Educação (PNE) 2014-2024. As correlações pretendidas pelo Inep entre aferições e resultados do Saeb e do Pisa são questionáveis sob vários aspectos (Villani; Oliveira, 2018). Entretanto, a legitimidade atribuída ao Pisa, em escala internacional, fortalece o entendimento de que haveria um conjunto de habilidades consideradas necessárias para que jovens, ao final da escolarização obrigatória, participem de modo produtivo das economias de seus países, e de que essas

habilidades podem ser aferidas por provas (Meyer, 2014). O foco das provas do Saeb na Língua Portuguesa e na Matemática é compatível com a centralidade que o Pisa atribui às provas de letramento e de letramento matemático, ainda que o exame comporte outras áreas e temas.

O engajamento do Brasil no exame foi fortalecido ao longo de sucessivos governos, reforçando a dimensão da centralidade curricular nele implícita. A aceitação do modelo do Saeb entre educadores matemáticos, no Brasil, de certo modo espelha também o endosso às provas de letramento matemático do Pisa. Esse endosso é favorecido pelas referências, nos documentos da OCDE, às ideias de modelagem matemática, de resolução de problemas autênticos da vida real e de *empowerment* e pelo engajamento, na concepção do exame, de educadores matemáticos eminentes como Mogens Niss, secretário-geral da *International Commission on Mathematical Instruction* (ICMI), de 1991 a 1998 (Stacey; Turner, 2015).

5 A validação das provas pela produção de efeitos esperados

A medição de aprendizagens por provas foi foco de controvérsias nos debates educacionais dos anos 1970 e 1980, e apontada como um dos motivos para elevados índices de repetência e evasão nas etapas iniciais de escolaridade. Não é por acaso que a Lei de Diretrizes e Bases da Educação Nacional estabelece, como critério para verificação do rendimento escolar, a “avaliação contínua e cumulativa do desempenho do aluno, com prevalência dos aspectos qualitativos sobre os quantitativos e dos resultados ao longo do período sobre os de eventuais provas finais” (Brasil, 1996, s.p.).

A avaliação por questões de múltipla escolha foi ainda mais contestada nos debates pedagógicos do final do século XX, pois sequer oportuniza que o aluno explicita sua interpretação e modo de pensar, e contabiliza do mesmo modo erros decorrentes de incompreensão ou equívoco e erros acidentais ou sistemáticos na aplicação de um procedimento (Buriasco, 2002). O próprio Inep, em discussão sobre instrumentos de avaliação, reproduzindo trecho dos Parâmetros Curriculares Nacionais, adverte os professores de que “devem também contemplar as explicações, justificativas e argumentações orais, uma vez que estas revelam aspectos do raciocínio que muitas vezes não ficam explícitos nas avaliações escritas” (Inep, 2002, p. 23). Em estudo sobre novas matrizes para o Saeb, especialistas do Inep (2014) indicaram a necessidade do uso de calculadora durante a prova, reconhecendo o peso desmesurado dos cálculos nos itens da prova de Matemática. Itens de “resposta construída” foram propostos pelo Inep (2019) para avaliar habilidades que não podem ser aferidas nos chamados itens de “resposta objetiva”.

Como, nesse contexto em que as provas tiveram seu peso refutado pela literatura e reduzido no âmbito do cotidiano escolar, elas puderam obter reconhecimento como instrumentos legítimos de verificação de aprendizagens?

Pettersson, Popkewitz e Lindblad (2016) situam a emergência e a validação das avaliações em larga escala em uma corrente mais antiga e persistente de uso de estatísticas e de confiança nos números como capazes de traduzir fenômenos para um amplo público com confiabilidade. O uso de números para a descrição e regulação de fenômenos políticos e sociais esteve associado, desde o século XIX, a uma pretensão de objetividade, associada a discursos e aspirações democráticas. Um longo processo de reificação permite que os números sejam vistos como expressão direta da realidade, como se não fossem dependentes dos processos pelos quais foram construídos ou dos instrumentos pelos quais foram aferidos.

