

Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola

Alicia Bonamino

Pontifícia Universidade Católica do Rio de Janeiro

Sandra Zákia Sousa

Universidade de São Paulo

Resumo

Analisa-se, neste artigo, três gerações de avaliação da educação em larga escala, a partir dos objetivos e desenhos usuais em iniciativas implementadas no Brasil. A primeira geração consiste na avaliação diagnóstica da qualidade da educação, sem atribuição de consequências diretas para as escolas e para o currículo escolar. As outras duas gerações articulam os resultados das avaliações a políticas de responsabilização, com atribuição de consequências simbólicas ou materiais para os agentes escolares. Tomando como parâmetro de análise os objetivos e desenhos dessas avaliações, bem como estudos e pesquisas que produziram evidências sobre o tema, exploram-se possíveis implicações para o currículo escolar. Por um lado, discutem-se os riscos de as provas padronizadas, com avaliações que referenciam políticas de responsabilização envolvendo consequências fracas e fortes, exacerbarem a preocupação de diretores e professores com a preparação para os testes e para as atividades por estes abordadas, levando a um estreitamento do currículo escolar. Por outro lado, aponta-se o potencial das avaliações de segunda e terceira gerações em propiciarem uma discussão informada sobre o currículo escolar, em termos das habilidades fundamentais de leitura e matemática que ainda não têm sido garantidas a todos os alunos.

Palavras-chave

Avaliação da educação – Responsabilização – Currículo escolar.

Correspondência:

Alicia Bonamino

**Pontifícia Universidade Católica
do Rio de Janeiro**

Departamento de Educação

Rua Marquês de São Vicente, 225, 1050 L

22451-900 – Rio de Janeiro/RJ

alicia@puc-rio.br

Three generations of assessments of basic education in Brazil: interfaces with the curriculum in/of the school

Alicia Bonamino

Pontifical Catholic University of Rio de Janeiro

Sandra Zákia Sousa

University of São Paulo

Abstract

Three generations of large scale assessments of education are analyzed in this article based on the objectives and designs usual in initiatives of this kind implemented in Brazil. The first generation consists in the diagnostic assessment of the quality of education, without attribution of direct consequences for schools and for school curricula. The other two generations articulate the results of the assessments to accountability policies, with the attribution of symbolic or material consequences for the school agents. Taking as a parameter of analysis the objectives and designs of these assessments, as well as studies and researches that produced results about this theme, possible implications for the school curriculum are explored. On the one hand, a discussion is made of the risks that standardized tests, with evaluations that make reference to accountability policies involving weak and strong consequences, may exacerbate the concern of teachers and principals with the preparation for the tests and for the activities they include, leading to a narrowing of the school curriculum. On the other hand, the text points out the potential of the second and third generation assessments to stimulate an informed discussion of school curriculum in terms of the fundamental abilities of reading and mathematics which have not yet been guaranteed to all pupils.

Keywords

Assessment of education – Accountability – School curriculum.

Contact:

Alicia Bonamino

Pontifícia Universidade Católica

do Rio de Janeiro

Departamento de Educação

Rua Marquês de São Vicente,

225, 1050 L

22451-900 – Rio de Janeiro/RJ

alicia@puc-rio.br

Dentre os marcos presentes na formulação e na implementação das políticas educacionais brasileiras nas duas últimas décadas, ganham destaque as avaliações com elementos comuns a propostas realizadas em outros países, expressando uma agenda mundial. Além de outros objetivos, as iniciativas de avaliação associam-se à promoção da qualidade do ensino, estabelecendo, no limite, novos parâmetros de gestão dos sistemas educacionais.

Em relação ao currículo, na maioria dos países, e independentemente do grau de descentralização ou centralização das formas de regulação dos currículos escolares, o que se constata é uma tendência à utilização de avaliações centralizadas para mensurar o desempenho escolar dos alunos, sob os mesmos parâmetros curriculares aos quais se considera que todos os estudantes deveriam ter acesso.

Essa perspectiva mais universalista é reforçada pelo consenso que parece existir em escala mundial a respeito da pequena variabilidade das propostas curriculares, o que se reflete nos conteúdos das avaliações nacionais e na participação recente de 65 países no *Programa Internacional de Avaliação de Estudantes* (PISA), a partir da ideia de que o currículo de cada país é comparável aos dos outros países envolvidos.

No caso do Brasil, a análise dos desenhos das avaliações em andamento leva a que se identifiquem três gerações de avaliações da educação em larga escala, com consequências diferenciadas para o currículo escolar. Ao tempo em que se sucedem, essas gerações coexistem no âmbito das redes de ensino; daí a necessidade de se tomar tal classificação como um recurso analítico.

A primeira geração enfatiza a avaliação com caráter diagnóstico da qualidade da educação ofertada no Brasil, sem atribuição de consequências diretas para as escolas e para o currículo. No estágio atual das iniciativas de avaliação em larga escala, emergem outros dois novos modelos de avaliação com a finalidade de subsidiar, a partir dos resultados dos alunos,

políticas de responsabilização com atribuição de consequências para os agentes escolares. Na literatura sobre o tema, quando as consequências dessas políticas são apenas simbólicas, elas são chamadas de *low stakes* ou de *responsabilização branda*. Já quando as consequências são sérias, elas são chamadas de *high stakes* ou de *responsabilização forte* (CARNOY; LOEB, 2002; BROOKE, 2006). Tais avaliações são respectivamente identificadas, neste texto, como avaliações de segunda e terceira geração.

No Brasil, avaliações de primeira geração são aquelas cuja finalidade é acompanhar a evolução da qualidade da educação. De um modo geral, essas avaliações divulgam seus resultados na Internet, para consulta pública, ou utilizam-se da mídia ou de outras formas de disseminação, sem que os resultados da avaliação sejam devolvidos para as escolas.

Avaliações de segunda geração, por sua vez, contemplam, além da divulgação pública, a devolução dos resultados para as escolas, sem estabelecer consequências materiais. Nesse caso, as consequências são simbólicas e decorrem da divulgação e da apropriação das informações sobre os resultados da escola pelos pais e pela sociedade. Esse tipo de mecanismo de responsabilização tem como pressuposto que o conhecimento dos resultados favorece a mobilização das equipes escolares para a melhoria da educação, bem como a pressão dos pais e da comunidade sobre a escola (ZAPONI; VALENÇA, 2009).

Avaliações de terceira geração são aquelas que referenciam políticas de responsabilização forte ou *high stakes*, contemplando sanções ou recompensas em decorrência dos resultados de alunos e escolas. Nesse caso, incluem-se experiências de responsabilização explicitadas em normas e que envolvem mecanismos de remuneração em função de metas estabelecidas (ZAPONI; VALENÇA, 2009).

