

Avaliações em larga escala: uma sistematização do debate

Adriana Bauer¹

Ocimar Munhoz Alavarse¹

Romualdo Portela de Oliveira¹

Resumo

As reformas educativas implantadas nas últimas décadas caracterizam-se, entre outros traços, pela utilização de avaliações em larga escala como instrumento de gestão de redes de ensino e de responsabilização de profissionais da educação. Isso tem ocorrido de modo peculiar em cada país e, no Brasil, observa-se a difusão dessas avaliações, por meio de provas padronizadas. Tais avaliações geram na comunidade acadêmica e educacional um debate no qual há posições que vão da contraposição extremada ao reconhecimento da contribuição das avaliações em larga escala e das medidas educacionais como norteadoras de políticas e programas educacionais, incluindo as restrições ao uso que vem sendo feito de seus resultados. Diante dos argumentos favoráveis e contrários, incluindo as dimensões técnicas e políticas, neste trabalho, apresenta-se uma reflexão que reconhece a utilidade dessas avaliações, ainda que questionando alguns de seus usos para a gestão como critério, por exemplo, para a alocação de recursos nas escolas com melhores resultados, a definição de bônus para professores, o estabelecimento de *rankings* estimulando a competição entre escolas e redes de ensino, e seu entendimento como indicador único e principal de qualidade de ensino. Assim, o objetivo deste artigo é discutir as principais críticas e ponderações às avaliações em larga escala presentes na literatura nacional e internacional e sistematizar o debate em torno de suas potencialidades. Para tanto, dois aspectos recorrentes na literatura foram destacados, o papel e a validade das avaliações em larga escala nas reformas educacionais e o uso de seus resultados para a gestão de sistemas e escolas.

Palavras-chave

Avaliação em larga escala – Avaliação externa – Política educacional.

1- Universidade de São Paulo, São Paulo, SP, Brasil.
Contatos: dri.bauer@usp.br, ocimar@usp.br e romualdo@usp.br.

Large-scale assessment: debate systematization

Adriana Bauer¹
Ocimar Munhoz Alavarse¹
Romualdo Portela de Oliveira¹

Abstract

Educational reforms in recent decades have been characterized, among other features, by the use of large-scale assessments as a management tool of school systems and of educator accountability. This phenomenon has occurred differently in each country. In Brazil there has been widespread use such assessments with standardized tests, which generates in the academic and educational community not consensus but a debate in which there are different positions, ranging from the extreme opposition to the recognition of the contribution of large-scale assessments and educational measures to guide educational policies and programs, including restrictions on the use made of their results. Considering positive and negative arguments, including technical and political dimensions, this article, on the one hand, presents a reflection that recognizes the usefulness of these evaluations, and, on the other, questions some of its uses for educational management, for example, allocating resources in schools with best results, setting bonuses for teachers, establishing rankings which stimulate competition between schools and school systems, and understanding test results as a unique and primary indicator of education quality. The objective of this article is to discuss the main pros and cons to large-scale assessments in the national and international literature and systematize the debate on their potential. For this, it focuses on two recurrent themes in literature: the role and validity of large-scale assessments in the education reforms and the use of their results for the management of educational systems and schools.

Keywords

Large-scale assessment – External evaluation – Educational policy

1- Universidade de São Paulo, São Paulo, SP, Brasil.
Contact: dri.bauer@usp.br, ocimar@usp.br e romualdo@usp.br.

Introdução

As reformas educativas implantadas nas últimas décadas caracterizam-se por um conjunto de medidas que articulam os seguintes aspectos:

a) centralização dos sistemas de avaliação, que passam a ser utilizados como instrumentos de gestão e alimentam políticas de responsabilização aliadas a desenhos censitários de avaliação externa;

b) descentralização dos processos de gestão e financiamento, que fortalecem o discurso da autonomia e da gestão democrática da escola, numa perspectiva de melhoria dos resultados, o que inclui a autonomia financeira para buscar novas fontes de recursos, que não as fontes públicas tradicionais, e novas formas de gerenciamento da educação pública, o que inclui autonomia de gestão financeira e autonomia de gestão (*school based management*);

c) ampliação das possibilidades de escolha (*choice*), estimulando mecanismos de competição entre as escolas, o que induziria à melhoria de sua qualidade; e

d) valorização dos resultados e busca de maior efetividade do serviço ofertado (*school effectiveness*) (Cf. BONAMINO, 2013; LEVIN, 2001; OLIVEIRA, 1999, 2000).

