

O processo de desenvolvimento e institucionalização do Saeb

The Saeb development and institutionalization process

Talita Emidio Andrade Soares¹, Denilson Junio Marques Soares², Wagner dos Santos³

RESUMO: Este artigo foi desenvolvido com o objetivo de discutir a teoria adotada pelo Saeb, considerando o que e como se avalia, e também como esses aspectos contribuíram para o seu processo de desenvolvimento e institucionalização. Emprega uma abordagem metodológica qualitativa e exploratória, baseada na pesquisa bibliográfica e no paradigma indiciário. Como resultados, destaca-se a influência da adoção da Teoria de Resposta ao Item e das Matrizes de Referência para a utilização do sistema como indutor de políticas públicas, como os Planos Nacionais de Educação. Sinaliza, ainda, para o esvaziamento da participação das Secretarias Estaduais e Municipais de Educação, decorrente, sobretudo, da centralidade dos serviços oriundos da avaliação externa aplicada em larga escala que passaram a ser coordenados pelo Inep.

PALAVRAS-CHAVE: Sistema de Avaliação da Educação Básica; Avaliação Educacional; Política de Avaliação.

ABSTRACT: This article was developed with the objective of discussing the theory adopted by Saeb, considering what and how it is evaluated, and also how these aspects contributed to its development and institutionalization process. It uses a qualitative and exploratory methodology, assuming the bibliographic research and the evidential paradigm as theoretical-methodological approaches. As a result, it shows the influence of the adoption of the Item Response Theory and the Reference Matrices for the use of the system as an inducer of public policies, such as national education plans. It also signals the fading of the participation of State and Municipal Departments of Education, resulting, above all, from the centrality of services arising from external evaluation applied on a large scale, which started to be coordinated by Inep.

KEYWORDS: Basic Education Evaluation System; Educational Assessment; Evaluation Policy.

INTRODUÇÃO

Desde o final da década de 1990, o modelo de avaliação por meio de testagens em larga escala ganhou notoriedade no Brasil, inserindo-se na administração e gestão das redes de ensino (estaduais e municipais) do país (BAUER, 2019). Essa centralidade é evidenciada nos Planos

¹ Doutoranda em Educação (UFES). Professora vinculada à Secretaria de Educação do Estado do Espírito Santo, Vitória-ES, Brasil. E-mail: talitaeandrade@gmail.com Orcid: <https://orcid.org/0000-0003-2692-4941>

² Doutor em Educação (UFES). Professor do Ensino Básico, Técnico e Tecnológico no Instituto Federal de Minas Gerais, campus Ouro Preto, MG, Brasil. E-mail: denilson.marques@ifmg.edu.br Orcid: <https://orcid.org/0000-0003-3075-3532>

³ Doutor em Educação (UFES). Professor e Coordenador do Programa de Pós-Graduação em Educação - PPGE/UFES, Vitória-ES, Brasil. E-mail: wagnercefd@gmail.com Orcid: <https://orcid.org/0000-0002-9216-7291>

Nacionais de Educação, em que a avaliação é definida como uma forma de garantir a supervisão e inspeção da qualidade do ensino ofertado pelas escolas (BRASIL, 2001; 2014).

Nesse sentido, para o monitoramento e o cumprimento das metas estabelecidas pelos Planos, utilizam-se os indicadores fornecidos pelo Sistema de Avaliação da Educação Básica (Saeb), cuja primeira aplicação ocorreu em 1990. O Saeb é um sistema de avaliação externa em larga escala, composto por um conjunto de instrumentos e realizado periodicamente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) (BRASIL, 2021), autarquia do Ministério da Educação (MEC).

As avaliações externas, conforme Vianna (2010, p. 18), “[...] são recomendáveis, na medida em que representam um trabalho não comprometido com a administração educacional e as políticas que o orientam”. Em outras palavras, são operacionalizadas por órgãos externos, uma vez que se pretende aferir uma visão externa e supostamente isenta em relação aos sistemas educacionais (VIANNA, 2010).

Atualmente, o Saeb é aplicado na educação infantil e no 2º ano do ensino fundamental de maneira amostral e às etapas finais de cada ciclo da educação básica (5º e 9º anos do ensino fundamental e 3º do ensino médio) com caráter censitário. O sistema avalia os estudantes por meio de testes nas disciplinas de Língua Portuguesa e Matemática, mas também avança para a aplicação em outras áreas do conhecimento, como Ciências da Natureza e Ciências Humanas.

Durante esses trinta anos, o Saeb passou por algumas modificações. A primeira foi em 1995, quando o Inep encomendou uma avaliação externa do sistema que apontou uma série de questões a serem consideradas para o seu aperfeiçoamento. Entre essas, destacam-se a adoção de uma nova metodologia e a preocupação com os conteúdos mínimos comuns referenciados das provas.

Nessa perspectiva, este artigo tem como objetivo analisar a teoria adotada, considerando o que e como se avalia, e também como esses aspectos contribuíram para o processo de desenvolvimento e institucionalização do Saeb. Procurou-se, ainda, esclarecer quais são os seus desdobramentos na constituição de uma política educacional. Para tal, estabeleceu-se, como recorte temporal, o período compreendido entre os anos de 1990 e 2001, marcado pelas principais modificações do sistema. O esforço, assim, é responder à seguinte questão: de que maneira essas alterações metodológicas contribuíram para a efetivação de uma política de avaliação e quais são suas implicações para a educação básica?

Desse modo, além desta introdução, este artigo está estruturado em outras cinco seções: na primeira, apresenta-se a metodologia utilizada; na segunda, discute-se acerca do processo de desenvolvimento e institucionalização do Saeb; na terceira e na quarta seções, abarcam-se as duas principais modificações do período de estudo: a adoção da Teoria de Resposta ao Item (TRI) e a

adoção das Matrizes de Referência; na quinta e última seção, apresentam-se os apontamentos finais específicos, encerrando assim o artigo.

METODOLOGIA

O presente trabalho adota uma metodologia de natureza qualitativa e do tipo exploratória, assumindo a pesquisa bibliográfica e o paradigma indiciário (GINZBURG, 1989) como abordagens teórico-metodológicas. Utiliza como fontes: as publicações acadêmicas relacionadas com o Saeb; entrevistas jornalísticas com os sujeitos envolvidos no processo; e os documentos governamentais produzidos com o intuito de regulamentar e direcionar as ações dos governos brasileiros em prol das políticas de avaliação da educação básica brasileira, disponibilizadas, na íntegra, na internet. Desse modo, discute as principais modificações sofridas pelo sistema que contribuíram para o seu processo de desenvolvimento e institucionalização. Para isso, realiza um estudo das alterações mais importantes, desenvolvidas no período compreendido entre os anos de 1990 e 2001, que auxiliaram na efetivação de uma Política Nacional de Avaliação: a adoção da Teoria de Resposta ao Item (TRI) e das Matrizes de Referência. Esses documentos estão descritos a seguir no Quadro 1.