Uma dimensão crítica desse processo de representação da realidade por números é a crença e a valorização das comparações internacionais. Embora essas comparações fossem praticadas desde o século XIX, pelo menos, é no pós-guerra que se desenvolvem as *International Large-Scale Assessments* (Ilsa), com a pretensão de comparar aprendizagens entre estudantes de diferentes países.

Ainda que reconhecendo um percurso longo de validação de números e comparações, Meyer (2014) registra uma virada mais recente em favor das Ilsa no plano internacional. Até o início dos anos 1990, prevalecia no âmbito da OCDE o ceticismo em torno das avaliações: “As comparações internacionais de desempenho são populares, embora frequentemente de validade duvidosa” (Skilbeck, 1990, como citado em Meyer, 2014, p. 6). O autor argumenta que a mudança em favor das Ilsa integra um amplo redirecionamento da OCDE na direção das políticas neoliberais, em favor de uma educação escolar orientada para o mercado.

A adoção da Teoria de Resposta ao Item (TRI) é um outro elemento importante para a ampla aceitação das avaliações em larga escala desde os anos 1990, no âmbito internacional e com repercussões no Brasil. Tavares (2013) explica que os estudos sobre a TRI tiveram início nos anos 1950, como parte de um esforço de buscar medições de inteligência que não fossem dependentes da acuidade do instrumento utilizado; entretanto, foi só nos anos 1980 que a aplicação da TRI se tornou possível, graças ao desenvolvimento de softwares suficientemente sofisticados. A TRI aporta um refinamento à capacidade dos números hierarquizarem desempenhos, associando níveis de proficiência ao grau de dificuldade estimado das questões: indivíduos, grupos e questões são situados em uma mesma escala. A promessa da comparação temporal dos resultados reforça a potência atribuída às medições e, ao mesmo tempo, fundamenta o sigilo em torno dos itens. A sofisticação dos métodos estatísticos opera para o

consentimento às provas, segundo pelo menos duas tendências. De um lado, a TRI é apresentada como mais capaz de representar níveis de proficiência do que a Teoria Clássica dos Testes; como a apuração do escore associa erros e acertos a um cálculo de probabilidade de expressão de uma determinada habilidade, corrigido por outros acertos e erros, as limitações das chamadas provas objetivas tendem a ser minimizadas, até mesmo porque soterradas sob uma parafernália de cálculos. De outro lado, a complexidade dos métodos, agravada pelo sigilo em torno dos itens e respectivos parâmetros, reduz a auditabilidade do exame.

Desse modo, a sofisticação da TRI e o seu uso disseminado em avaliações em larga escala têm obscurecido inconsistências entre o modelo e sua aplicação. Uma inconsistência flagrante está relacionada ao pressuposto da unidimensionalidade, isto é, de que todos os itens de uma mesma prova mediriam o mesmo traço latente ou a mesma habilidade; esse pressuposto permite situar todos os itens em uma mesma escala, segundo graus variados de dificuldade (Tavares, 2013). As próprias Matrizes de Referência adotadas pelo Saeb já contradizem o pressuposto, apresentando dezenas de habilidades a serem avaliadas em itens diversos. Se reconhecemos que a prova de Matemática não é unidimensional – pois apresenta, por exemplo, itens que avaliam operações com formas e outros que exigem apenas processamento de números –, o ordenamento dos itens segundo graus de dificuldade não será o mesmo para todos os indivíduos avaliados, comprometendo a estimação de proficiência do respondente, que corresponde à proficiência necessária para a resolução do item estimado como mais difícil. Uma outra provável inconsistência está relacionada à qualidade dos itens, que deveriam ser submetidos, como aponta Tavares (2013), a pré-testagem segundo amostras representativas. Como aponta Schwartzman (2013), a opacidade do processo de elaboração das provas não permite avaliar a qualidade e confiabilidade dos itens.