Ainda que estas características tenham se difundido de modo peculiar em cada país, e que nem sempre apareçam ao mesmo tempo em todas as reformas implantadas, no caso brasileiro é bastante evidente a difusão e ampliação de avaliações que têm como traço comum a utilização de provas padronizadas, aplicadas em larga escala, configurando nitidamente um processo de avaliações externas, com crescente importância no desenho das políticas educacionais de todos os entes federados.

A aceitação de tais avaliações, entre nós, longe está de ser consensual. Ao contrário, ainda encontramos parcelas significativas da comunidade acadêmica e educacional que, de maneira geral, desconsideram as eventuais contribuições que podem se originar das medidas educacionais em larga escala.

Com algumas exceções que tiveram maior repercussão, como os livros de Diane Ravitch (2010) e o de Almerindo Afonso (2000), a contribuição internacional ao debate a respeito de avaliações em larga escala é praticamente ausente no debate brasileiro acerca do tema, com predominância da difusão de posições críticas a estas.

Por outro lado, há trabalhos de autores que não negam a contribuição das avaliações em larga escala e das medidas educacionais como norteadoras de políticas e programas educacionais, mas que se contrapõem ao modo como estas têm se difundido no território nacional, bem como ao uso que vem sendo feito de seus resultados, como encontramos, entre outros autores, em Freitas (2013); Sousa, Oliveira (2010).

Os autores deste trabalho reconhecem a utilidade dessas avaliações, ainda que questionem alguns de seus usos para a gestão, por exemplo, como critério para alocação de recursos nas escolas com melhores resultados, a definição de bônus para professores, o estabelecimento de *rankings* estimulando a competição entre escolas e redes de ensino entre outros, e seu entendimento como indicador único e principal de qualidade de ensino (Cf. OLIVEIRA, 2013; BAUER, 2010; 2012; 2013; MACHADO; ALAVARSE, 2014).

Assim sendo, o objetivo deste artigo é discutir as principais críticas às avaliações em larga escala presentes na literatura nacional e internacional e sistematizar o debate em torno dos possíveis usos de seus resultados. Para tanto, organizamos o texto a partir de dois aspectos recorrentes na literatura: 1. O papel e a validade das avaliações em larga escala nas reformas educacionais, em que discutiremos, também, a fundamentação e conceituação dos testes em larga escala; e 2. O uso dos resultados das avaliações em larga escala, para a gestão do sistema escolar e das escolas em particular, abrangendo o que os sistemas de ensino e as escolas realizam com os resultados de suas avaliações.

O papel das avaliações em larga escala nas reformas educacionais

O papel de destaque da avaliação padronizada nas políticas públicas educacionais, geralmente, aparece justificado pela necessidade de mudança nas concepções de gestão na educação *pari passu* à mudança nas organizações em geral. Ao se passar do controle burocrático e hierárquico do modo de execução para o controle do produto, as testagens revelaram-se instrumentos bastante adequados na educação. Além de possibilitar uma redução do pessoal total envolvido com o controle dos processos educativos, essa mudança permitiu o que Lício Lima (2011) chamou de “administração por controle remoto” dos sistemas de ensino. De fato, quem controla a avaliação, à medida que essa ganha importância na definição da alocação de recursos, dos salários etc., torna-se mais poderoso. Assim sendo, por meio das avaliações, busca-se o poder de induzir o(s) outro(s) a fazer o que se deseja. Ademais, possibilita um discurso progressista, agora reconceituado, de possibilitar a autonomia da escola, a descentralização da gestão, a participação etc. Não é coincidência que algumas reformas ancoradas no fortalecimento dos mecanismos de testagens em larga escala defendam medidas democratizadoras da gestão, como a eleição de diretores etc.

Configura-se assim um novo discurso acerca da gestão da escola pública, posto que as testagens propiciam melhor gerenciamento dos recursos disponíveis (a partir da possibilidade de auxiliar no direcionamento dos investimentos na escola pública), bem como pela demanda por dados educacionais e sobre as aprendizagens a serem utilizados para a melhoria do ensino público (Cf. AFONSO, 1998; BARRERA, 2000; CASTRO, 2009a; KLEIN; FONTANIVE, 1995; VIANNA, 2003).