Quadro 1. Documentos Normativos utilizados como fontes para o desenvolvimento desta pesquisa

Ano	Documento	Resumo
1988	Ministério da Educação	Constituição da República Federativa do Brasil de 1988
1993	Ministério da Educação	Plano Decenal de Educação para Todos 1993-2003
1994	Ministério da Educação	Sistema Nacional de Avaliação da Educação Básica: objetivos, diretrizes, produtos e resultados
1994	Portaria nº. 1.795, de 27 de dezembro de 1994	Cria o Sistema de Avaliação da Educação Básica (Saeb)
1996	Lei nº. 9.934, de 20 de dezembro de 1996	Lei de Diretrizes e Bases da Educação Nacional (LDB)
1998	Ministério da Educação	Matrizes Curriculares de Referência para o Saeb/97
2001	Lei nº. 10.172, de 9 de janeiro de 2001	Aprova o Plano Nacional de Educação (PNE 2001)
2014	Lei nº. 13.005, de 25 de junho de 2014	Aprova o Plano Nacional de Educação (PNE 2014)
2020	Ministério da Educação/Instituto	Matrizes de Referência de Língua Portuguesa e

	Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira	Matemática do Saeb: documento de referência do ano de 2001
2021	Portaria n°. 10, de 8 de janeiro de 2021	Estabelece parâmetros e fixa diretrizes gerais para implementação do Saeb, no âmbito da Política Nacional de Avaliação da Educação Básica
2021	Portaria n°. 250, de 5 de julho de 2021	Estabelece as diretrizes de realização do Saeb em 2021

Fonte: Autores, 2022

Pretende, assim, buscar “pistas”, “indícios”, “vestígios” (GINZBURG, 1989), na tentativa de compreender quais são os interesses defendidos durante o processo de institucionalização do Saeb e suas possíveis implicações para o desenvolvimento da avaliação estandardizada no Brasil. Também se esforça para compreender como as condutas não ditas oficialmente podem demonstrar a influência e/ou imposição de organismos multilaterais, e outras instâncias, no seu processo de origem.

O PROCESSO DE DESENVOLVIMENTO E INSTITUCIONALIZAÇÃO DO SAEB (1990-1995)

Após a primeira aplicação nacional do Saeb, em 1990, o MEC definiu que novas avaliações ocorreriam a cada dois anos. Se, por um lado, a escolha desse intervalo se deu por este ser um período suficiente para efetuar a coleta de dados e a divulgação dos resultados (WAISELFISZ; HORTA NETO, 2016), por outro, pretendia-se cumprir uma das condicionantes do Banco Mundial, que incentivou a adoção de sistemas de avaliação educacional com características amostrais e bianuais nos países latino-americanos em troca de financiamentos (MENDONÇA, 2014) que, por sua vez, eram visados pelo Governo Federal.

Entretanto, o segundo ciclo⁴ de avaliação do Saeb, que estava previsto para 1992, precisou ser reprogramado para o ano seguinte, pois não haviam recursos financeiros suficientes para sua realização e nem financiamentos externos (HORTA NETO, 2007; WAISELFISZ; HORTA NETO, 2016).

Dessa forma, o segundo ciclo de avaliação ocorreu em 1993, financiado pelos próprios Estados da Federação. Entre as atribuições por eles assumidas, destaca-se o levantamento de dados que, segundo Pestana (2016, p. 75), foi o “[...] aspecto decisivo e o que mais contribuiu para a

⁴ Cada avaliação do Saeb, segundo Horta Neto (2007), é conhecida como ciclo de avaliação por apresentar um trabalho árduo: na elaboração e aplicação das provas; no processamento e análise dos dados; e na elaboração dos relatórios e informes para divulgação dos resultados.

institucionalização do sistema e, principalmente, para introdução e disseminação de uma cultura de avaliação no setor educacional”. Isso gerou, nas Secretarias Estaduais de Educação, um maior interesse em dar continuidade à obtenção de informações, visando à melhoria da qualidade da educação (BRASIL, 1994a).

Diante disso, percebeu-se a necessidade de aprimorar os instrumentos de avaliação que constituíam o Saeb, tendo em vista a sua consolidação e aperfeiçoamento. Assim, o Saeb passou a receber apoio institucional do Plano Decenal de Educação para Todos 1993-2003 (BRASIL, 1993), que definiu o sistema como ação prioritária das diversas instâncias envolvidas, como Secretarias Estaduais e Municipais de Educação, universidades, conselhos e órgãos de classe (BRASIL, 1994a).

De acordo com o referido plano, o Saeb possuía a finalidade de “[...] aferir a aprendizagem dos alunos e o desempenho das escolas de primeiro grau e prover informações para avaliação e revisão de planos e programas de qualificação educacional” (BRASIL, 1993, p. 59). Além disso, seus objetivos,

[...] compreendem a pesquisa e desenvolvimento de métodos e instrumentos de avaliação escolar, a implementação, em universidades e centros de pesquisa, de núcleos ou centros regionais de estudos em avaliação, a formação e capacitação de recursos humanos para a área e, por fim, a produção de pesquisa aplicada para subsidiar políticas públicas de qualidade educacional (BRASIL, 1993, p. 59).

Esses objetivos aos poucos foram sendo alcançados. Os resultados obtidos pelo Saeb possibilitaram a consolidação de um banco de dados em nível nacional e por unidade federativa. Além disso, conforme Brasil (1994a), houve, também, a incorporação dos resultados na formulação de políticas públicas e no planejamento do ensino, bem como na elaboração de propostas curriculares desenvolvidas pelas Secretarias Estaduais e Municipais de Educação.

Conjuntamente, ainda segundo Brasil (1994a), deu-se a estruturação de núcleos regionais ligados a instituições universitárias e de pesquisa. A partir dos dados do Saeb e de outras pesquisas educacionais, esses núcleos contribuíram para o desenvolvimento teórico e metodológico da avaliação e para o treinamento de recursos humanos em nível de pós-graduação na área.

Verificou-se, ainda, uma progressiva valorização da avaliação na educação básica realizada por alguns órgãos, como: o Conselho Nacional de Secretários de Educação (Consed); a União Nacional dos Dirigentes Municipais de Educação (Undime); o Fórum dos Conselhos Estaduais de Educação; o Conselho Federal de Educação e outros (BRASIL, 1994a).

Desse modo, como as políticas educacionais deveriam oportunizar uma educação básica de qualidade com equidade e eficiência (BRASIL, 1993), para a sua efetividade, entendeu-se que era necessário o monitoramento permanente e a avaliação dos resultados dessas políticas. Para a realização desses processos (monitoramento e avaliação), o Saeb foi institucionalizado por meio da

Portaria n°. 1.795, de 27 de dezembro de 1994 (BRASIL, 1994b), e passou a vigorar com os seguintes objetivos gerais:

[...] 1. contribuir para o desenvolvimento, em todos os níveis do sistema educativo, de uma cultura avaliativa que estimule a melhoria dos padrões de qualidade e equidade da educação básica e adequados controles sociais de seus resultados; 2. implementar e desenvolver, em articulação com as Secretarias de Educação, processos permanentes de avaliação, apoiados em ciclos regulares de aferições das aprendizagens e competências obtidas pelos alunos e do desempenho dos estabelecimentos que compõem os sistemas de educação básica; 3. mobilizar os recursos humanos, técnicos e institucionais do Ministério, das Secretarias e de universidades e centros de estudos e pesquisas sociais e educacionais, para gerar e difundir os conhecimentos, técnicas e instrumentos requeridos pelas práticas da aferição e avaliação educacional; 4. proporcionar, aos responsáveis pela gestão educacional e pelo ensino, às famílias e aos cidadãos em geral, informações seguras e oportunas a respeito do desempenho e dos resultados dos sistemas educativos (BRASIL, 1994b, p. 20.768).

Além disso, é preciso destacar que a referida portaria previa, também, a elaboração de um Plano Nacional de Avaliação da Educação Básica com diretrizes, normas básicas e objetivos específicos dos ciclos de avaliação, bem como seus padrões de desempenho e qualidade, cronogramas e orçamentos de execução (BRASIL, 1994b).