Este artigo procura caracterizar experiências de avaliação da educação básica em curso no país e explora possíveis relações com o currículo escolar. O interesse, assim, é

responder às seguintes questões: em que condições a avaliação da educação em larga escala tem consequências para o currículo escolar?; há evidências disponíveis sobre as interferências da avaliação no currículo escolar? Para tanto, adotam-se como parâmetro para análise os objetivos e desenhos de tais avaliações, bem como estudos e pesquisas que produziram evidências sobre o tema.

Além desta introdução, o texto está organizado em outras quatro seções. A primeira discute as principais características do *Sistema Nacional de Avaliação da Educação Básica* (Saeb) que permitem identificá-lo como uma avaliação de primeira geração. A seção seguinte aborda, em suas subseções, as avaliações de segunda e terceira geração e as relações entre formas diferenciadas de responsabilização e currículo escolar, tomando como exemplos a *Prova Brasil* e as avaliações estaduais de São Paulo e Pernambuco. A terceira seção apresenta uma revisão de estudos e pesquisas sobre as relações entre tais avaliações e o currículo escolar em diferentes contextos. O artigo encerra-se com a apresentação das conclusões.

Primeira geração de políticas de avaliação em larga escala: o Saeb

Embora se tenha evidência de que, desde os anos 1930, havia interesse do Estado em tomar a avaliação como parte do planejamento educacional, é no final dos anos 1980 que a avaliação passa paulatinamente a integrar políticas e práticas governamentais direcionadas à educação básica. Como afirma Dirce Nei Teixeira de Freitas (2007),

[...] foram necessárias mais ou menos cinco décadas para que a avaliação (externa, em larga escala, centralizada e com foco no rendimento do aluno e no desempenho dos sistemas de ensino) viesse a ser introduzida como prática sistemática no governo da educação básica brasileira. (p. 51)

Desde a década de 1960, tem-se a ampliação do uso de testes educacionais¹; no entanto, situa-se nos anos finais da década de 1980 a primeira iniciativa de organização de uma sistemática de avaliação dos ensinos fundamental e médio em âmbito nacional. Essa sistemática é denominada pelo Ministério da Educação (MEC), a partir de 1991, como *Sistema Nacional de Avaliação da Educação Básica* (Saeb)².

O Saeb, principal sistema de avaliação da qualidade da educação básica, avalia, a cada dois anos, uma amostra dos alunos regularmente matriculados na 4ª e na 8ª série (6º e 9º ano) do ensino fundamental e no 3º ano do ensino médio, em escolas públicas e privadas localizadas em área urbana e rural.

Os testes de desempenho aplicados aos alunos são conjugados com questionários sobre fatores associados a esses resultados, tendo por foco a escola e seus diferentes agentes. Até 2009, foram realizados dez ciclos de avaliação.

Desde sua criação³, o Saeb configura-se como uma avaliação com desenho apropriado para diagnosticar e monitorar a qualidade da educação básica nas regiões geográficas e nos Estados brasileiros. Em 1995, foram introduzidas inovações metodológicas em seu desenho, as quais consolidaram sua configuração atual; são elas: i) inclusão da rede particular de ensino na amostra; ii) adoção da *Teoria de Resposta ao Item* (TRI)⁴, que permite estimar as habilidades

1- Cf. GATTI, 1987.

2- Uma caracterização mais detalhada da institucionalização do Saeb é apresentada por Alicia Bonamino (2002).

3- Embora o Saeb tenha-se iniciado anos antes, só em 1994 ele foi criado formalmente por meio da Portaria nº 1.795, de 27 de dezembro.

4- A *Teoria de Resposta ao Item* (TRI) é um modelo matemático que permite estimar a capacidade dos indivíduos em determinada área ou disciplina a partir da premissa de que ela é unidimensional. Vale dizer: presume-se, por exemplo, que os alunos tenham uma capacidade ou competência para a matemática que define a probabilidade de que determinado aluno realize adequadamente as diferentes atividades incluídas no banco de questões. Ela tem algumas vantagens sobre o enfoque clássico, pois permite pôr questões e alunos em uma mesma escala; fazer estimativas mais precisas das mudanças ao longo do tempo, mediante equiparação das pontuações; estimar uma medida da capacidade dos alunos que leva em conta a dificuldade das questões, isto é, as questões mais difíceis têm peso maior na determinação da pontuação individual. Nas pontuações de TRI, o SAEB adota uma média de 250 pontos, o que corresponde à média nacional dos alunos da 8ª série em 1997. A partir das premissas da TRI, é possível construir uma escala única de pontuações para populações de

dos alunos independentemente do conjunto específico de itens respondidos; iii) opção de trabalhar com as séries conclusivas de cada ciclo escolar (4ª e 8ª série do ensino fundamental e inclusão da 3ª série do ensino médio); iv) priorização das áreas de conhecimento de língua portuguesa (foco em leitura) e matemática (foco em resolução de problemas); v) participação das 27 unidades federais; vi) adoção de questionários para os alunos sobre características socioculturais e hábitos de estudo. A partir da introdução dessas inovações, o Saeb tornou comparáveis os desempenhos dos alunos entre anos e séries.

Os testes cognitivos do Saeb são elaborados com base em matrizes de referência, desenhadas a partir de uma *síntese* do que é comum a diferentes propostas curriculares estaduais, municipais e nacionais, além da consulta a professores e especialistas nas áreas de língua portuguesa e matemática e do exame dos livros didáticos mais utilizados nas redes e séries avaliadas.

Ainda que a elaboração dos testes leve à definição do que deve ser considerado fundamental em termos de aprendizagem escolar e, portanto, do que todos os alunos deveriam saber e ser capazes de fazer ao final de determinados ciclos de escolarização, por ser de base amostral, o Saeb apresenta baixo nível de interferência na vida das escolas e no currículo escolar.

Seu desenho mostra-se adequado para diagnosticar e monitorar a evolução da qualidade da educação básica, mas não permite medir a evolução do desempenho individual de alunos ou escolas. Seus resultados são divulgados de forma bastante agregada e, portanto, não permitem apoiar a introdução de políticas de responsabilização de professores, diretores e gestores por melhorias de qualidade nas unidades escolares.

Paralelamente, enquanto o MEC desenvolvia uma avaliação da educação básica de

distintos níveis de escolaridade: no caso do Brasil, 4ª e 8ª séries do ensino fundamental e 3ª do ensino médio. Isso permite comparar as médias de *proficiência* em cada disciplina entre os diversos níveis do sistema educativo, entre as regiões do país e entre os vários anos, situando todos os níveis em uma mesma escala.

base amostral, Estados e municípios sentiam a necessidade de implantar avaliações que atinxissem todas as escolas. Tal necessidade fez com que vários Estados adotassem seus próprios sistemas de avaliação. O Estado de Minas Gerais, por exemplo, criou, em 1991, o *Sistema de Avaliação da Educação Pública* (Simave), e o Ceará, o *Sistema Permanente de Avaliação da Educação Básica* (Spaece), em 1992. Várias outras iniciativas estaduais e municipais vêm sendo conduzidas desde então. No caso das unidades federadas, 14 das 27 possuíam, em 2007, sistemas próprios de avaliação (LOPES, 2007).