Assim, após a virtual universalização do acesso ao ensino fundamental, as testagens seriam a chave para se verificar a qualidade do ensino, entendida apenas como elevação

dos níveis de aprendizagem dos alunos (Cf. CASASSUS, 2013; CASTRO, 2009; FERRER; ARREGUI, 2003; MONS, 2009). Tal controle tem sido exercido por meio dos resultados obtidos pelos estudantes em provas padronizadas e comparáveis ao longo do tempo, considerados evidência empírica dessa qualidade (Cf. CASASSUS, 2013). Dessa forma, as avaliações propiciariam informação, diagnóstico, regulação, monitoramento e controle (tanto do indivíduo quanto do sistema educacional) e legitimação das políticas (Cf. AFONSO, 2000; BONAMINO, 2001; BROADFOOT, 1996; CASTRO, 2009a; HOUSE, 1998; KELLAGHAN; GREANEY, 1992).

Outros argumentos favoráveis à adoção de avaliações padronizadas em larga escala nas políticas educacionais realçam, ainda que potencialmente, que essas avaliações:

1. Responsabilizam professores e escolas pelos resultados obtidos, sendo essa responsabilização considerada um dos maiores benefícios das políticas avaliativas. A prestação de contas do trabalho realizado nas unidades escolares e a responsabilização de gestores e professores pelos resultados, fariam com que esses ficassem mais comprometidos em melhorar sua prática e garantir o aprendizado dos alunos, de acordo com os padrões estabelecidos, ao mesmo tempo em que poderiam relativizar a responsabilidade dos governantes (Cf. EVERS; WALBERG, 2002; MADAUS; RUSSELL; HIGGINS, 2009; MONS, 2009);

2. Instauram uma cultura de avaliação dos serviços públicos e de transparência sobre seus processos e resultados;

3. Com a publicização dos resultados obtidos pelos estudantes nessas avaliações, permite-se que os pais acompanhem como os alunos de uma determinada escola estão se desenvolvendo, possibilitando que façam escolhas sobre os estabelecimentos de ensino de preferência para que os filhos estudem, ao mesmo tempo em que este processo pressionaria as escolas a melhorar;

4. Produzem diversas comparações entre alunos de uma mesma escola e entre alunos de diferentes escolas da região ou, até mesmo, do país, que não seriam possíveis se elas não existissem. Da mesma forma, em nível de rede de ensino, um determinado gestor poderá comparar o desempenho das escolas de sua rede com as de outros municípios, estados e com as médias nacionais. Tais comparações são importantes em contextos em que há uma matriz de referência ou um currículo básico comum;

5. São, por natureza, mais neutras e objetivas e, normalmente, corrigidas com a utilização de programas computacionais, possibilitando manter o anonimato dos estudantes e minimizando a subjetividade inerente às correções realizadas pelos professores em sala de aula. Além disso, o processo de elaboração de itens e instrumentos, usualmente, envolve especialistas para a eliminação de imperfeições e vieses de resposta e para garantir sua validade – medir o que se deseja medir. Nesse sentido, a busca pelo aprimoramento da dimensão técnica das avaliações, principalmente da validade dos testes e das técnicas de análise dos resultados, conferir-lhes-ia, segundo alguns autores, como, por exemplo, Fernandes; Gremaud (2008), maior precisão e credibilidade dos resultados. Essa discussão técnica foi bastante enfatizada na literatura dedicada ao tema no início da expansão e consolidação de diversas avaliações;

6. Garantem a constituição de bases de dados objetivos e de um sistema de informações que possibilitam acompanhar a evolução da educação e favorecem a tomada de decisões no âmbito educacional das políticas públicas com maior consistência. Além disso, o posterior estudo e análise desses dados, em relação a fatores diversos, tanto referentes a características dos sujeitos (etnia, nível socioeconômico, sexo, *background* cultural etc.), quanto com relação às características das escolas, do professorado, das redes de ensino, podem subsidiar o desenvolvimento de programas educacionais específicos, com vistas à melhoria dos resultados obtidos (Cf. CASTRO, 2009b);

7. Podem responsabilizar os próprios estudantes por sua aprendizagem, desafiando-os constantemente a melhorar seus resultados e possibilitando que se criem incentivos para melhorar as aprendizagens (Cf. EVERS; WALBERG, 2002);

8. Impulsionam mudanças em currículos inadequados (Cf. MADAUS; RUSSELL; HIGGINS, 2009); e

9. Subsidiem programas de melhoria referentes à idade de ingresso no ensino superior, com vistas a obter-se maior equidade nos resultados (Cf. EVERS; WALBERG, 2002).