Um efeito também pouco discutido da adoção da TRI é a exigência, decorrente do próprio modelo, da comparação de proficiências, o que pressupõe itens capazes de discriminar desempenhos e, por conseguinte, que predominantemente não serão de fácil resolução para a maioria dos respondentes. A própria dinâmica de elaboração da prova impõe, desse modo, um grau previsível de insucesso para uma grande parcela da população avaliada.

A aceitação das provas do Saeb como capazes de aferir aprendizagens, portanto, não está ancorada na correspondência entre currículo e itens avaliados nem na qualidade técnica do processo pelo qual itens e provas são construídos.

Entretanto, como apontam Pettersson, Popkewitz e Lindblad (2016), a legitimidade das avaliações em larga escala encontra respaldo em uma longa tradição de comparação de desempenhos. No Saeb amostral, que vigorou até 2003, a comparação era apenas temporal, pois

não havia dados sobre populações. O exame censitário inaugurou o inevitável ranqueamento resultante da medição e divulgação de proficiências de escolas, municípios e estados e, com ele, como aponta Ball (2001, p. 110), “um fluxo de performatividades contínuas e importantes, isto é, um espetáculo”. Se, de um lado, o escalonamento provoca uma competição incessante, e, portanto, uma importante instabilidade, com a tendência à continuada elevação das metas, de outro lado ele tem como efeito a confirmação de desigualdades esperadas: entre regiões do país, entre redes com mais ou menos recursos, entre escolas que atendem camadas médias e camadas populares. A confirmação dessas desigualdades esperadas de desempenho – porque assentadas em desigualdades nas condições de oferta ou no acesso ao conhecimento por parte das populações envolvidas – empresta verossimilhança aos escores de proficiência do Saeb, a despeito de toda opacidade no processo pelo qual são produzidos.

Mas, as comparações também carecem de sustentação técnica. Travitzki (2020) observa que a divulgação do Ideb por parte do Inep não tem sido acompanhada da divulgação do respectivo intervalo de confiança. Em estudo que utilizou os microdados da Prova Brasil e do Censo Escolar aplicados às séries finais do Ensino Fundamental em 2015, incluindo os resultados de cada uma das provas para cada respondente, o autor conclui que o Ideb não é adequado para comparar resultados de escolas, dada a elevada margem de erro na estimativa do escore para essa escala.

6 Resolução de problemas: evocações e deslizamentos

A Matriz de Referência de Matemática para o Saeb, aplicada desde 2001, tem sido apresentada pelo Inep como norteadora pela resolução de problemas. No campo da Educação Matemática, no Brasil, a resolução de problemas tem se constituído como tendência desde os anos 1980, articulando a investigação com a reflexão em torno da prática docente (Onuchic; Allevato, 2011). Os documentos do Inep que tratam da Matriz (Inep, 2002, 2006, 2018; Brasil, 2008) não apresentam uma conceituação do que se entende por *resolução de problemas*, mas considerações que evocam e de algum modo buscam se aproximar dos debates no campo da Educação Matemática.

Uma consideração reiterada nos documentos do Inep é a de que a resolução de problemas demanda o desenvolvimento de estratégias de resolução por parte dos alunos: “o conhecimento matemático ganha significado quando os alunos têm situações desafiadoras para resolver e trabalham para desenvolver estratégias de resolução” (Inep, 2002, p. 23; Brasil, 2008, p. 106; Inep, 2018, p. 29). Desse modo, os documentos sugerem que as provas de Matemática

do Saeb poderiam avaliar o desenvolvimento de estratégias de resolução de problemas pelos alunos.

Como já apontado na seção anterior, essa pretensão de avaliar estratégias de resolução de problemas é contraditória com o formato da prova, em que o aluno apenas marca uma alternativa de resposta e não tem oportunidade de registrar os caminhos pelos quais construiu ou mesmo escolheu uma resposta. Há, além dessa, uma contradição mais flagrante com a resolução de problemas como é entendida por autores da Educação Matemática, pois o tempo limitado e a diversidade de itens são incompatíveis com a realização de uma “tarefa ou atividade para a qual não se tem métodos ou regras prescritas ou memorizadas, nem a percepção de que haja um método específico para chegar à solução correta” (Onuchic; Allevato, 2011, p. 81).