A coexistência do Saeb com avaliações estaduais e, anos mais tarde, com a *Prova Brasil* faz com que a ênfase inicial na finalidade diagnóstica no uso dos resultados da avaliação perca força em face da tendência de focalizar esse uso como subsídio a políticas de responsabilização, o que leva ao reconhecimento de duas novas gerações de avaliação da educação básica no Brasil. Como será visto nas seções seguintes, tais avaliações envolvem a publicidade dos resultados dos testes por redes e/ou escolas e, no caso da terceira geração, também o estabelecimento de prêmios atrelados aos resultados dos alunos.

Avaliação em educação e responsabilização

No campo educacional, as avaliações que subsidiam políticas de responsabilização operam crescentemente dentro de um referencial que associa gestão democrática da educação, avaliação e responsabilização. A definição aí subjacente de democracia apoia-se em dois princípios orientadores. Por um lado, há a participação que acontece, em grande medida, mas não exclusivamente, por meio do processo eleitoral e do sistema partidário. Para tanto, todo cidadão deve desfrutar de direitos políticos fundamentais: direito de expressão, de associação, de votar e de candidatar-se a cargos públicos. Por outro lado, há a contestação pública entre vários atores políticos, não apenas no sentido

da competição política, mas, sobretudo, como controle dos governantes pelos governados. Ou seja, “os governantes (enquanto agentes da soberania popular) devem responsabilizar-se perante o povo por seus atos e omissões no exercício do Poder Público” (CENEVIVA, 2005, p. 12).

Esses dois ideais de sistemas democráticos – participação e contestação pública – correspondem a duas formas básicas de responsabilização. A primeira delas é o processo eleitoral, que expressa o controle vertical sobre os governantes, sendo um instrumento de participação política, de garantia da soberania popular, e que assegura, mediante eleições periódicas, a expressão das preferências do povo por meio de mandatos. A segunda forma de responsabilização é o controle institucional durante os mandatos, o qual garante a contestação pública e a fiscalização contínua dos representantes políticos, eleitos ou não, no exercício do Poder Público (CENEVIVA, 2005).

Impulsionada pela constatação de que a democratização do Poder Público deve ir além do voto, essa forma de responsabilização vem sendo considerada na perspectiva do aperfeiçoamento das instituições estatais, o que envolve, ao mesmo tempo, a melhoria das políticas e dos programas governamentais e uma maior transparência e responsabilidade nas ações de política pública.

Nos últimos anos, ganhou relevância a relação entre a qualidade das ações dos governos e os controles e incentivos a que estão submetidos os governantes e a burocracia, bem como entre o fortalecimento dos mecanismos de responsabilização e o aperfeiçoamento das práticas administrativas.

Dois mecanismos, em particular, têm sido apontados no estabelecimento de novas formas de participação e controle da sociedade sobre as ações do Estado: o controle social e o controle de resultados. A introdução de mecanismos de controle social e de responsabilização da administração pública pelo desempenho de políticas e programas governamentais aparece como uma promessa de substituição de um modelo

em que impera o controle burocrático baseado na observância a normas e procedimentos, sem a participação dos cidadãos, por outro no qual se estabelece o controle *a posteriori* dos resultados da ação governamental, com a participação da sociedade. Esse mecanismo pode envolver, ainda, a definição de metas e índices de desempenho, bem como a avaliação direta dos bens e serviços públicos que estão sendo ofertados (CENEVIVA, 2005).

A avaliação de políticas e programas públicos ganha, assim, um lugar de destaque como meio para mensurar seu desempenho e exercer a prestação de contas à sociedade. Nessa perspectiva, a avaliação aparece diretamente ligada ao desempenho da gestão pública, à promoção de maior transparência e à criação de mecanismos de responsabilização.

Nas próximas seções, veremos como a evolução das avaliações educacionais articula-se com essa perspectiva, no interior da qual podemos identificar o surgimento da segunda e da terceira geração de avaliações em larga escala.

Segunda geração de avaliação da educação – responsabilização e currículo: a *Prova Brasil*

A fim de aumentar o conteúdo informacional da avaliação e suas consequências sobre as escolas, foi implementada, a partir de 2005, a *Prova Brasil*, que permite agregar à perspectiva diagnóstica a noção de responsabilização (FERNANDES; GREMAUD, 2009). A justificativa para sua implementação indicava as limitações do desenho amostral do Saeb em retratar as especificidades de municípios e escolas e em induzir dirigentes públicos estaduais e municipais na formulação de políticas para a melhoria do ensino.

A *Prova Brasil*, que ocorre a cada dois anos, foi idealizada para produzir informações a respeito do ensino oferecido por município e escola, com o objetivo de auxiliar os governantes nas decisões sobre o direcionamento de recursos técnicos e financeiros e no estabelecimento de metas e implantação de ações pedagógicas

e administrativas, visando à melhoria da qualidade do ensino. De outra parte, considera-se que essa avaliação pode funcionar como um elemento de pressão, para pais e responsáveis, por melhoria da qualidade da educação de seus filhos, uma vez que, a partir da divulgação dos resultados, eles podem cobrar providências para que a escola melhore.

A introdução da *Prova Brasil* em 2005 e sua repetição, a cada dois anos, permitem a comparação, ao longo do tempo, entre as escolas que oferecem o ensino fundamental. Em sua primeira edição, ela avaliou mais de 3 milhões de alunos em aproximadamente 45.000 escolas urbanas de 5.398 municípios; foi muito além, portanto, do Saeb, que avalia, em média, uma amostra de 300.000 alunos.

A *Prova Brasil* foi censitária para as escolas urbanas em 2005 e 2007. Em 2007, alterou-se o número mínimo de alunos na série avaliada, que passou de trinta para vinte. Essa alteração foi realizada para possibilitar que aproximadamente quatrocentos municípios que não participaram da primeira edição pudessem ser incluídos na avaliação. Já na terceira edição, em 2009, houve a ampliação do universo avaliado, de modo a incluir também todas as escolas rurais que tivessem, no mínimo, vinte alunos nas séries avaliadas.

Os resultados da *Prova Brasil* de 2007 passaram a integrar o *Indicador de Desenvolvimento da Educação Básica* (Ideb⁵), referência para a definição de metas a serem alcançadas, gradualmente, pelas redes públicas de ensino até 2021. O princípio básico de tal indicador é o de que a qualidade da educação envolve que o aluno aprenda e passe de ano. Com o Ideb, o desempenho passa a ser medido por meio da *Prova Brasil* e a aprovação, por meio do *Censo Escolar*. Os índices de aprovação permitem levar em conta o número de anos que, em média, os alunos levam para completar uma série.