Entretanto, há os que se posicionam criticamente a respeito do papel político das avaliações externas. Para alguns, como Arelaro (2003) e Esteban (2012), tais avaliações vêm responder a pressões para mudanças nos modos de administração e controle das redes de ensino, estando aliadas a um novo modelo de gestão educacional (nova gestão pública). Tais pressões, oriundas de organismos multilaterais, seriam no sentido de impor uma agenda educacional transnacional.

Há também os que criticam as avaliações estandardizadas devido ao papel que têm assumido no direcionamento de políticas de responsabilização e prestação de contas (*accountability*), principalmente quando passam a compor políticas de alto impacto (*high stakes tests*), como, por exemplo, a atribuição de bônus e/ou premiação a professores e escolas¹ (Cf. FERRER, 1996; FREITAS, 2013).

Assim, como as avaliações padronizadas em larga escala possuem um papel político que não pode ser desconsiderado, ganha importância a análise tanto de suas finalidades e propósitos quanto dos efeitos que essas podem ter nos sistemas educativos em que são aplicadas (Cf. BROADFOOT, 1996; COSTA, 1998; OZGA, 2000). Como ilustra Patrícia Broadfoot (1996, p. 12-13):

1- Popham (1987) utilizou o termo *high stakes* para referir-se a testes cujos resultados tinham consequências severas sobre os alunos. Complementarmente, o termo *low stakes* é utilizado para caracterizar testes padronizados que não têm consequências sobre a população avaliada, direta (alunos) ou indiretamente (professores, gestores etc.).

A maior parte da literatura sobre avaliação de desempenho se preocupa com as técnicas de avaliação. Enquanto muitos estudos se preocupam com o desenvolvimento e refinamento de diferentes abordagens de medida educacional, muitos outros oferecem poderosas críticas às técnicas de avaliação de desempenho e falhas dessas mesmas técnicas quando aplicadas. [...] Então, poderosos e intensos debates sobre técnicas concorrentes, os quais têm tradicionalmente caracterizado a pesquisa sobre avaliação educacional auxiliaram a excluir da discussão à questão mais fundamental relacionada aos propósitos e efeitos dessa atividade.

A preocupação com os propósitos e efeitos das avaliações, ou com o debate educacional mais amplo, tem sido alvo de diversos estudos e reflexões, mesmo porque os propósitos condicionam as opções de técnica a ser utilizada e, conseqüentemente, condicionam a interpretação e o uso que será feito de seus resultados.

Para avançarmos nessas análises e dado o lugar privilegiado que as avaliações em larga escala têm nos Estados Unidos da América (EUA), incluindo seu poder indutor em todo o mundo, pode-se destacar o livro *More than a score: the new uprising against high-stakes testing*, editado por Jesse Hagopian (2014), sintomaticamente, apresentado por Diane Ravitch, introduzido por Alfie Kohn e posfaciado por Wayne Au, autores estadunidenses que têm se notabilizado por apresentarem um conjunto de objeções aos testes de alto impacto, quadro que seria agravado pelo alcance dos mesmos pelo fato de serem avaliações externas. No livro, que reúne um conjunto de aportes que retratam o movimento de resistência aos testes de alto impacto nos EUA, Hagopian (2014, p. 7-27), em seu prefácio, destaca que a disseminação dos testes se fez em nome de reformas da educação pública, tendo à sua frente grandes interesses econômicos, quer do ponto de vista da direção da nação, quer do ponto de vista das empresas

que aplicam os testes, associadas, aliás, à venda de materiais didáticos.