O documento *PDE: Plano de Desenvolvimento da Educação: Prova Brasil: ensino fundamental: matrizes de referência, tópicos e descritores*, dirigido aos professores, procura contornar essa contradição, apontando, de um lado, que os resultados da prova não devem ser usados para comparar escolas e, de outro, sugerindo que a intencionalidade da prova seria, sobretudo, induzir estratégias de ensino mais eficazes nas escolas:

Esses mesmos itens mostram que a Prova Brasil concentra-se em medir competências básicas e essenciais e que, portanto, qualquer estratégia que dote os alunos de capacidade de responder corretamente aos itens da Prova Brasil estará lhes permitindo consolidar competências fundamentais para o exercício de sua cidadania (Brasil, 2008, p. 16)

Uma perspectiva diversa foi apontada no documento *Fundamentação das matrizes de avaliação da educação básica: estudos e propostas* (Inep, 2014), com o reconhecimento de que havia várias inconsistências nas matrizes adotadas desde 2001. Esse reconhecimento da inadequação da prova como instrumento para avaliar resolução de problemas foi adensado em 2019: “Nas avaliações de larga escala não se tem acesso à estratégia utilizada e, portanto, não se tem informação se foi um exercício ou um problema” (Inep, 2019, p. 86). Entretanto, como apontado anteriormente, a Matriz e o formato das provas aplicadas aos quinto e nono anos do ensino fundamental e ao terceiro ano do ensino médio não foram modificados, até o início de 2024.

7 Considerações finais

Ao final da vigência do Plano Nacional de Educação 2014-2024, o peso do Saeb na configuração das políticas educacionais é ainda maior do que ao tempo da instituição da Prova Brasil, em 2005, e do Ideb, em 2007. As pressões para a elevação do Ideb e, portanto, dos

desempenhos medidos no Saeb, só crescem com a aprovação da Emenda Constitucional n. 108 e da Lei n. 14.113 (Brasil, 2020a, 2020b), que regulamenta o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb), estabelecendo que parcela da complementação da União ao Fundeb referente ao chamado Valor Anual por Aluno (VAAR) considerará os resultados médios dos estudantes de cada rede pública estadual e municipal nos exames nacionais do sistema nacional de avaliação da educação básica.

A constituição de um Sistema Nacional de Avaliação da Educação Básica (Sinaeb), mais amplo e complexo que o Saeb, previsto pelo PNE 2014-2024, ainda está por ser empreendida. O Documento Final da Conae 2024, realizada em janeiro de 2024, aponta a superação de “uma avaliação pautada na métrica, na quantificação, no ranqueamento e na classificação de instituições educativas” (Brasil, 2024, p. 148). Entretanto, prevê a manutenção das avaliações externas baseadas em testes aplicados aos estudantes.

Em tempos de formulação de um novo Plano Nacional de Educação, é crucial empreender uma avaliação cuidadosa dos instrumentos de avaliação adotados nestes últimos trinta anos e, em particular, das provas de Língua Portuguesa e de Matemática consideradas no cálculo do Ideb. As educadoras matemáticas e os educadores matemáticos temos uma responsabilidade importante para com esse processo.

Neste texto, trazemos algumas das vozes que contestam a capacidade do Saeb medir aprendizagens e, em particular, a capacidade da prova de Matemática medir a competência dos estudantes na resolução de problemas. São vozes rarefeitas, mas não tão periféricas: pois incluem, ainda que sem muito alarde, manifestações dos especialistas do próprio Inep, responsável desde os anos 1990 pela elaboração das matrizes e das provas.

Trazemos, também, um esforço de descrição dos caminhos pelos quais o modelo de provas do Saeb tem sido aceito e naturalizado. Com essa reconstituição, pretendemos contribuir para a desnaturalização dessas provas e, mais amplamente, para a discussão da capacidade das avaliações em larga escala retratarem aprendizagens e, sobretudo, a qualidade da educação.