5- O Ideb foi criado pelo Instituto Nacional de Estudos e de Pesquisas Educacionais Anísio Teixeira (Inep), órgão vinculado ao MEC. Para informações mais detalhadas, consultar <<http://www.inep.gov.br>> e FERNANDES 2007.

Os resultados da *Prova Brasil* passaram a ser amplamente divulgados e, atualmente, o Ideb é o principal indicador adotado pelo Governo Federal para traçar metas educacionais a serem alcançadas por escolas e redes estaduais e municipais. A ideia central do sistema de metas foi obter um maior comprometimento das redes e escolas com o objetivo de melhorar os indicadores educacionais. Supõe-se que um sistema de metas pactuado entre o MEC e as secretarias de educação de Estados e municípios serviria para aumentar a mobilização da sociedade em favor da qualidade da educação. Nessa perspectiva, é preciso lembrar que o Brasil possui um sistema educacional descentralizado, com mais de 5.000 redes de ensino com autonomia para gerir suas escolas (FERNANDES; GREMAUD, 2009).

A divulgação dos resultados da primeira edição da *Prova Brasil* ocorreu em julho de 2006 por intermédio dos principais meios de comunicação e de um boletim disponibilizado na Internet e enviado a cada uma das escolas participantes. Esse boletim apresentava, entre outras informações, os resultados das escolas em uma escala de desempenho e as médias alcançadas pelas escolas das redes municipal, estadual e federal.

Enquanto a mídia divulgava *rankings* de escolas, com destaque para os melhores e piores resultados, nos sites do Inep e do MEC, enfatizava-se, como novidade da *Prova Brasil*, a devolução dos resultados para as escolas a fim de colaborar com o planejamento das ações pedagógicas (OLIVEIRA, 2011).

Em 2009, meses antes de acontecer a terceira edição da *Prova Brasil*, o Inep e o MEC distribuíram duas publicações em todas as escolas públicas: a *Matriz de Referência da Prova Brasil e do Saeb – Ensino Fundamental* e a *Matriz de Referência do Saeb – Ensino Médio e de Ensino Fundamental*, ambas com exemplos de itens de edições anteriores comentados. No caso da *Prova Brasil*, no entanto, por razões administrativas, os resultados não foram divulgados às escolas, podendo ser encontrados apenas nas

tabelas que apresentam as pontuações obtidas pelas escolas no Ideb de 2009 (OLIVEIRA, 2011).

Com a divulgação, pelo Governo Federal, de resultados nacionais da *Prova Brasil*, associada a iniciativas de governos estaduais nessa mesma direção – por exemplo, Minas Gerais, Ceará e Rio Grande do Sul –, passa-se a contar com experiências de avaliação da educação de segunda geração, caracterizadas por inovações que incorporam a divulgação de resultados de modo a permitir comparações não apenas entre redes, mas entre escolas.

A estratégia da mídia de divulgação, por meio de *rankings*, embora não oficial, juntamente com a distribuição nas escolas da matriz de conteúdos e habilidades utilizada na elaboração dos testes de língua portuguesa e matemática, introduz perspectivas concretas de interferência mais direta no *que* as escolas fazem e em *como* o fazem.

Em termos de responsabilização, no entanto, a *Prova Brasil* e o uso de seus resultados para composição do Ideb integram uma política de *responsabilização branda*, uma vez que se limitam a traçar metas e a divulgar os resultados dos alunos por escola e rede de ensino, sem atrelar prêmios ou sanções a esses resultados, como é característico das políticas de *responsabilização sólida* (HANUSHEK, 2004; HANUSHEK; RAYMOND, 2005). Tais políticas e suas relações com a terceira geração das avaliações educacionais serão tratadas na próxima seção

Terceira geração de avaliação da educação – responsabilização e currículo: as avaliações estaduais de São Paulo e Pernambuco

Vários sistemas estaduais e municipais de ensino básico vêm desenvolvendo propostas próprias de avaliação – usualmente caracterizadas pela avaliação censitária de suas escolas – por meio de aplicação de provas aos alunos de 4ª e 8ª séries do ensino fundamental e 3º ano do ensino médio, com frequência bianual. Cotejando-se as propostas em curso, nota-se grande similaridade nos delineamen-

tos adotados pelos sistemas de avaliação, os quais tendem a assumir, na elaboração dos itens das provas, a matriz de referência do Saeb e da *Prova Brasil*⁶.

No entanto, há especificidades nas avaliações educacionais e no uso de seus resultados que ilustram as características das relações entre avaliações de terceira geração, políticas de responsabilização e currículo escolar.

Evoluções recentes na avaliação estadual de São Paulo

O *Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo* (SARESP) foi implantado em 1996, apresentando-se com os seguintes objetivos:

- Subsidiar a Secretaria de Educação na tomada de decisão quanto à política educacional;
- Verificar o desempenho dos alunos da educação básica para fornecer informações a todas as instâncias do sistema de ensino que subsidiem a capacitação dos recursos humanos do magistério; a reorientação da proposta pedagógica das escolas, de modo a aprimorá-la; a viabilização da articulação dos resultados da avaliação com o planejamento escolar, capacitação e o estabelecimento de metas para o projeto de cada escola. (SÃO PAULO, 1996, p. 7)

Os objetivos explicitados indicam que a avaliação tinha dupla orientação: servir de referência para a elaboração de políticas, por parte da Secretaria de Educação, e orientar a construção da proposta pedagógica e a elaboração do planejamento pelas escolas. Associando a avaliação à melhoria da qualidade do ensino, o documento de implantação revela que tal qualidade é dependente, por um lado, do compromisso dos gestores do sistema de ensino e, por outro, das escolas, sendo estas particularmente responsabilizadas pelo desempenho dos alunos.

6 - Cf. LOPES, 2007; SOUSA; OLIVEIRA, 2007.

A noção de responsabilização, direcionada aos professores e demais profissionais da educação, concretizou-se no ano de 2000, com a instituição do *Bônus Mérito*, cuja distribuição levou em conta os resultados da avaliação em larga escala⁷.

Em 2007, quando a então Secretária de Educação Maria Helena Guimarães Castro e o Governador José Serra anunciaram o *Plano de Metas*, evidenciou-se a importância que a avaliação em larga escala assumiria para essa gestão. A 5ª meta estabelecida no *Plano* previa um aumento de 10% nos índices de desempenho dos ensinos fundamental e médio nas avaliações nacionais e estaduais. Ao estabelecer essa meta, a Secretária indica a continuidade do SARESP, e, dentre as *10 Metas para uma Escola Melhor*, duas enfatizavam o papel da avaliação em larga escala no desenvolvimento da política educacional paulista. As metas foram divulgadas com a seguinte redação:

Meta 8 - Sistemas de Avaliação:

- A avaliação externa das escolas estaduais (obrigatória) e municipais (por adesão) permitirá a comparação dos resultados do SARESP com as avaliações nacionais (SAEB e a Prova Brasil), e servirá como critério de acompanhamento das metas a serem atingidas pelas escolas.
- Participação de toda a rede na Prova Brasil (novembro de 2007).
- Capacitação dos professores para o uso dos resultados do SARESP no planejamento pedagógico das escolas em fevereiro de 2008.