Hagopian (2014) salienta que, acopladas aos programas *No child left behind*, da administração Bush, e *Race to the top*, no governo Obama, as avaliações em larga escala constituíram-se em um instrumento para impor um modelo de reforma educacional que, em seus resultados, aprofundam as desigualdades educacionais entre os alunos com desvantagens e entre aqueles provenientes de comunidades pauperizadas, como é o caso de alunos negros. Para o autor, não bastando essas conseqüências, que desmascaram as propaladas intenções dos reformadores de aumentar a equidade, os testes trouxeram pressões sobre professores e gestores de escolas, atualizando o que se convencionou denominar de *accountability*.

Hagopian (2014), ao recuperar o fio condutor da obra, sublinha a resistência aos testes de alto impacto que começa a ganhar corpo não só entre professores, mas, especialmente, envolvendo alunos, pais e gestores de escolas e de redes. Esse movimento, sem negar a importância da avaliação para o trabalho pedagógico, tem, entre outros aspectos, procurado evidenciar as limitações que tais testes possuem intrinsecamente, pelo modo como são construídos e validados, e, sobretudo, pelo tipo de orientação e legitimação de política educacional que ignora as principais variáveis do processo educacional e que se voltam contra os professores, considerados nessas políticas como os únicos responsáveis pelos resultados escolares.

Embora produzindo mais de uma década antes, ainda que sem um posicionamento cabalmente contrário aos testes de alto impacto, Stecher (2002), ao estudar a experiência dos EUA e estabelecer um tipo de estado da arte dos efeitos dos testes, considera que os mesmos podem ter efeitos positivos, negativos e até mesmo nulos ou ambíguos. O autor diferencia os impactos em quatro níveis – alunos, professores, gestores de escolas e gestores de redes –, demarcando que esses últimos são os que, em última instância, têm a possibilidade de adotar – ou não – os testes como um elemento de suas políticas educacionais.

Stecher (2002, p. 99-100), contudo, indica a necessidade de aprofundamentos investigativos a respeito da disseminação desses testes, pois:

O efeito líquido de testes de alto impacto sobre a política e a prática é incerto. Os pesquisadores não têm documentado as consequências desejáveis – de mais instrução, trabalho com mais afinco e trabalho mais eficaz –, tão claramente como as indesejáveis – alinhamento curricular negativo, alocação negativa do tempo de aula para enfatizar tópicos abordados por um teste, treinamento excessivo, e “trapaça”. Mais importante, os pesquisadores não têm, em geral, medido a extensão ou magnitude das mudanças na prática que eles identificam como um resultado de testes de alto impacto.

No geral, as evidências sugerem que os testes em larga escala de alto impacto têm sido uma política relativamente potente em termos de trazer mudanças dentro das escolas e salas de aula. [...] Vai demorar mais tempo e mais pesquisa para determinar que os impactos positivos no ensino e no aprendizado do aluno superam os negativos.

David Nevo (1995), mesmo reconhecendo que muitas das avaliações externas estão inseridas em políticas educacionais que ignoram em suas formulações as necessidades das escolas, trabalha com a necessidade e a possibilidade de diálogo entre as avaliações externas e as avaliações internas – estas conduzidas pelos professores ou pelos profissionais da educação –, pois cada uma das avaliações possuem potencialidades e limitações, ressaltando, entretanto, que esse diálogo deve ser baseado na escola. As avaliações internas poderiam ser identificadas como avaliações institucionais, conduzidas com vistas ao conjunto de elementos que caracterizam o trabalho escolar, não se restringindo, necessariamente, à avaliação da aprendizagem de seus alunos, ainda que essa

seja a avaliação mais importante da escola para o autor.

Para Nevo (1995), no entanto, as políticas educacionais precisariam favorecer e criar condições para esse diálogo. Deveriam começar pelo reconhecimento de que se as avaliações externas podem contar com pessoal mais especializado tecnicamente e trazer resultados decorrentes de um olhar menos condicionado pelas rotinas da escola e, portanto, de maior amplitude, podendo envolver comparações com um conjunto de escolas, essas avaliações não gozam, via de regra, de reconhecimento acentuado no processo de trabalho de cada uma das escolas, algo que só uma avaliação interna dispõe, ainda que à custa de menor precisão sobre alguns objetos de avaliação. Nesses termos, até mesmo para que a ideia de avaliação formativa possa ser posta em prática, o autor defende que ambas as avaliações se complementam, superando suas limitações e adicionando suas potencialidades.