Contribuições de autoria

Todos os autores contribuíram, substancialmente, na concepção e no planejamento do estudo; na obtenção, análise e interpretação dos dados; na redação e revisão crítica; e aprovaram a versão final a ser publicada.

Referências

- ALMEIDA, C.; XABREGAS, D. A Avaliação no Plano Nacional de Educação–PNE (2014–2024). **Examen: Política, Gestão e Avaliação da Educação**, Brasília, v. 1, n. 1, p. 127-147, 2017.
- ADDEY, C.; SELLAR, S.; STEINER-KHAMSI, G.; LINGARD, B.; VERGER, A. The rise of international large-scale assessments and rationales for participation. **Compare: A Journal of Comparative and International Education**, London, v. 47, n. 3, p. 434-452, 2017.
- BALL, S. J. Diretrizes políticas globais e relações políticas locais em educação. **Currículo sem fronteiras**, Rio de Janeiro, v. 1, n. 2, p. 99-116, 2001.
- BAUER, A.; ALAVARSE, O. M.; OLIVEIRA, R. P. D. Avaliações em larga escala: uma sistematização do debate. **Educação e Pesquisa**, São Paulo, v. 41, [s.n.], p. 1367-1384, 2015.
- BONAMINO, A.; FRANCO, C. Avaliação e política educacional: o processo de institucionalização do Saeb. **Cadernos de Pesquisa**, Campinas, [s.v.], n. 108, p. 101-132, 1999.
- BONAMINO, A.; SOUSA, S. Z. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. **Educação e Pesquisa**, São Paulo, v. 38, [s.n.], p. 373-388, 2012.
- BRASIL. **Constituição Federal**. Promulgada em 5 de outubro de 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm. Acesso em: 10 fev. 2025.
- BRASIL. **Lei n. 9.394**, de 20 de dezembro de 1996. Estabelece as Diretrizes e Bases da Educação Nacional. Brasília: Casa Civil, 1996. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19394.htm. Acesso em: 10 fev. 2025.
- BRASIL. Conselho Nacional de Educação. Câmara de Educação Básica. **Parecer CNE/CEB n. 3**, de 12 de março de 1997. Aprecia os Parâmetros Curriculares Nacionais. Brasília: MEC; CEB; CNE, 1997. Disponível em: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=14615-pceb003-97&category_slug=novembro-2013-pdf&Itemid=30192. Acesso em: 20 jan. 2025.
- BRASIL. **Resolução CEB n. 3**, de 26 de junho de 1998. Institui as Diretrizes Curriculares Nacionais para o Ensino Médio. Brasília: MEC; CEB, 1998. Disponível em: http://portal.mec.gov.br/cne/arquivos/pdf/rceb03_98.pdf. Acesso em: 20 jan. 2025.
- BRASIL. **Decreto n. 6.094**, de 24 de abril de 2007. Dispõe sobre a implementação do Plano de Metas Compromisso Todos pela Educação, pela União Federal, em regime de colaboração com Municípios, Distrito Federal e Estados, e a participação das famílias e da comunidade, mediante programas e ações de assistência técnica e financeira, visando a mobilização social pela melhoria da qualidade da educação básica. Brasília: Casa Civil, 2007. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/decreto/d6094.htm. Acesso em: 20 jan. 2025.
- BRASIL. Ministério da Educação. **PDE: Plano de Desenvolvimento da Educação: Prova Brasil: ensino fundamental: matrizes de referência, tópicos e descritores**. Brasília: MEC; SEB; Inep, 2008.
- BRASIL. Ministério da Educação. **Documento final da Conferência Nacional de Educação 2010**. Brasília: MEC, 2010a. Disponível em: http://pne.mec.gov.br/images/pdf/CONAE2010_doc_final.pdf. Acesso em: 12 mai. 2024.
- BRASIL. Câmara dos Deputados. **Projeto de Lei n. 8035/2010**. Aprova o Plano Nacional de Educação para o decênio 2011-2020 e dá outras providências [Projeto de lei de iniciativa do Poder

Executivo]. Brasília: 2010b. Disponível em:

<https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=490116>. Acesso em: 10 fev. 2025.