Para Mendonça (2014), o desenvolvimento do Saeb e a aprendizagem técnica oferecida por ele foram essenciais para a evolução das políticas educacionais durante a década de 1990. O sistema contribuiu, inclusive, para o surgimento de iniciativas estaduais e municipais de avaliação externa aplicada em larga escala (GATTI, 2013).

Vale a pena destacar as iniciativas estaduais que, de acordo com Freitas (2013), utilizaram o desenho censitário (ou quase) com base no Saeb. Conforme a autora, além de servir como referência, o sistema também disponibilizava seus instrumentos, principalmente os itens das provas. Dessa maneira, reproduziam-se, portanto, a própria lógica utilizada pelo Saeb. Dentre elas, destacam-se as avaliações realizadas pelos Estados do Paraná, a partir de 1989; de Minas Gerais e São Paulo, a partir de 1992; e do Ceará, a partir de 1996 (GATTI, 2013).

Outra iniciativa estimulada pelo surgimento do Saeb foi a avaliação das escolas padrão, realizada pela Fundação Carlos Chagas (FCC), em 1992 e 1993. Encomendada pela Secretaria de Educação do Estado de São Paulo, avaliou-se o desempenho dos estudantes da 8ª série do ensino fundamental em Língua Portuguesa, Matemática, Conhecimentos Gerais e Redação. Segundo Gatti (2013), o objetivo era estudar o nível de habilidade envolvido, a dificuldade e discriminação dos itens e o seu comportamento, visando a futuras aplicações.

Dando continuidade à trajetória do Saeb, para Pestana (1998), começar o sistema utilizando provas únicas em todo o território nacional não foi uma tarefa fácil. Apesar disso, a autora destaca

que, a partir de 1995, todos os Estados e as suas respectivas redes de ensino passaram a participar voluntariamente do sistema.

Nesse mesmo ano, início da gestão do governo Fernando Henrique Cardoso (FHC), foram levantados questionamentos acerca da validade da prova realizada pelo sistema. Por esse motivo, o Inep/MEC encomendou a especialistas internacionais uma avaliação externa do próprio Saeb, que trouxe à tona algumas recomendações, sugerindo modificações nos instrumentos e nas técnicas utilizadas nos ciclos anteriores (PESTANA, 1998; HORTA NETO, 2007; CASTRO, 2016).

Com relação às recomendações do estudo, Mendonça (2014) ressalta que muitas foram enriquecidas por consultas a especialistas do Banco Mundial e a entidades brasileiras da área de avaliação educacional. Em especial, destacam-se a FCC e o Centro de Seleção de Candidatos do Grande Rio (Cesgranrio) que, além da experiência em exames standardizados, também já começavam a utilizar a TRI (CASTRO, 2016).

Dessa forma, graças às parcerias do Inep/MEC com organizações multilaterais internacionais⁵ e nacionais, como as Secretarias Estaduais e Municipais de Educação, o Consed e as UFs, passou-se a compartilhar perspectivas sobre o processo educacional dessas instituições (MENDONÇA, 2014). Como consequência, houve avanços significativos para o desenvolvimento do sistema avaliativo do país.

Como resultado, o ciclo de avaliação de 1995 passou por uma série de modificações, podendo ser considerado um marco significativo nas alterações do Saeb. A partir daí, o sistema: passou a avaliar as séries finais de cada etapa da educação básica da rede pública (4ª e 8ª séries do ensino fundamental e 3º ano do ensino médio); incluiu as escolas da rede privada, cuja participação era restrita ao ensino fundamental; e procurou focar a avaliação do desempenho escolar dos estudantes nas disciplinas Língua Portuguesa e Matemática (BRASIL, 2020).

Ademais, nesse ciclo foram incorporadas outras técnicas e instrumentos de avaliação, como o levantamento de dados sobre as características socioeconômicas e culturais dos estudantes por meio de questionários contextuais, preenchidos por eles e pelos seus professores. Além disso, o sistema também sofreu uma alteração metodológica na construção do teste e na análise dos seus resultados com a adoção da TRI, cuja principal contribuição consiste na possibilidade de comparar os resultados das avaliações ao longo do tempo (BRASIL, 2020).

⁵ As instituições, segundo Mendonça (2014), eram: a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco), a *Oficina Regional de Educacion para a America Latina y el Caribe* (Orealc), a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), o *Trends in International Mathematics and Science Study* (TIMSS), o Programa das Nações Unidas para Desenvolvimento/Organização das Nações Unidas (PNUD/ONU), o Bird e o Banco Interamericano de Desenvolvimento (BID).

A ADOÇÃO DA TEORIA DE RESPOSTA AO ITEM (TRI)

Em busca de uma maior sofisticação técnica, que pudesse auxiliar no processo de ampliação do Saeb, a partir do ciclo de 1995, o sistema incorporou a TRI como metodologia para a construção das avaliações e análise dos seus resultados. Essa foi uma iniciativa inédita no cenário nacional.

Para tanto, considerando a ausência de profissionais especializados na área, a sofisticação técnica exigida para sua implementação e as dificuldades operacionais próprias de uma avaliação de porte nacional (HORTA NETO, 2007), o Inep optou por realizar parcerias com fundações de pesquisas que já possuíam certa experiência na área (BONAMINO; FRANCO, 1999), como a FCC e a Cesgranrio. Cabe ressaltar que essas fundações contribuíram de forma significativa no desenvolvimento das primeiras experiências de avaliação educacional (em nível nacional) no país (GATTI; VIANNA; DAVIS, 1991; VIANNA, 1995; SOARES, T.; SOARES, D.; SANTOS, 2022).

Desse modo, essas fundações ficaram responsáveis pela elaboração, aplicação e análise dos resultados das provas, bem como a definição de matrizes de conteúdo que, por sua vez, serviriam de base para a construção dos itens que passariam a compor as avaliações (BONAMINO; FRANCO, 1999; MENDONÇA, 2014; PESTANA, 2016).

De acordo com Ruben Klein (2013), pesquisador da Cesgranrio e entusiasta do uso da TRI nos exames padronizados, a adoção dessa teoria no Saeb introduziu uma verdadeira revolução metodológica nas avaliações brasileiras. Essa revolução impulsionou outras iniciativas semelhantes, como: o Exame Nacional de Cursos (ENC), conhecido como Provão, a partir de 1996; o Exame Nacional do Ensino Médio (Enem), a partir de 1998; o Exame Nacional para Certificação de Competências de Jovens e Adultos (Encceja), a partir de 2001; o Exame de Desempenho dos Estudantes (Enade), que substituiu o Provão a partir de 2004; e o Sistema Nacional de Avaliação da Educação Superior (Sinaes), também a partir de 2004 (CASTRO, 2016; RABELO, 2013).

A adoção da TRI como metodologia para a construção de testes e análises de resultados implicou a transferência do objeto central das análises. Passou-se, então, a analisar cada item individualmente. Antes, porém, considerava-se o teste como um todo, cujas análises eram pautadas, exclusivamente, na Teoria Clássica dos Testes (TCT).