Deve-se considerar, ainda, que no debate e na literatura há questionamentos em torno da validade, da adequação dos instrumentos utilizados e da confiabilidade dos resultados relativos às avaliações padronizadas. Tais questionamentos colocam em dúvida o que são capazes de medir e da própria medição, em termos de aprendizagem dos alunos, o cerne daquilo para o que foram projetadas e com a precisão anunciada de seus resultados (Cf. BROADFOOT, 1996; CASASSUS, 2013).

Se tais dúvidas apontam para a necessidade de discussão de aspectos técnicos, consideramos que esta deveria ser realizada em conjunto com a discussão sobre os objetivos das avaliações, sua disseminação e usos dos resultados. O debate técnico, inclusive para que se garanta a meta-avaliação dessas avaliações, exigência decorrente da necessidade de controle social sobre as mesmas, como destacou Alavarse (2013, p. 148), forçosamente deve ser aliado ao debate pedagógico, a fim de que ajustes pudessem ser realizados em prol da consolidação de avaliações mais significativas,

ou seja, cujos resultados pudessem alimentar debates e impulsionar o desenvolvimento do sistema educacional.

Outra questão que tem sido discutida é a da seleção dos conteúdos e habilidades que embasam essas avaliações, bem como a definição de níveis ou padrões de rendimento, tanto em seu caráter técnico – o que medir, como medir e como avaliar os resultados –, quanto pedagógico – as interpretações em relação ao conteúdo de ensino e suas condições. Nesse sentido, tem-se questionado as orientações na elaboração das matrizes e na definição de parâmetros de avaliação, como são os pontos de corte em relação às escalas de medida, elementos que não, necessariamente, seriam condizentes com as necessidades sociais e possibilidades de desenvolvimento dos alunos (Cf. BROADFOOT, 1996; CASASSUS, 2013; IAIES, 2003).

Ademais, tem-se alertado para o fato de que a definição de padrões (*standards*) de proficiência gera um empobrecimento curricular, impulsionando redes de ensino a direcionarem esforços para garantir o ensino do que é cobrado nas avaliações, deixando de lado outros conteúdos fundamentais para a educação básica das crianças e adolescentes em idade escolar (Cf. HYPÓLITO, 2013; SANTOS, 2013).

É possível encontrar, também, na literatura de referência, questionamentos acerca do potencial de as avaliações em larga escala melhorarem a qualidade do ensino, visto que, apesar dos esforços realizados, especialmente pelos países latino-americanos, os dados obtidos por meio dessas avaliações mostram que pouco tem se alterado no quadro educacional da maioria dos países que as empregam enquanto instrumento de gestão (Cf. TEDESCO, 2003). A isso se soma a própria definição do que seria a qualidade da educação e as possibilidades e termos de sua mensurabilidade, produzindo polêmicas adicionais que, no caso brasileiro, envolvem o Índice de Desenvolvimento da Educação Básica (IDEB), como se encontra, entre outras fontes, em Machado; Alavarse (2014) e Oliveira (2011, 2013).

Tentando entender essa aparente contradição, alguns autores têm alertado para o fato de que urge recuperar o sentido de avaliar os sistemas educacionais e analisar os modelos que têm sido utilizados para essas avaliações. Há quem argumente que os resultados das provas não têm tido o impacto esperado na melhoria da qualidade do sistema educacional, porque as avaliações configuram-se em modelos cuja prioridade é medir o rendimento dos alunos e não avaliar² o sistema como um todo. A inexistência de clareza a respeito dos objetivos da avaliação e a dificuldade de definir e produzir consenso em torno de padrões de qualidade claros, que permitam comparar longitudinalmente os resultados obtidos e que possam ser utilizados para analisar as possíveis mudanças que se operam a partir das políticas e programas implementados, faz com que a preocupação recaia sobre os resultados das provas e sua dimensão técnica, não se enfatizando as análises contextuais que permitiriam um melhor entendimento da situação educacional e uma intervenção mais efetiva, como destacou Iaies (2003, p. 18), para quem

Os sistemas educativos deixaram de trabalhar para melhorar a qualidade e a equidade educativa e passaram a trabalhar para o melhoramento dos resultados das avaliações. [...] Os dispositivos e seus produtos têm tido maior impacto na construção do imaginário educativo da sociedade, que na transformação das estratégias educativas.