7- O *Bônus Mérito* foi instituído no Governo de Mario Covas (1999-2001), pela Lei Complementar nº 891/00, e mantido durante as gestões de Geraldo Alckmin (2001-2002 e 2003-2006). Atualmente, na gestão de José Serra, a Secretaria de Estado da Educação instituiu novo *Bônus Mérito*, cujo cálculo baseia-se no *Índice de Desenvolvimento da Educação do Estado de São Paulo* (IDESP), o qual tem como um de seus critérios o desempenho dos alunos nas provas de língua portuguesa e matemática do SARESP. Essa vinculação fortalece a relação entre o pagamento do bônus e a avaliação em larga escala. Para saber mais sobre a bonificação por resultados, ver Lei Complementar nº 1.078/08 e Resoluções SEE nº 21/09; 22/09; 23/09 e 26/09. Para informações sobre o IDESP, consultar <<http://idesp.edunet.sp.gov.br>>.

- Divulgação dos resultados do SARESP 2007 para todas as escolas, professores, pais e alunos em março de 2008.

Meta 9 - Gestão de Resultados e Política de Incentivos:

- Implantação de incentivos à boa gestão escolar valorizando as equipes.
- O SARESP 2005 e as taxas de aprovação em 2006 serão a base das metas estabelecidas por escola.
- Também serão considerados indicadores como a assiduidade dos professores e a estabilidade das equipes nas escolas.
- Cada escola terá metas definidas a partir da sua realidade, e terá que melhorar em relação a ela mesma.
- As escolas com desempenho insuficiente terão apoio pedagógico intensivo e receberão incentivos especiais para melhorarem seu resultado.
- As equipes escolares que cumprirem as metas ganharão incentivos na remuneração dos profissionais.⁸

Essas metas ilustram a importância atribuída aos resultados das avaliações em larga escala nas gestões da Secretária Maria Helena Guimarães Castro (2007 a 2009) do Secretário Paulo Renato de Souza (abril de 2009 a 2010). As metas 8 e 9 permitem concluir que os objetivos indicados para o SARESP em 1996 permanecem até os dias atuais, evidenciando que a avaliação deve servir tanto para uso dos gestores dos sistemas, quanto na orientação do planejamento e do trabalho pedagógico nas escolas.

A política iniciada na gestão da Secretária Maria Helena e que vem tendo continuidade revela preocupação com a apropriação dos resultados pelos órgãos gestores do sistema e pelas escolas. Dentre as medidas tomadas em tal gestão, está a implantação de um currículo unificado que se apresenta como

8- Disponível em: <www.educacao.sp.gov.br>. Acesso em: 15 jan. 2010.

norteador da organização do ensino, pautando os parâmetros da avaliação. Essa unificação do currículo relaciona-se diretamente com as mudanças implementadas no SARESP a partir de 2007, sobretudo com a adoção da TRI.

A análise do currículo oficial e das matrizes do SARESP revela a correspondência entre o currículo, as matrizes e os materiais didáticos disponibilizados para professores (desde 2008) e para alunos (desde 2009), denominados cadernos do professor e do aluno. Esses materiais apresentam situações de aprendizagem que visam orientar e apoiar, a partir do currículo, o trabalho docente em sala de aula.

Evoluções recentes na avaliação estadual de Pernambuco

Um aspecto central da política educacional no Estado de Pernambuco foi a introdução, pelo Governador Eduardo Campos e pelo Secretário de Educação Danilo Cabral, de um sistema de responsabilização educacional. Esse sistema inclui a condução anual do *Sistema de Avaliação Educacional de Pernambuco* (Saepe), a ampla divulgação de seus resultados, avaliações bimestrais de estudantes por notas e o monitoramento bimestral de indicadores educacionais de cada escola da rede estadual por meio de um sistema informatizado.

O Saepe foi realizado pela primeira vez em 2000. Em 2005, ele foi novamente aplicado, mas seus resultados somente foram consolidados e divulgados em 2007. A partir de 2008, começou a ser realizado anualmente, e seus resultados passaram a compor o *Índice de Desenvolvimento da Educação Básica de Pernambuco* (Idepe).

O Sistema também coleta informações sobre as condições socioeconômicas e culturais dos estudantes, dos professores e da equipe gestora. Seus principais objetivos são:

- produzir informações sobre o grau de domínio dos estudantes nas habilidades e competências consideradas essenciais, em

cada período de escolaridade avaliado, não apenas para a continuidade dos estudos, mas para a vida em sociedade;

- monitorar o desempenho dos estudantes ao longo do tempo, como forma de avaliar continuamente o projeto pedagógico de cada escola, possibilitando a implementação de medidas corretivas quando necessário;
- contribuir diretamente para a adaptação das práticas de ensino às necessidades dos alunos, diagnosticadas por meio dos instrumentos de avaliação;
- associar os resultados da avaliação às políticas de incentivo com a intenção de reduzir as desigualdades e elevar o grau de eficácia da escola;
- compor, em conjunto com as taxas de aprovação verificadas pelo Censo Escolar, o Idepe.

A característica central dessa política é a formulação de metas para cada escola e a concessão do *Bônus de Desempenho Educacional* (BDE) às escolas que cumprirem suas metas. Dessa forma, as notas dos alunos nas provas de proficiência do Saepe são utilizadas, juntamente com o Idepe, para a definição das metas a serem alcançadas.

O Idepe considera tanto os resultados da avaliação do Saepe em língua portuguesa e matemática dos alunos de 4ª e 8ª séries do ensino fundamental e do 3º ano do ensino médio, quanto a média de aprovação dos alunos, medida pelo *Censo Escolar*. Assim, para elevar o Idepe, a escola deve, necessariamente, apresentar melhorias na média da proficiência dos alunos no Saepe e na média da taxa de aprovação.

As metas a serem alcançadas pelos alunos são pactuadas entre a Secretaria de Educação e a escola. Cada escola possui sua própria meta, calculada de acordo com suas particularidades. Além disso, as metas são compatíveis com o estágio da escola: para cada série avaliada, é estabelecida uma meta para língua portuguesa e uma meta para matemática. A diferença entre o Idepe utilizado como referência e o esperado é a meta para cada disciplina e cada série avaliada,

e a média dos resultados efetivamente alcançados demonstra o percentual obtido pela escola em relação às suas metas.