A TCT é uma vertente da Psicometria que, conforme Pasquali (2009, p. 993), procura explicar o escore total obtido em um teste, isto é, a soma das respostas dadas “[...] a uma série de tarefas, tipicamente chamadas de itens”. Essa teoria foi amplamente utilizada no início do século XX e ainda possui grande relevância nas avaliações internas, graças à simplicidade dos cálculos por

ela exigidos e pela facilidade das interpretações dos parâmetros com ela relacionados, que consideram, sobretudo, a dificuldade e a capacidade discriminatória dos itens.⁶

O parâmetro de dificuldade é definido pela TCT como o percentual de acerto do item, de modo que, quanto maior for esse percentual, mais fácil ele será. Além disso, a nota (ou escore) de um respondente nada mais é do que o total de itens respondidos corretamente. Ou seja, o que é feito, majoritariamente, em todo o percurso da avaliação escolar.

Para avaliar o poder de discriminação de um item, utilizado para analisar a sua capacidade de identificar grupos com diferentes níveis de conhecimento,⁷ a TCT apresenta algumas opções. Entre as mais utilizadas, destacam-se o coeficiente de correlação ponto-bisserial e o índice de discriminação (SOARES, D.; SOARES, T.; SANTOS, 2021).

O primeiro é derivado do coeficiente de correlação de Pearson e visa a mensurar o grau de associação entre a resposta dada a um determinado item e o escore do teste (PASQUALI, 2018; SOARES, 2018). Assim, quanto maior o valor mensurado, maior a discriminação do item. Segundo Tôres (2015), valores acima de 0,30 são considerados satisfatórios.⁸

O índice de discriminação, por sua vez, é obtido com base na “regra dos 27” proposta por Kelley (1939). Para Pasquali (2018), trata-se da diferença entre o índice de dificuldade do item calculado para os 27% de respondentes com melhor (dito grupo superior) e pior (dito inferior) desempenho no teste. Em suma, quanto maior o valor dessa diferença, mais discriminativo será o item.

Embora ainda muito utilizadas em análise de itens, as estatísticas que envolvem a TCT apresentam limitações, principalmente no que se refere à comparabilidade de desempenho de indivíduos que se submetem a diferentes testes. Isso dificulta a realização de estudos longitudinais, que são úteis para avaliar o comportamento de uma política pública ao longo dos anos, por exemplo.

Na TCT, os resultados das análises estatísticas, com fins de avaliar as características psicométricas dos itens do teste, como o índice de dificuldade e discriminação, são dependentes da amostra de sujeitos para os quais eles são calculados (subject-dependent) (PASQUALI, 2018).

⁶ Existem outros parâmetros psicométricos que compõem a TCT que avaliam, por exemplo, a confiabilidade ou fidedignidade do teste como instrumento de medida. Maiores informações acerca desses parâmetros podem ser obtidas em Pasquali (2018).

⁷ Essa característica é muito útil em exames seletivos que precisam classificar os respondentes. Em suma, quanto maior a discriminação de um item, maior a sua capacidade em separar os estudantes com maior e menor habilidade.

⁸ De acordo com Andrade, Laros e Gouveia (2010, p. 426), no que se refere aos procedimentos metodológicos adotados no Saeb 2005: “[...] itens com coeficiente bisserial do gabarito menor ou igual a 0,15; itens com dois coeficientes bisseriais de distratores (alternativas erradas) maiores que 0,10 ou coeficiente bisserial de um distrator maior que a bisserial da alternativa correta, são encaminhados para a análise pedagógica”. Entretanto, não foram encontrados documentos oficiais mais recentes que tratam desses parâmetros.

Dessa forma, um mesmo item pode ser considerado fácil para um determinado grupo com alto grau de conhecimento e difícil para um grupo com grau de conhecimento menor. Analogamente, um item pode apresentar problemas de discriminação se aplicado a uma amostra homogênea, composta por sujeitos com habilidades similares.

Seguindo essa vertente, a TRI foi desenvolvida no sentido de superar essas limitações. Entretanto, deve-se ressaltar que uma teoria não substitui a outra, ambas se complementam. No caso do Saeb, por exemplo, primeiramente os itens são analisados de forma exploratória pela TCT. Se apresentarem problemas, será realizada uma análise pedagógica por especialistas da área antes da estimação pela TRI (ANDRADE; LAROS; GOUVEIA, 2010).

Diferentemente da TCT, cujo interesse estava na produção de “testes de qualidade”, a TRI se inclina para a produção de “itens de qualidade”. De acordo com Andrade, Tavares e Valle (2000, p. 8), trata-se de “[...] um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente”.

Contudo, esses modelos apresentam uma complexidade que torna o uso da TRI inacessível sem o auxílio de softwares computacionais especializados e de uma amostra de tamanho significativo. Isso justifica a sua utilização tardia (a partir de 1980), ao considerar que o seu surgimento, como teoria, data da década de 1950 (SOARES, 2018).

Além disso, o uso da TRI pressupõe a presença de uma habilidade dominante que responde por todos os itens do teste (unidimensionalidade), bem como independência entre as respostas a diferentes itens do teste (independência local). Esses pressupostos são de difícil verificação sem a utilização de técnicas computacionais (ANDRADE; TAVARES; VALLE, 2000; PASQUALI, 2018; SOARES, 2018).

Na prática, conforme Muñiz (1990), é comum o uso de modelos logísticos na TRI, justamente por permitirem um melhor tratamento matemático. Esses modelos levam em consideração, respectivamente: a dificuldade do item (modelo logístico de um parâmetro); a dificuldade e a capacidade de discriminação do item (modelo logístico de dois parâmetros); a dificuldade, a capacidade de discriminação e o acerto casual do item (modelo logístico de três parâmetros).

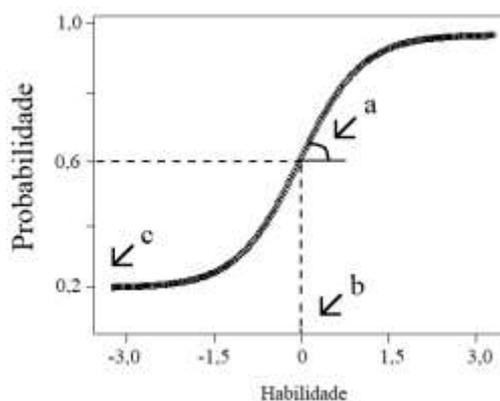
O modelo da TRI utilizado pelo Saeb é o logístico de três parâmetros (ML3P). Esse modelo, proposto por Birnbaum (1968), é expresso matematicamente pela equação:

$$P(U_{ij} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}}$$

Nesse caso, $P(\theta_j, a_i, b_i, c_i)$ representa a probabilidade de o indivíduo j com habilidade θ_j acertar o item i . Além disso, a_i , b_i e c_i são os parâmetros de discriminação, dificuldade e probabilidade de acerto ao acaso do item i , respectivamente.

Essa relação entre a probabilidade de acerto a um item, seus parâmetros psicométricos e a habilidade de o respondente pode ser representada por um gráfico denominado Curva Característica do Item (CCI), cujo comportamento é mostrado na Figura 1.

Figura 1. Curva Característica do item



Fonte: Autores, 2022.

Desse modo, observa-se que os respondentes com maior (menor) habilidade possuem maior (menor) probabilidade de acerto do item, como esperado. Por meio de uma análise geométrica, é possível verificar que o parâmetro de discriminação do item (a_i) é proporcional à inclinação da CCI, no ponto em que a abscissa representa o parâmetro de dificuldade do item (b_i). Esse valor é obtido quando a probabilidade de resposta correta ao item é igual a $(1 + c_i)/2$, no qual c_i representa a probabilidade de acerto casual do item i .