Nesse sentido, há críticas que apontam que as decisões tomadas a partir dos resultados obtidos, muito relacionadas à redefinição de conteúdos curriculares, na formulação de indicadores de uma boa educação, da

2- A diferenciação entre avaliação e medida tem sido alvo de atenção de inúmeros pesquisadores da área e pode ser encontrada em Vianna (2000). A avaliação consiste em um processo mais amplo que pode tomar a medida como uma de suas dimensões, mas se associa à elaboração de juízos de valor sobre a medida e a proposição de ações a partir dela.

definição de competências e habilidades a serem desenvolvidas nos alunos, parecem não ecoar na escola, não tendo sentido para aqueles que, em última instância, fazem as reformas acontecerem, ou seja, os professores e técnicos escolares. Essas avaliações, em geral, não atentam para outras demandas que as escolas enfrentam, como as necessidades sociais do público que tem acessado a escolarização formal e que, anteriormente, estava excluído dela (Cf. OLIVEIRA, 2007, 2011), as características de formação dos profissionais que se dedicam à educação, a garantia de condições de educabilidade mínimas, para que os estudantes possam produzir suas aprendizagens (Cf. IAIES, 2003).

Dessa forma, haveria uma distância entre as considerações técnicas que embasam a avaliação e o debate político-educacional que precisa ser enfrentado tanto na escola como nos níveis centrais, pelos gestores do sistema. Como provoca Iaies (2003, p. 20-21),

Construíram-se indicadores que se definiram tecnicamente, e que consideram quase com exclusividade as habilidades acadêmicas. Nossos índices não consideram o aumento das taxas de escolarização, a capacidade do sistema para homogeneizar atores de uma sociedade cada dia mais segmentada, dar conta dos novos públicos que a escola tem sido capaz de abrigar, da capacidade de conter outras realidades sociais, etc. E essas definições implicam uma tomada de postura ideológica, utilizar umas variáveis e abandonar outras; o certo é que a experiência dos '90 faz pensar mais em uma "não tomada" de posição política, no sentido de que os tomadores de decisões não se posicionaram nesse ponto.

Outros argumentos contrários às avaliações padronizadas que aparecem na literatura, consideram que:

1. A responsabilização de professores e escolas pode levar a medidas punitivas injustas, como a perda do emprego de gestores

e professores ou, até mesmo, o fechamento de escolas ou sua maior supervisão pelo Estado, entendida como perda de autonomia;

2. As avaliações em larga escala interferem na autonomia dos docentes, influenciando até mesmo a forma como os conteúdos são ensinados;

3. Provas padronizadas e *rankings* incentivam a competição entre escolas e alunos, substituindo o aprendizado pela melhoria nos resultados em provas, gerando processos de ensinar para os testes, o que, no limite, substitui o aprendizado pela difusão de macetes que potencializam os resultados (SANTOS, 2013);

4. As avaliações externas padronizadas, usualmente realizadas em um ou dois dias, não consideram fatores externos que podem afetar o desempenho dos alunos. Há estudos que mostram que muitos alunos não apresentam bom desempenho em testes com itens de múltipla escolha, bem como que os resultados dos alunos são influenciados por fatores como ansiedade, por exemplo;

5. As avaliações podem gerar o fenômeno do afunilamento curricular, levando os professores a ensinar somente os conteúdos cobrados nas provas, sem se preocupar em desenvolver outros conteúdos e habilidades importantes à formação dos alunos (Cf. MADAUS; RUSSEL; HIGGINS, 2009);

6. As avaliações são parciais, normalmente realizadas para poucas disciplinas curriculares, e não conseguem captar o crescimento geral no decorrer do ano letivo, induzindo, se muitos valorizados os resultados, ao afunilamento curricular (Cf. BAUER, 2013; MADAUS; RUSSELL; HIGGINS, 2009);

7. Produzem injustiças relativas à bonificação de professores e premiação das melhores escolas (Cf. BAUER, 2012; BROOKE, 2013; FREITAS, 2013; SOUSA, 2008);

8. A pressão para melhoria de resultados pode levar os dirigentes escolares a tomar decisões sobre o gerenciamento dos tempos e conteúdos a serem ensinados que podem ter influência negativa sobre professores e alunos