As metas de 2008 foram definidas de forma que as escolas alcançassem o ponto médio entre o Idepe inicial, em 2005, e a meta de 2009. Essas metas variaram entre os grupos de *baixo desempenho*, *desempenho intermediário* e *alto desempenho*. No entanto, a partir de 2009, as metas são únicas para todas as escolas dentro de um mesmo grupo.

O Saepe constitui uma avaliação de terceira geração em apoio a mecanismos de responsabilização forte cuja expressão mais consistente é o referido BDE. A bonificação varia de 50% a 100%: as escolas estaduais que obtiverem um índice global abaixo de 50% não recebem o bônus; a escola que alcançar 50% da meta estipulada recebe a metade do bônus; a partir daí, o valor é proporcional ao percentual da meta atingido.

A Secretaria de Educação também incentiva o reconhecimento dos professores que permanecem numa mesma escola. Para tanto, o cálculo do bônus considera a proporcionalidade no cumprimento da meta a partir de 50% e a lotação do professor na escola por, no mínimo, seis meses do ano letivo de referência para a concessão da remuneração. O BDE é coletivo, uma vez que todos os funcionários lotados e em exercício na escola têm direito ao bônus. Ele é também proporcional ao salário e ao percentual de cumprimento da meta.

Em Pernambuco, o valor máximo a ser recebido por cada servidor não está definido *a priori*. Enquanto em São Paulo esse valor chega, no máximo, a 2,4 salários, em Pernambuco, apenas o valor total alocado pelo governo para pagamento do bônus é fixo. Além disso, a condição para que os servidores de uma escola tenham acesso ao bônus é a existência de uma das séries testadas pelo Saepe (4ª e 8ª série do ensino fundamental e 3º ano do ensino médio). Desse modo, uma escola que atenda exclusivamente a modalidades de ensino não testadas não tem acesso ao bônus. Porém, profes-

sores que atuam em modalidades de ensino não testadas, como educação de jovens e adultos e educação infantil, têm acesso ao bônus caso a escola possua alguma série testada. Para as escolas que não alcançarem as metas, a legislação prevê reforço técnico, pedagógico e estrutural, com o objetivo de elas reenquadrarem-se nos critérios do BDE no ano letivo seguinte.

Além do resultado das escolas, o *site* da avaliação apresenta as matrizes utilizadas na elaboração dos itens dos testes, com explicações detalhadas sobre os descritores, os conhecimentos e as competências esperadas para cada série dos ensinos fundamental e médio, além de boletins de resultados contendo a análise contextual com informações nacionais e por município e Estado.

O Saepe constitui-se, assim, como uma avaliação capaz de pautar o currículo e o que os alunos devem aprender em cada fase do ciclo escolar.

Avaliação em larga escala e currículo escolar: o que dizem as pesquisas?

Avaliações de segunda e terceira geração, associadas à introdução de políticas de responsabilização baseadas em consequências simbólicas e materiais, têm o propósito de criar incentivos para que o professor se esforce no aprendizado dos alunos. No entanto, evidências nacionais e internacionais mostram que principalmente o uso de resultados das avaliações de terceira geração para informar iniciativas de responsabilização forte pode envolver riscos para o currículo escolar. Um deles é a situação conhecida como *ensinar para o teste*, que ocorre quando os professores concentram seus esforços preferencialmente nos tópicos que são avaliados e desconsideram aspectos importantes do currículo, inclusive de caráter não cognitivo.

É difícil discordar da alegação de que as avaliações em larga escala lidam com uma visão estreita de currículo escolar diante do que as es-

colas se propõem como objetivos para a formação de seus estudantes. Também é complexo o uso de testes padronizados para aferir objetivos escolares relacionados a aspectos não cognitivos.

O problema decorre do fato de os currículos escolares possuírem múltiplos objetivos, ao passo que as medidas de resultados utilizadas pelas avaliações em larga escala tipicamente visam a objetivos cognitivos relacionados à leitura e à matemática. Essa não é exatamente uma limitação das avaliações, mas demanda atenção para riscos relativos ao estreitamento do currículo, os quais podem acontecer quando há uma interpretação distorcida do significado pedagógico dos resultados da avaliação.

Os estudos sobre o tema no Brasil são ainda limitados e bastante recentes. Mesmo assim, eles colaboram para o entendimento sobre como escolas e secretarias de educação interpretam e articulam as relações entre as três gerações da avaliação em larga escala e o currículo escolar.

Os resultados da pesquisa realizada por João Luiz Horta Neto (2006) sobre o uso dos dados do Saeb 2003 pela Secretaria de Educação do Distrito Federal para orientar o planejamento da rede de ensino, por exemplo, ajudam a ilustrar a escassa interferência de avaliações de primeira geração no contexto da gestão educacional. A pesquisa indicou que, apesar de os gestores da Secretaria de Educação defenderem a importância do Saeb, eles têm pouco conhecimento e praticamente não utilizam os dados gerados pela avaliação nos processos de gestão, principalmente devido a limitações para compreender os resultados produzidos. Essa pesquisa, ao mesmo tempo em que demonstra a baixa capacidade do Saeb para impactar a gestão educacional e as atividades escolares, contribui para a compreensão da emergência de uma segunda geração de avaliação educacional que, como a *Prova Brasil*, permite que as escolas enxerguem-se nos resultados produzidos.

Ana Paula Oliveira (2011) pesquisou em que medida os resultados da *Prova Brasil* 2007 têm servido de subsídio para a gestão da rede de

ensino pela Secretaria de Educação do Distrito Federal. Tal questão também foi investigada em duas escolas do Distrito Federal, as quais apresentavam o maior e o menor Ideb. Por um lado, os gestores educacionais mostraram conhecer a *Prova Brasil*, inclusive do ponto de vista técnico, além de percebê-la como um instrumento que estabelece os parâmetros de qualidade de ensino que devem ser alcançados pelas escolas. Nessa perspectiva, os gestores declaram que os melhores resultados obtidos por algumas escolas nas avaliações são utilizados por outras escolas da rede como indicativo da qualidade do trabalho que estão desenvolvendo. Por outro lado, as equipes da Secretaria de Educação percebem a *Prova Brasil* como uma iniciativa que possibilita unificar o processo de ensino-aprendizagem, no sentido de dar a conhecer o que está sendo ensinado e aprendido em todas as escolas do país.

Complementarmente, os professores das duas escolas investigadas afirmam que a *Prova Brasil* contribui para a unificação do ensino em suas instituições. Isso é tido como um fato positivo, uma vez que a unificação curricular poderia vir a contribuir para que estudantes de todo o país tivessem acesso aos mesmos conhecimentos, independentemente da localidade em que vivem e da escola que frequentam. Dessa forma, professores e gestores percebem a *Prova Brasil* como uma referência para a possível implantação de um currículo comum nacional.