É preciso destacar que os detalhes metodológicos que envolvem o cálculo dos parâmetros psicométricos dos itens do Saeb, bem como a estimação das notas de seus respondentes não foram encontrados em documentos oficiais. Porém, sabe-se que para tal se utiliza o método *Expected a Posteriori* (EAP), estimador bayesiano, que atribui estimativas de proficiências a todos os padrões de respostas, inclusive, os certos ou errados (KLEIN, 2013).

A ESCALA DE PROFICIÊNCIA DO SAEB

Uma das principais vantagens da TRI em relação à TCT está na comparabilidade de resultados de estudantes submetidos a provas distintas. Essa comparabilidade é possível mediante a

construção de uma escala de proficiência, a partir do nível de proficiência de cada participante, estimado via TRI. A escala de proficiência do Saeb está baseada nos resultados do Saeb 1997, cuja média e desvio padrão para cada ciclo da educação básica estão representados na Tabela 1.

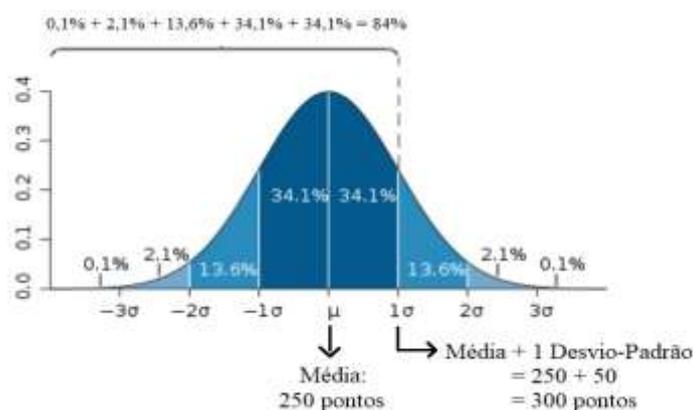
Tabela 1. Limites inferiores e superiores e amplitudes das proficiências do Saeb 1997

Série	Matemática		Língua Portuguesa	
	Média	Desvio padrão	Média	Desvio padrão
4ª série/5º ano do EF	190,8	44	186,5	46
8ª série/9º ano do EF	250,0	50	250,0	50
3ª série do EM	288,7	59	283,9	56

Fonte: Nota Técnica Índice de Desenvolvimento da Educação Básica (BRASIL, 2009, p. 3).

Em termos práticos, um estudante do 9º ano do ensino fundamental que obtiver nota 300 em Língua Portuguesa, por exemplo, estaria a um desvio padrão acima da média (SOARES, D.; SOARES, T.; SANTOS, 2020). Conforme ilustra a Figura 2, esse estudante estaria entre os 16% com melhor desempenho no exame.

Figura 2. Distribuição normal e escala de proficiência da Prova Brasil



Fonte: Soares, Soares e Santos (2020, p. 5).

A partir da construção dessas escalas, são descritas as habilidades que os estudantes são capazes de demonstrar em cada um dos seus níveis. Assim, é possível compreender quais habilidades eles já construíram, estão desenvolvendo ou ainda precisam ser alcançadas. Segundo Rabelo (2013, p. 33), isso permite fazer “[...] comparações e acompanhar progressos de um sistema educacional ao longo dos anos”. Para isso, as provas de um determinado ano incluem alguns itens, ditos âncoras, aplicados no ano anterior.

A partir desses itens, é realizado um processo de equalização (*equating*),⁹ uma espécie de transformação de medidas, de modo a alocar os resultados em uma mesma escala. Para Klein (2009), a metodologia de escolha e a identificação dos níveis âncoras no Saeb deram-se a partir de duas metodologias: a primeira para as edições de 1995 e 1997 e a segunda para as edições a partir de 1999.

Em 1995 e 1997, o Saeb utilizou a metodologia proposta por Beaton e Allen (1992). Nela os itens âncoras caracterizam os pontos ou níveis da escala em que: a grande maioria dos respondentes (65%) situados em torno do nível acerta o item; menos da metade dos estudantes no nível inferior também acerta o item; e a diferença entre os percentuais de acerto é maior do que 30%. Um dos problemas observados com a aplicação dessa metodologia, de acordo com Klein (2009), era o não aparecimento de alguns descritores na descrição, pois, embora obtivessem bons parâmetros psicométricos, não configuravam como âncoras.

Assim, a partir do ciclo de 1999, uma nova metodologia foi aplicada, de modo a identificar itens âncoras e a verificar em quais níveis os estudantes estariam dominando o conteúdo do item. Essa identificação era realizada a partir do agrupamento em intervalos de 25 pontos de uma escala com média de 250 pontos e desvio padrão de 50 pontos.¹⁰ Nessa nova abordagem, passou-se a considerar como âncora em um nível o item em que:

[...] i) O número de alunos no nível que respondeu ao item é maior que 50. ii) O percentual de acerto do item nos níveis anteriores é menor que 65%. iii) O percentual de acerto do item no nível considerado e nos níveis acima é maior que 65%. iv) O ajuste da curva é bom (KLEIN, 2009, p. 139).

Em síntese, conforme Klein (2013), o conceito de interpretação da escala de proficiência está entre as contribuições mais importantes do Saeb, cujo intuito é descrever o que os estudantes sabem e são capazes de fazer em determinados níveis da escala. Para orientar o processo de elaboração dos itens da prova, permitindo a interpretação qualitativa dessas escalas, a partir do ciclo de avaliação de 1997, desenvolveram-se as Matrizes de Referência para o sistema, consideradas um grande marco para os elaboradores de instrumentos (FINI, 2009).

A ADOÇÃO DAS MATRIZES DE REFERÊNCIA

A ideia inicial para a construção de uma matriz de referência para o Saeb, a partir do ciclo de avaliação de 1997, centrava-se na necessidade de definir qual seria o conteúdo mínimo que o

⁹ Segundo Klein (2013), no Saeb tem sido usado o método EAP a priori, sendo a distribuição normal com média 0 e variância 1, $N \sim (0,1)$, para todos os grupos (séries/anos).

¹⁰ Esses valores foram obtidos a partir dos resultados do Saeb 1997. A descrição dos itens da escala de proficiência do Saeb pode ser consultada no site: <http://portal.inep.gov.br/educacao-basica/saeb/matrizes-e-escalas>.

estudante deveria aprender em cada etapa escolar (WAISELFISZ; HORTA NETO, 2016). Em teoria, os resultados da avaliação externa seriam capazes de identificar o quão distante ou próximo os estudantes se encontravam do mínimo estabelecido.

Para a elaboração da primeira versão, o documento Matrizes Curriculares de Referência para o SAEB/97 (BRASIL, 1998) partiu de uma ampla consulta nacional sobre os conteúdos abordados nas escolas brasileiras. Além disso, consideraram: a Lei de Diretrizes e Bases da Educação Básica (LDB, Lei nº 9.394/96); a disseminação das Diretrizes Curriculares Nacionais (DCNs); e as reflexões dos professores e pesquisadores de cada área/disciplina avaliada, bem como as das equipes das Secretarias de Educação (PESTANA, 1998; BRASIL, 1998, 2002; ARAÚJO; LUZIO, 2005; CASTRO, 2016). Essa metodologia também foi adotada no ciclo de 1999.

Em 2001, mais uma vez o Saeb passou por modificações. Com a promulgação da LDB, marco na agenda de reformas do governo FHC, o Saeb referiu-se pela primeira vez à coleta de informações e à avaliação, cujo objetivo era atuar sobre a qualidade do ensino (HORTA NETO, 2007). Conforme consta no art. 9º, coube à União

[...] V - coletar, analisar e disseminar informações sobre a educação; VI - assegurar processo nacional de avaliação do rendimento escolar no ensino fundamental, médio e superior, em colaboração com os sistemas de ensino, objetivando a definição de prioridades e a melhoria da qualidade do ensino (BRASIL, 1996, p. 4).