BRASIL. Câmara dos Deputados. Comissão Especial destinada a proferir parecer ao Projeto de Lei N. 8.035, de 2010 – Plano Nacional De Educação [Parecer às emendas apresentadas ao Substitutivo do Relator, Dep. Angelo Vanhoni (PT-PR), ao Projeto de Lei n. 8035/2010]. Brasília: Câmara dos Deputados, 2011. Disponível em:

https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=947081&filename=Tramitacao-PL%208035/2010. Acesso em: 10 fev. 2025.

BRASIL. **Lei n. 13.005**, de 25 de junho de 2014. Aprova o Plano Nacional de Educação - PNE e dá outras providências. Brasília: Casa Civil, 2014. Disponível em:

https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/113005.htm. Acesso em: 20 jan. 2025.

BRASIL. **Resolução CNE/CP n. 2**, de 22 de dezembro de 2017. Institui e orienta a implantação da Base Nacional Comum Curricular, a ser respeitada obrigatoriamente ao longo das etapas e respectivas modalidades no âmbito da Educação Básica. Disponível em:

http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=79631-rcp002-17-pdf&category_slug=dezembro-2017-pdf&Itemid=30192. Acesso em: 20 jan. 2025.

BRASIL. **Lei n. 14.113**, de 25 de dezembro de 2020. Regulamenta o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb), de que trata o art. 212-A da Constituição Federal; revoga dispositivos da Lei nº 11.494, de 20 de junho de 2007; e dá outras providências. Brasília: Casa Civil, 2020a. Disponível em:

https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/114113. Acesso em: 20 jan. 2025.

BRASIL. **Emenda Constitucional n. 108**, de 26 de agosto de 2020. Altera a Constituição Federal para [...] dispor sobre o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb); altera o Ato das Disposições Constitucionais Transitórias; e dá outras providências. Brasília: Casa Civil, 2020b. Disponível em:

https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc108.htm. Acesso em: 20 jan. 2025.

BRASIL. Ministério da Educação. **Documento final da Conferência Nacional de Educação 2024**. Brasília: MEC, 2024. Disponível em: <https://campanha.org.br/acervo/documento-final-conae-2024/>. Acesso em: 10 fev. 2025.

BURIASCO, R. Sobre Avaliação em Matemática: uma reflexão. **Educação em Revista**, Belo Horizonte, v. 28, n. 36, p. 255-263, 2002.

BÚRIGO, E. Z. Quando os números falam mais alto: imposições, consentimentos e contestações ao reducionismo curricular. **Revista e-Curriculum**, São Paulo, v. 19, n. 4, p. 1513-1541, 2021.

CARVALHO, D. L. A Educação Matemática de Jovens e Adultos e o Ensino Médio [Texto da intervenção apresentada à Mesa 4]. In: ENCONTRO NACIONAL DE EDUCAÇÃO MATEMÁTICA, 7., 2001, Rio de Janeiro. **Anais...** Rio de Janeiro: SBEM, 2001. p. 1-6. CD-Rom. Disponível em: <https://www.sbembrasil.org.br/sbembrasil/index.php/anais/enem>. Acesso em: 10 fev. 2025.

CASTRO, M. H. G. O Saeb e a Agenda de Reformas Educacionais: 1995- 2002. **Em Aberto**, Brasília, v. 29, n. 96, p. 85-98, 2016.

FERNANDES, R. **Índice de Desenvolvimento da Educação Básica (IDEB)**: metas intermediárias para a sua trajetória no Brasil, estados, municípios e escolas. Brasília: INEP; MEC, 2007.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA – INEP. **SAEB 2001**: novas perspectivas. Brasília: Inep, 2002.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA – INEP. **Relatório Nacional Saeb 2003**. Brasília: Inep, 2006.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA – INEP. **Fundamentação das matrizes de avaliação da educação básica**: estudos e propostas. Brasília: Inep, 2014.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA – INEP. **Relatório SAEB (ANEB e ANRESC) 2005-2015**: panorama da década. Brasília: Inep, 2018.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA – INEP. **Sistema de Avaliação da Educação Básica**. Documentos de Referência. Versão Preliminar. Brasília: Inep, 2019.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA – INEP. **SAEB**: diretrizes da edição de 2023. Brasília: Inep, 2023.