Coordenadores e professores das escolas pesquisadas por Oliveira (2011) também declararam buscar redefinir o conteúdo programático de modo a atender ao que é avaliado pela *Prova Brasil*. Como a avaliação é geralmente aplicada antes do término do ano letivo, as escolas antecipam os conteúdos para que os alunos consigam responder aos testes, a fim de garantir uma boa média de desempenho para a escola. Ainda na perspectiva do que a literatura denomina *ensinar para o teste*, os professores afirmam ter incorporado a prática de preparar os alunos para se habituarem aos textos, aos comandos e à extensão dos testes de leitura da *Prova Brasil*.

Pesquisas que analisaram a implementação da avaliação no Estado de São Paulo destacam elementos que ilustram as transformações que vêm ocorrendo no currículo⁹. Os estudos feitos há mais tempo tendem a indicar poucos efeitos do SARESP no cotidiano escolar, embora destaquem reações de desconfiança e de resistência ao Sistema por parte dos profissionais da educação (OLIVEIRA, 1998; ESTEVES, 1998; FELIPE, 1999; KAWAUCHI, 2001). Esses estudos, que tiveram como foco eventuais reflexos do SARESP na escola, considerando, em especial, opiniões e reações de professores, tendem a não identificar influências de seus resultados no currículo escolar.

É a partir dos estudos realizados na segunda metade da década de 2000 que são identificadas evidências de efeitos do SARESP no contexto escolar. Lilian Rose Freire (2008) realça, nas conclusões de sua pesquisa, realizada em uma escola da rede estadual, alguns usos dos resultados do SARESP, tais como: i) utilização na composição das notas bimestrais dos alunos; ii) reprodução de questões na prova unificada criada pela escola, com a intenção de *treinar* os alunos para a avaliação; iii) utilização, pelos professores de português, das orientações do SARESP relativas à correção de redações, para orientar os alunos nas redações escolares, o que pode representar aprimoramento de práticas vigentes; iv) incentivo à participação dos alunos no dia da aplicação das provas do SARESP, por meio de atribuição de nota a ser considerada na média bimestral.

As informações apresentadas por essa pesquisa indicam que o significado assumido pelo SARESP e por seus resultados no contexto da escola pesquisada não se associa à ideia de uma avaliação que traga subsídios para a orientação e o replanejamento do trabalho escolar. Com exceção da iniciativa no uso de critérios para correção de redações, a interação com o SARESP, ao que parece, é mais instrumental, no sentido de implantar iniciativas que possam ajudar os alunos a

obterem melhores resultados, como ensinar a preencher gabaritos e aplicar provas com questões semelhantes às provas do Sistema.

Uma pesquisa realizada por Paulo Henrique Arcas (2009) focalizou eventuais repercussões do SARESP na avaliação escolar, procurando apreender características e tendências da avaliação após sua implantação. Para tanto, realizou-se um estudo em uma diretoria regional de ensino na região metropolitana de São Paulo, buscando-se a opinião de professores-coordenadores. Por meio de questionário e entrevista, o objetivo foi analisar como viam o SARESP e como foram construindo suas opiniões acerca dele no período de sua vigência – no caso da referida pesquisa, até 2007. Essas opiniões possibilitaram identificar como a avaliação em larga escala vem incidindo na avaliação e no currículo escolar.

Ficou evidente, na fala dos professores-coordenadores entrevistados, que há uma tendência de aceitação do SARESP, embora, inicialmente, o Sistema tenha sido visto por eles com desconfiança. Há evidências de que os dados obtidos na avaliação são analisados e discutidos no planejamento escolar, no início do ano, e também no replanejamento, no início do segundo semestre letivo. Os resultados da escola e das turmas são analisados e orientam o trabalho escolar, definindo habilidades, competências e conteúdos a serem ensinados.

Assim, pode-se afirmar que o SARESP vem-se fazendo presente gradualmente, influenciando práticas, definindo metas, estabelecendo rumos, orientando o trabalho pedagógico. Outra revelação importante sobre as implicações de tal avaliação no contexto escolar é que ela tem incidido sobre as práticas avaliativas desenvolvidas na escola. As evidências da referida pesquisa demonstraram que a avaliação da aprendizagem realizada na escola toma a avaliação em larga escala como referência.

Entretanto, o que pôde ser constatado é que o SARESP, ao servir como referência para as práticas avaliativas empreendidas

9 - As considerações relativas ao SARESP foram extraídas de SOUSA; ARCAS, 2010.

nas escolas, acabou por reforçar práticas tradicionais de avaliação da aprendizagem. [...] o SARESP reforça a aplicação de provas testes, objetivando, na maioria dos casos, simular a aplicação da avaliação externa, supondo-se, desse modo, estar preparando os alunos. (ARCAS, 2009, p.120)

A centralidade que o SARESP está adquirindo na organização do trabalho escolar, ao nortear práticas avaliativas, permite afirmar que a avaliação em larga escala vem sendo crescentemente apropriada pelas escolas. Nesse sentido, ao orientar os procedimentos avaliativos, o SARESP vem induzindo a uma ênfase na aplicação de provas e exames simulados como meios de preparar os alunos para se saírem bem na avaliação estadual.

Conclusões

O presente artigo buscou caracterizar desenhos, objetivos e usos dos resultados de experiências de avaliação da educação básica em curso no país, tendo em vista suas relações com o currículo escolar. O estudo de três gerações de avaliação da educação básica permitiu identificar as avaliações de segunda e terceira geração – isto é, que se articulam, respectivamente, a políticas de responsabilização fraca e forte – como aquelas com consequências mais expressivas para o currículo escolar.

Embora seja relativamente cedo para fazer afirmações mais consistentes sobre essas avaliações e suas repercussões no currículo escolar, os primeiros estudos mostram que o novo desenho introduzido pela segunda geração produz resultados que servem como indicadores capazes de fornecer informações a respeito dos componentes do currículo que estão chegando aos alunos e daqueles que não estão. Nessa perspectiva, tal tipo de avaliação parece estar reforçando o alinhamento, nas escolas e secretarias de educação, entre o currículo ensinado e o currículo avaliado.

De fato, a breve revisão de pesquisas sobre o tema apresentada aqui aportou contribui-

ções, por vezes recorrentes, para a compreensão das interferências da avaliação sobre o currículo escolar. O que esses trabalhos evidenciam, em conjunto, é a importância que vêm assumindo as avaliações de segunda e terceira geração no delineamento das políticas educacionais e, em consequência, seu potencial de direcionar o *que, como e para que* ensinar.

De acordo com os resultados de pesquisa, é o uso de provas padronizadas no contexto de avaliações referentes a políticas de responsabilização com consequências fracas e fortes para as escolas – principalmente as fortes – que exacerbaria a preocupação de diretores e professores em preparar os alunos para os testes e para o tipo de atividade neles presente.