Dessarte, tendo em vista a garantia do padrão de qualidade da educação básica (BRASIL, 1988, 1996), coube à avaliação o objetivo de “monitorar a qualidade” (LOCATELLI, 2002), sobretudo por meio do Saeb, que se tornou responsável por fornecer informações sobre o desempenho escolar e os fatores a ele associados. Por esse motivo, segundo Castro (2016), a influência da LDB foi fundamental, uma vez que se passou a exigir avaliações periódicas de todos os níveis do sistema educacional.

Para se adaptar às novas exigências, houve, portanto, uma atualização das Matrizes de Referência do Saeb, em 2001. Dessa forma, realizou-se uma nova consulta às UFs, para que suas equipes de ensino e professores regentes das séries/anos avaliados verificassem a compatibilidade entre as matrizes vigentes e o currículo proposto pelos sistemas estaduais de ensino (BRASIL, 2002). Partindo dessa realidade, o Inep e os grupos de especialistas das duas áreas (Língua Portuguesa e Matemática) se reuniram para a elaboração das novas matrizes.

É digno de nota que, em consonância com o documento SAEB 2001: novas perspectivas (BRASIL, 2002), as matrizes de referência não englobam a totalidade do currículo escolar. De acordo com Brasil (2002, p. 15), “[...] foi feito um recorte com base no que pode ser possível aferir por meio do tipo de instrumento de medida utilizado no Saeb e que, ao mesmo tempo, seja representativo do que está contemplado nos currículos vigentes no Brasil”. Por isso, essas matrizes

não devem ser confundidas com os procedimentos, estratégias de ensino ou orientações metodológicas.

As Matrizes de Referência do Saeb, utilizadas para a elaboração dos itens das provas, reúnem o conteúdo a ser avaliado em cada disciplina e série/ano, informando as competências e habilidades esperadas dos estudantes (BRASIL, 2002). Pressupõe-se, portanto, que “[...] o aluno deve desenvolver determinadas competências cognitivas no processo de construção do conhecimento e apresentar habilidades a partir de tais competências” (ARAÚJO; LUZIO, 2005, p. 15).

Essa perspectiva, para Bonamino (2000), configura-se em um deslocamento da preocupação com o domínio dos conteúdos escolares – que demarcou os dois primeiros ciclos do Saeb, em direção à valorização do “saber” e do “saber fazer” do aluno, que se deu a partir do ciclo de 1997, com a adoção das matrizes de referência. Desde então, conforme observado pela autora, as competências cognitivas e as habilidades instrumentais passaram a nortear a formulação dos itens dos testes de desempenho do sistema. Competências cognitivas são entendidas como

[...] as modalidades estruturais da inteligência, operações que o sujeito utiliza para estabelecer relações com e entre os objetos, situações, fenômenos e pessoas (observar, representar, imaginar, reconstruir, comparar, classificar, ordenar, memorizar, interpretar, inferir, criticar, supor, levantar hipóteses, escolher, decidir, etc.) (BRASIL, 1998, p. 7).

As habilidades instrumentais, por sua vez, “[...] referem-se especificamente ao plano do saber fazer e decorrem diretamente do nível estrutural das competências adquiridas e que se transformam em habilidades” (BRASIL, 1998, p.7). Com efeito, a partir da descrição das competências e habilidades que os estudantes deveriam dominar em cada etapa avaliada, foram desenvolvidos os respectivos descritores. Esses associam o objeto de conhecimento (conteúdos curriculares) às operações desenvolvidas pelos estudantes (competências e habilidades) (PESTANA, 1998; BRASIL, 2002; RABELO, 2013).

De acordo com Araújo e Luzio (2005, p. 15), “[...] cada descritor dá origem a diferentes itens e, a partir das respostas dadas a eles, verifica-se o que os alunos sabem e conseguem fazer com os conhecimentos adquiridos”. Assim, conforme Rabelo (2013), os descritores são capazes de expressar quais foram os saberes significativos desenvolvidos pelos alunos durante o processo de ensino-aprendizagem. Desse modo, a primeira versão das matrizes de referência do Saeb, criada em 1997, englobava as disciplinas Língua Portuguesa, Matemática, Ciências, Física, Química e Biologia, avaliadas nesta edição. Já a segunda versão, em 2001, passou a enfatizar somente Língua

Portuguesa e Matemática, devido ao “novo foco” dado pelo sistema. Essas matrizes de referência podem ser consultadas no portal eletrônico do Inep/MEC.¹¹

CONSIDERAÇÕES FINAIS

Levando em consideração a trajetória do Saeb, este artigo empenhou-se em discutir as principais modificações sofridas pelo sistema que contribuíram para o seu processo de desenvolvimento e institucionalização. Para isso, realizou-se um estudo das alterações mais importantes, desenvolvidas no período compreendido entre os anos de 1990 e 2001, que auxiliaram na efetivação de uma Política Nacional de Avaliação: a adoção da TRI e das Matrizes de Referência.

Nota-se que uma das maiores contribuições da preferência pelo uso da metodologia da TRI em relação à TCT consiste na possibilidade de comparação dos resultados ao longo do tempo. Dessa alteração surgiu uma nova utilidade significativa para o sistema que, além de fornecer um diagnóstico da educação básica, também é indutor de políticas públicas que exercem funções importantes para a avaliação, como o juízo de valor e a tomada de decisão.

Além do mais, no que se refere à adoção dessas políticas, o ideal seria que fossem disponibilizadas unidades comparativas, já que assim seria viável avaliar se está ou não havendo melhorias em determinado indicativo de desempenho por meio de sua implementação. Do contrário, não seria possível promover discussões que permitissem corrigir a própria política.

Compreende-se, portanto, que, apesar dos avanços com a adoção da TRI e do caráter censitário da Prova Brasil, a comparação dos resultados pode levar ao uso exacerbado do sistema como política de responsabilização (do inglês, *accountability*), apequenando-se, assim, o que o Saeb se propõe fazer. Dessa forma, aponta-se para a necessidade de estudos que investiguem os *usos e consumos* (GINZBURG, 1989) que se têm feito a partir das médias de desempenho escolar provenientes da Prova Brasil.

Percebeu-se, ainda, que aos poucos a participação das Secretarias Estaduais e Municipais de Educação foi se esvaindo, uma vez que, com a adoção das Matrizes de Referência do Saeb, não foi mais necessária a realização de consulta às Secretarias. Aliás, todos os serviços passaram a ser coordenados pelo Inep/MEC e/ou pelas fundações responsáveis por sua operacionalização.

As Matrizes de Referência do Saeb, apesar de não englobar a totalidade dos currículos (BRASIL, 2002), são responsáveis pelo estabelecimento das competências e habilidades em cada disciplina e ciclo escolar avaliado. É importante destacar que, desde a atualização de 2001,

¹¹ Portal eletrônico do Inep: <https://www.gov.br/inep/pt-br>.

manteve-se a configuração para as disciplinas de Língua Portuguesa e Matemática, que nunca deixaram de ser avaliadas pelo sistema e também compuseram o quadro de outras avaliações que surgiram posteriormente.