LIMA, I. G.; GANDIN, L. A. O contexto da consolidação das avaliações em larga escala no cenário brasileiro. **Educação & Sociedade**, Campinas, v. 40, e0204183, p. 1-28, 2019. Disponível em: <https://doi.org/10.1590/ES0101-73302019204183>. Acesso em: 10 fev. 2025.

MATHEUS, D. D. S.; LOPES, A. C. Sentidos de qualidade na política de currículo (2003-2012). **Educação & Realidade**, Porto Alegre, v. 39, n. 2, p. 337-357, 2014.

MEYER, H. D. The OECD as pivot of the emerging global educational accountability regime: How accountable are the accountants? **Teachers College Record**, New York, v. 116, n. 9, p. 1-20, 2014.

MIGUEL, A. Entre jogos de luzes e de sombras: uma agenda contemporânea para a educação matemática brasileira. **Perspectivas da Educação Matemática**, Campo Grande, v. 9, n. 20, p. 323-365, 2016.

MOREIRA, A. F. B. Os Parâmetros Curriculares Nacionais em questão. **Educação & Realidade**, Porto Alegre, v. 21, n. 1, p. 9-22, 1996.

ONUCHIC, L. D. L. R.; ALLEVATO, N. S. G. Pesquisa em Resolução de Problemas: caminhos, avanços e novas perspectivas. **Bolema**, Rio Claro, v. 25, n. 41, p. 73-98, 2011.

PESTANA, M. I. G. S.; OKUDA, M. M.; UEMA, L. M.; FIGUEIREDO, M.; TRIGO, M. C. L. M. **Matrizes curriculares de referência para o SAEB**. Brasília: Inep, 1999.

PETTERSSON, D.; POPKEWITZ, T. S.; LINDBLAD, S. On the use of educational numbers: Comparative constructions of hierarchies by means of large-scale assessments. **Espacio, Tiempo y Educación**, Salamanca, v. 3, n. 1, p. 177-202, 2016.

RIBEIRO, R. M.; SOUSA, S. Z. A controvérsia sobre avaliações em larga escala no Brasil: continuum argumentativo. **Educação e Pesquisa**, São Paulo, v. 49, e250287, 2023. Disponível em: <https://doi.org/10.1590/S1678-4634202349250287>por. Acesso em: 10 fev. 2025.

SANTOS, A. A.; HORTA NETO, J. L.; JUNQUEIRA, R. D. **Sistema Nacional de Avaliação da Educação Básica** (Sinaeb). Proposta para atender ao disposto no Plano Nacional de Educação. Brasília: Inep, 2017.

SCHWARTZMAN, S. Uses and abuses of education assessment in Brazil. **Prospects**, Geneva, v. 43, n. 3, 269-288, 2013.

STACEY, K.; TURNER, R. **Assessing Mathematical Literacy**. The PISA Experience. Cham: Springer, 2015.

TAVARES, C. Z. Teoria da Resposta ao Item: uma análise crítica dos pressupostos epistemológicos. **Estudos em Avaliação Educacional**, São Paulo, v. 24, n. 54, p. 56-76, 2013.

TRAVITZKI, R. Qual é o grau de incerteza do Ideb e por que isso importa? **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 28, n. 107, p. 500-520, 2020.

VILLANI, M.; OLIVEIRA, D. A. Avaliação Nacional e Internacional no Brasil: os vínculos entre o PISA e o IDEB. **Educação & Realidade**, Porto Alegre, v. 43, n. 4, 1343-1362, 2018.

Submetido em: 13 de Maio de 2024.

Aprovado em: 21 de Outubro de 2024.

Editor-chefe responsável: Prof. Dr. Marcus Vinicius Maltempi

Editor associado responsável: Profa. Dra. Maria Laura Magalhães Gomes