Além disso, as pesquisas também mostram que a primeira geração de avaliação em larga escala, ou seja, a avaliação *sem consequências*, minimiza esses problemas, porque os diretores e professores veem-se menos *ameaçados* pela avaliação e podem assumi-la, ou não, com maior liberdade. Nesse contexto, em contrapartida, tais profissionais raramente se sentem obrigados a prestar conta dos resultados de seu trabalho ou têm motivação para inteirar-se dos resultados das avaliações e para levá-los em consideração em sua atuação educacional e pedagógica.

Em síntese, este estudo discutiu os riscos e potenciais das avaliações de segunda e terceira geração para o currículo escolar. Apontou, por um lado, os riscos de as avaliações relativas a políticas de responsabilização exacerbarem a preocupação de diretores e professores em preparar seus alunos para os testes, levando a um estreitamento do currículo escolar. Indicou, ainda, as implicações para a avaliação da aprendizagem quando as escolas passam a organizá-la tomando como referência o tipo de teste utilizado pela avaliação em larga escala.

Por outro lado, o presente estudo indicou o potencial das avaliações de segunda e terceira geração para propiciar uma discussão mais informada sobre o currículo escolar, em termos das habilidades fundamentais de leitura e matemática que ainda não têm sido

garantidas a todos os alunos de ensino fundamental e médio.

Diante disso, o desafio parece ser a compatibilização dos objetivos, desenhos e usos dos resultados das três gerações de avaliação em larga escala a fim de propiciar uma discussão informada sobre os aspectos específicos de língua portuguesa e matemática que pre-

cisariam ser aprendidos por todos os alunos, bem como uma definição mais clara do que esses alunos deveriam ter aprendido ao final de cada ciclo nessas duas áreas do saber escolar. É aí que se insere a necessidade de aumentar o acervo de pesquisas que contribuam para a compreensão dos impactos das novas gerações da avaliação educacional no currículo escolar.

Referências

ARCAS, Paulo Henrique. **Implicações da progressão continuada e do SARESP na avaliação escolar: tensões, dilemas e tendências.** Tese (Doutorado) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2009.

BONAMINO, Alicia. **Tempos de avaliação educacional: o SAEB, seus agentes, referências e tendências.** Rio de Janeiro: Quartet, 2002.

BROOKE, Nigel. O futuro das políticas de responsabilização no Brasil. **Cadernos de Pesquisa**, São Paulo, v. 36, n. 128, p. 377-401, 2006.

CARNOY, Martin; LOEB, Susanna. Does external accountability affect student outcomes? A cross-state analysis. **Educational Evaluation and Policy Analysis**, v. 24, n. 4, p. 305-331, 2002.

CENEVIVA, Ricardo. **Democracia, accountability e avaliação: a avaliação de políticas públicas como mecanismo de controle democrático.** Dissertação (Mestrado) – Fundação Getúlio Vargas, São Paulo, 2005.

ESTEVES, Maria Eunice de P. **Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP: uma ação planejada.** Dissertação (Mestrado) – Pontifícia Universidade Católica de São Paulo, São Paulo, 1998.

FELIPE, Jesse Pereira. **Uma análise crítica do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo: SARESP.** Dissertação (Mestrado) – Pontifícia Universidade Católica de São Paulo, São Paulo, 1999.

FERNANDES, Reynaldo. **Índice de Desenvolvimento da Educação Básica (Ideb).** Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2007.

FERNANDES, Reynaldo; GREMAUD, Amaury. Qualidade da educação básica: avaliação, indicadores e metas. In: VELOSO, Fernando et al. (Orgs.). **Educação básica no Brasil: construindo o país do futuro.** Rio de Janeiro: Elseiver, 2009.

FREIRE, Lilian Rose S. Carvalho. **SARESP 2005: as vicissitudes da avaliação em uma escola da rede estadual.** Dissertação (Mestrado) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2008.

FREITAS, Dirce Nei Teixeira de. **A avaliação da educação básica no Brasil: dimensão normativa, pedagógica e educativa.** Campinas: Autores Associados, 2007.

GATTI, Bernadete A. Testes e avaliações do ensino no Brasil. **Educação e Seleção**, São Paulo, n. 16, p. 12-21, 1987.

HANUSHEK, Eric A. United States lessons about school accountability. **CESifo DICE Report 4**, 2004.

HANUSHEK, Eric A.; RAYMOND, Margaret E. Does school accountability lead to improved student performance? **Journal of Policy Analysis and Management**, v. 24, n. 2, p. 297-327, 2005.

HORTA NETO, João Luiz. **Avaliação externa**: a utilização dos resultados do SAEB 2003 na gestão do sistema público de Ensino Fundamental no Distrito Federal. Dissertação (Mestrado) – Faculdade de Educação, Universidade de Brasília, Brasília, DF, 2006.

KAWAUCHI, Mary. **SARESP e ensino de história**: algumas questões. Dissertação (Mestrado) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2001.

LOPES, Valéria Virgínia. **Cartografia da avaliação educacional no Brasil**. Tese (Doutorado) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2007.

OLIVEIRA, Ana Paula M. **A Prova Brasil como política de regulação da rede pública do Distrito Federal**. Dissertação (Mestrado) – Faculdade de Educação, Universidade de Brasília, Brasília, DF, 2011.

OLIVEIRA, Duzolina Alfredo F. de. **Uma avaliação política do SARESP**. Dissertação (Mestrado) – Universidade Estadual de Campinas, Campinas, 1998.

SÃO PAULO. Secretaria de Estado de Educação. **SARESP**: documento de implantação. São Paulo: FDE, 1996.

SOUSA, Sandra Zákia; ARCAS, Paulo Henrique. Implicações da avaliação em larga escala no currículo: revelações de escolas estaduais de São Paulo. **Educação: Teoria e Prática**, Rio Claro, v. 20, n. 35, p. 181-199, jul./dez. 2010.

SOUSA, Sandra Zákia; OLIVEIRA, Romualdo Portela de. **Sistemas de avaliação educacional no Brasil**: características, tendências e uso dos resultados. Relatório de pesquisa apresentado à FAPESP, São Paulo, jul. 2007.

ZAPONI, Margareth; VALENÇA, Epifânia. **Política de responsabilização educacional**: a experiência de Pernambuco. Abr. 2009. Disponível em: <www.abave.org.br>.

Recebido em: 08.08.2011

Aprovado em: 20.09.2011

Alicia Bonamino é professora associada da Pontifícia Universidade Católica do Rio de Janeiro e coordenadora do Laboratório de Avaliação em Educação (LAEd).

Sandra Zákia Sousa é doutora em Educação, professora da Faculdade de Educação da Universidade de São Paulo, atuando na pós-graduação em Estado, Sociedade e Educação, e do mestrado em Educação da Universidade Cidade de São Paulo. Possui publicações predominantemente no campo da avaliação educacional. Email: sanzakia@usp.br