Ademais, há *indícios* (GINZBURG, 1989) de que a notoriedade das Matrizes de Referência do Saeb colaborou, inclusive, para a adequação dos sistemas de ensino à BNCC, que irá definir o conjunto de aprendizagens “essenciais” não só para a avaliação, mas também ao longo de todas as etapas e modalidades da educação básica (BRASIL, 2018). Inicialmente, ela foi utilizada pelo Saeb como referência para a elaboração dos itens do 2º ano do ensino fundamental (Língua Portuguesa e Matemática) e do 9º ano para os testes de Ciências da Natureza e Ciências Humanas, aplicados de forma amostral.

Vale a pena destacar que, conforme disposto na Portaria n.º. 10, de 8 de janeiro de 2021 (BRASIL, 2021), o Saeb passará por ajustes técnico-pedagógicos, tendo em vista a implementação da BNCC. Além disso, a portaria prevê que todos os anos/séries da educação básica, das escolas públicas e privadas, serão avaliados anualmente em todas as áreas de conhecimento (BRASIL, 2021c).

Verificou-se, também, a influência da legislação brasileira que estabelece o Saeb como ação prioritária (BRASIL, 1993, 2001), cuja avaliação tem o objetivo de monitorar a garantia do padrão de qualidade da educação básica (BRASIL, 1988, 1996, 2001). Atenta-se, portanto, para a uma crescente utilização dos seus dados como um forte auxílio na tomada de decisões e implementação de políticas, inclusive, para a distribuição de recursos educacionais.

Nesse sentido, o Plano Nacional de Educação (PNE) se caracteriza como um desdobramento do Saeb como política de avaliação. A partir do PNE 2001-2011, instituído pela Lei n.º. 10.172, de 9 de janeiro de 2001 (BRASIL, 2001), os aspectos que dizem respeito à avaliação, como os processos de coleta e difusão dos dados, são considerados “[...] instrumentos indispensáveis para a gestão do sistema educacional e melhoria do ensino” (BRASIL, 2001, p. 4).

Assim, a avaliação é, portanto, uma forma de garantir a supervisão e inspeção da qualidade do ensino ofertado pelas escolas (BRASIL, 2001). Configura-se como um elemento importante capaz de direcionar outras ações do plano, como objetivos e metas. Os resultados produzidos pelo Saeb são utilizados para identificar o cumprimento desses objetivos e metas, o que revela o impacto do sistema e seus desdobramentos na política educacional brasileira.

Esse aspecto também é evidenciado pelo PNE 2014-2024, aprovado por meio da Lei n.º. 13.005, de 25 de junho de 2014 (BRASIL, 2014), atualmente em vigor. O Saeb, além de acompanhar as metas, também é utilizado para fixá-las, juntamente com o Ideb, no qual também se utilizam as médias de desempenho escolar do sistema (Prova Brasil).

Dessa forma, aponta-se para a necessidade de estudos que possam investigar como o Saeb se torna um instrumento de monitoramento da qualidade da educação básica. Além disso, surgem alguns questionamentos, sobretudo com a intenção de compreender quais foram e como ocorreram as primeiras iniciativas de responsabilização e suas implicações para a educação no Brasil.

Finalmente, mais uma vez tornou-se evidente a importância das entidades brasileiras da área de avaliação, como a FCC e a Cesgranrio. Ademais, nota-se que o Saeb contribuiu, gradativamente, para uma valorização da avaliação que aconteceu por vários órgãos durante o período estudado, por exemplo, pelo Consed e pela Undime. Do mesmo modo, foram incentivadas iniciativas estaduais e municipais de avaliação externa, além de impulsionar o aparecimento de outras avaliações nacionais, como o Enem.

Espera-se, portanto, que este artigo esclareça como se deu o processo de desenvolvimento e institucionalização do Saeb e como a adoção da TRI e das Matrizes de Referência, além da legislação brasileira, contribuíram para a sua consolidação e aperfeiçoamento e, conseqüentemente, a efetivação de uma Política Nacional de Avaliação.

REFERÊNCIAS

ARAÚJO, C. H.; LUZIO, N. **Avaliação da educação básica**: em busca da qualidade e equidade no Brasil. Brasília: Inep, 2005.

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. Teoria da resposta ao item: conceitos e aplicações. In: Simpósio Nacional de Probabilidade E Estatística (SINAPE), 14. **Anais**. Caxambu: Associação Brasileira de Estatística, 2000.

BAUER, A. Avaliação de Redes de Ensino e Gestão Educacional: aportes teóricos. **Educação & Realidade**, Porto Alegre, v. 44, n. 1, dez./fev. 2019. DOI: <https://doi.org/10.1590/2175-623677006>

BEATON, A. E.; ALLEN, N. L. Chapter 6: Interpreting Scales Through Scale Anchoring. **Journal of Educational Statistics**, v. 17, n. 2, p. 191-204, 1992. DOI: <https://doi.org/10.2307/1165169>

BIRNBAUM, A. Some latent trait models and their use in inferring an examinee's ability. **Statistical Theories of Mental Test Scores**, Boston, p. 397-479, 1968.

BONAMINO, A.; FRANCO, C. Avaliação e política educacional: o processo de institucionalização do Saeb. **Cadernos de Pesquisa**, São Paulo, n. 108, p. 101-132, nov. 1999. DOI: <http://doi.org/10.1590/S0100-15741999000300005>

BRASIL. Constituição da República Federativa do Brasil de 1988. **Diário Oficial da União**, Brasília, DF: Presidência da República, 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 15 abr. 2022.

BRASIL. Ministério da Educação. **Plano Decenal de Educação para Todos 1993-2003**. Brasília, DF: MEC, 1993. Disponível em: <http://www.dominiopublico.gov.br/download/texto/me001523.pdf>. Acesso em: 18 fev. 2022.

BRASIL. Ministério da Educação. **Sistema nacional de avaliação da educação básica**: objetivos, diretrizes, produtos e resultados. Brasília, DF: MEC/Inep, 1994a.

BRASIL. Portaria n° 1.795, de 27 de dezembro de 1994. **Diário Oficial da União**, Brasília, DF: Presidência da República, 1994b. Disponível em: <https://pesquisa.in.gov.br/imprensa/servlet/INPDFViewer?jornal=1&pagina=91&data=28/12/1994&captchafield=firstAccess>. Acesso em: 15 maio 2022.

BRASIL. Lei n° 9.934, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. **Diário Oficial da União**, Brasília, DF: Presidência da República, 1996. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/19394.htm. Acesso em: 13 mar. 2022.

BRASIL. Ministério da Educação. **Matrizes Curriculares de Referência para o SAEB/97**. Brasília, DF: MEC/Inep, 1998.

BRASIL. Lei n° 10.172, de 9 de janeiro de 2001. Aprova o Plano Nacional de Educação e dá outras providências. **Diário Oficial da União**, Brasília, DF: Presidência da República, 2001. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/leis_2001/110172.htm. Acesso em: 23 maio 2022.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Saeb 2001**: novas perspectivas. Brasília, DF: Inep, 2002. Disponível em: http://portal.inep.gov.br/informacao-da-publicacao/-/asset_publisher/6JYIsGMAMkW1/document/id/484540. Acesso em: 10 fev. 2022.

BRASIL. Lei n° 13.005, de 25 de junho de 2014. Aprova o Plano Nacional de Educação (PNE) e dá outras providências. **Diário Oficial da União**, Brasília, DF: Presidência da República, 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/113005.htm. Acesso em: 22 fev. 2022.

BRASIL. Ministério da Educação. **Base Nacional Comum Curricular**. Brasília, DF: MEC, 2018. Disponível em: http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versaofinal_site.pdf. Acesso em: 22 fev. 2022.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Matrizes de Referência de língua portuguesa e matemática do Saeb**: documento de referência do ano de 2001. Brasília, DF: Inep, 2020. Disponível em: https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_exames_da_educacao_basica/matriz_de_referencia_de_lingua_portuguesa_e_matematica_do_saeb.pdf. Acesso em: 08 mar. 2021.

BRASIL. Portaria n° 10, de 8 de janeiro de 2021. Estabelece parâmetros e fixa diretrizes gerais para implementação do Sistema de Avaliação da Educação Básica (Saeb) no âmbito da Política Nacional de Avaliação da Educação Básica. **Diário Oficial da União**, Brasília, DF: Presidência da República, 2021a. Disponível em: <https://www.in.gov.br/en/web/dou/-/portaria-n-10-de-8-de-janeiro-de-2021-298322305>. Acesso em: 22 fev. 2022.

CASTRO. M. H. G. O Saeb e a agenda de reformas educacionais: 1995 a 2002. **Em Aberto**, Brasília, v. 29, n. 96, p. 85-98, maio/ago. 2016. DOI: <https://doi.org/10.24109/2176-6673.emaberto.29i96.%25p>

FINI, M. I. Currículo e avaliação: articulação necessária em favor da aprendizagem dos alunos da rede pública de São Paulo. **São Paulo em Perspectiva**, São Paulo, v. 23, n. 1, p. 57-72, jan./jun. 2009. Disponível em: http://produtos.seade.gov.br/produtos/spp/v23n01/v23n01_05.pdf. Acesso em: 15 abr. 2022.

FREITAS, D. N. T. Avaliação da educação básica no Brasil: características e pressupostos. In: BAUER, Adriana; GATTI, Bernadette; TAVARES, Marialva R. (org). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origens e pressupostos**. Florianópolis: Insular, 2013. p. 70-96.

GATTI, B. Possibilidades e fundamentos de avaliações em larga escala: primórdios e perspectivas. In: BAUER, Adriana; GATTI, Bernadette; TAVARES, Marialva R. (org). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origens e pressupostos**. Florianópolis: Insular, 2013. p. 47-69.

GATTI, B; VIANNA, H. M.; DAVIS, C. Problemas e impasses da avaliação de projetos e sistemas educacionais: dois casos brasileiros. **Estudos em Avaliação Educacional**, São Paulo, n. 4, p. 7-26, jul./dez. 1991. DOI: <http://dx.doi.org/10.18222/ae00419912374>

GINZBURG, C. **Mitos, emblemas, sinais: morfologia e história**. São Paulo: Companhia das Letras, 1989.

HORTA NETO, J. L. Um olhar retrospectivo sobre a avaliação externa no Brasil: das primeiras mediações até o Saeb de 2005. **Revista Iberoamericana de Educación**, Madrid, v. 42, n. 5, jan./abr. 2007. DOI: <https://doi.org/10.35362/rie4252398>

KELLEY, T. L. The selection of upper and lower groups for the validation of test items. **Journal of educational psychology**, Warwick & York, v. 30, n. 1, p. 17-24, 1939. Disponível em: <https://psycnet.apa.org/record/1939-03313-001>. Acesso em: 13 de abr. 2022.

KLEIN, R. Utilização da teoria de resposta ao item no Sistema Nacional de Avaliação da Educação Básica (Saeb). **Meta: Avaliação**, Rio de Janeiro, v. 1, n. 2, p. 125-140, jul./set. 2009. DOI: <http://dx.doi.org/10.22347/2175-2753v1i2.38>

KLEIN, R. Aspectos Metodológicos e Técnicos: delineamentos assumidos nas avaliações, limites e perspectivas de aprimoramento. In: BAUER, Adriana; GATTI, Bernadette; TAVARES, Marialva R. (org). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origens e pressupostos**. Florianópolis: Insular, 2013b. p. 109-130.

LOCATELLI, I. Construção de instrumentos para a avaliação de larga escala e indicadores de rendimento: o modelo SAEB. **Estudos em Avaliação Educacional**, São Paulo, n. 25, p. 3-21, jan./jun. 2002. DOI: <https://doi.org/10.18222/ae02520022189>

MENDONÇA, L. P. **A reforma educacional na América Latina e a disseminação de políticas públicas de avaliação de sistemas educacionais**. Dissertação (Mestrado em Ciências Sociais) – Universidade Federal de Juiz de Fora, Juiz de Fora, MG, 2014. Disponível em: <http://repositorio.ufjf.br:8080/jspui/handle/ufjf/1830>. Acesso em: 22 fev. 2022.

MUÑIZ, J. **Teoría de respuesta a los ítems: Un nuevo enfoque en la evolución psicológica y educativa**. Madri: Ediciones Pirámide, S. A, 1990.

PASQUALI, L. **TRI–Teoria de resposta ao item**: Teoria, procedimentos e aplicações. Editora Appris, 2018.

PESTANA, M. I. O sistema de avaliação brasileiro. **Revista brasileira de Estudos Pedagógicos**, Brasília, v. 79, n. 191, p. 65-73, jan./abr. 1998. DOI: <https://doi.org/10.24109/2176-6681.rbep.79i191.1044>

PESTANA, M. I. Trajetória do Saeb: criação, amadurecimento e desafios. **Em Aberto**, Brasília, v. 29, n. 96, p. 71-84, maio/ago. 2016. DOI: <http://doi.org/10.24109/2176-6673.emaberto.29i96.%25p>

RABELO, M. L. **Avaliação Educacional**: Fundamentos. Metodologia e Aplicações no Contexto Brasileiro. Rio de Janeiro: Coleção Profmat, Sociedade Brasileira de Matemática, 2013.

SOARES, D. J. M. **Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de matemática básica**. 2018. 121 f. Dissertação (Mestrado em Estatística) - Universidade Federal de Viçosa, Viçosa, MG, 2018.

SOARES, D. J. M.; SOARES, T. E. A.; SANTOS, W. Infraestrutura e desempenho escolar na Prova Brasil: aspectos e conexões. **Olhar de Professor**, Ponta Grossa, v. 23, p. 1-18, jul. 2020. DOI: <https://doi.org/10.5212/10.5212/OlharProfr.v.23.2020.15023.209209227242.0626>

SOARES, D. J. M.; SOARES, T. E. A.; SANTOS, W. Análise da qualidade psicométrica da prova de matemática do Exame Nacional do Ensino Médio brasileiro de 2018. **Actualidades Investigativas en Educación**, v. 21, n. 1, p. 86-115, 2021. DOI: <http://dx.doi.org/10.15517/aie.v21i1.42338>

SOARES, T. E. A.; SOARES, D. J. M.; SANTOS, W. Sistema de Avaliação da Educação Básica: revisão sistemática da literatura. **Estudos em Avaliação Educacional**, v. 32, 2021. DOI: <https://doi.org/10.18222/eae.v32.7839>

VIANNA, H. M. Avaliação educacional: uma perspectiva histórica. **Estudos em Avaliação Educacional**, São Paulo, n. 12, p. 7-24, jul./dez. 1995. DOI: <http://dx.doi.org/10.18222/eae01219952297>

VIANNA, H. M. **Avaliação em debate**: Saeb, Enem, Provão. Brasília: Liber Livro, 2010.

WASELFISZ, J. J.; HORTA NETO, J. L. As origens do Saeb. **Em Aberto**, Brasília, v. 29, n. 96, p. 177-193, maio/ago. 2016. DOI: <https://doi.org/10.24109/2176-6673.emaberto.29i96.2705>