(Cf. MADAUS; RUSSELL; HIGGINS, 2009). Dentre as ações que as escolas podem realizar para melhorar os resultados estão a recusa de matrícula a alunos de baixo rendimento ou que necessitem de atendimento educacional especializado ou utilização de mecanismos de transferência de alunos de baixo rendimento (Cf. BROADFOOT, 1996; OLIVEIRA et al., 2013);

9. Os estudantes e professores, ao se sentirem pressionados pela situação de avaliação, podem sofrer problemas de saúde (aumento dos níveis de stress) e, ainda, sentirem-se desmotivados ou criar um sentimento negativo em relação à escola (Cf. FARBER, 2010; MADAUS; RUSSELL; HIGGINS, 2009); e

10. Na busca por melhores resultados, podem induzir o aumento na desigualdade, posto que investir mais nos melhores alunos parece mais promissor do que enfrentar as dificuldades de aprendizado dos alunos com piores resultados (Cf. OLIVEIRA et al., 2013).

Estes argumentos indicam que se os resultados das avaliações em larga escala forem sobrevalorizados, suas eventuais contribuições podem ser sobrepujadas por indução de resultados indesejáveis ou mesmo deletérios para escolas, redes de ensino, alunos e profissionais da educação.

Entretanto, o papel que essas avaliações efetivamente assumirão no conjunto da política educacional é determinado pelos usos dos seus resultados, algo que ou não é definido a priori de modo mais completo ou sequer é cogitado na adoção de avaliações em larga escala. Assim, dada sua relevância, é pertinente a análise de como a literatura trata os usos dos resultados das avaliações em larga escala.

O caráter político das avaliações em larga escala

Dentre as muitas dimensões deste debate, encontra-se a influência da divulgação dos resultados sobre os seus usos, a relação entre os objetivos das avaliações e as informações por elas produzidas, bem como a discussão

política e ideológica em torno do assunto, o que sublinha o caráter político dessas avaliações. Considere-se, assim, primeiramente, a dimensão que relaciona os usos dos resultados à sua divulgação e disseminação, consideradas bastante marcadas política e ideologicamente.

Ainda que divulgar e disseminar os resultados obtidos seja de fundamental importância para fomentar a discussão a respeito de uma dimensão da qualidade da educação e, até mesmo, propiciar que escolas e gestores obtenham dados e informações para subsidiar a tomada de decisões, ao nível local ou regional, observa-se que as práticas de publicização e disseminação das avaliações padronizadas são amplamente discutidas.

Por um lado, encontram-se os defensores da publicização dos resultados e, até mesmo, dos ranqueamentos, que consideram que tais práticas dão transparência para o público, evidenciando a forma como as instituições atuam e utilizam os recursos públicos, como em Castro (2007).

Cabe destacar que, considerando os aspectos técnicos das avaliações, as generalizações são muitas vezes indevidas e a divulgação dos resultados brutos, sem os necessários detalhamentos, não se sustentam, pois impedem que auxiliem na gestão dos sistemas educacionais como um todo (Cf. BRITZ, 2003; KLEIN; FONTANIVE, 1995). Britz (2003, p. 94) alerta para possíveis problemas na divulgação dos resultados na forma como vem sendo realizada e aponta, como consequência, a limitação do uso da informação, bem como a tendência de recuo na divulgação de informações para os meios de comunicação em massa, dado que:

A massificação dos indicadores impedirá de circunscrever seu uso ao âmbito para o qual foram desenhados: cada vez mais será usado como um *ranking* para tomar decisões imediatas pela família, cada vez terá mais sentido transformá-lo em base de um debate político. Esta desnaturalização

dos alcances dos indicadores finalmente pode levar ao imobilismo: estariam dispostas as autoridades a seguir dando transparência a sua gestão sem que os instrumentos se voltem contra si? Estariam dispostos os instrumentos de comunicação a não disporem destes dados?

Complementarmente, Brunner (2003, p. 81-82), ao analisar a divulgação dos resultados das avaliações, especialmente nos jornais e periódicos, salienta que estes não atentam para a complexidade dos resultados e o contexto onde são produzidos. Para o autor, os periódicos

[...] se limitam a ressaltar unidimensionalmente os produtos finais destes processos – a saber, os resultados enquanto pontuação – e a organizá-los em um esquemático *ranking* de países ou tipos de estabelecimento, suprimindo toda referência ao contexto onde se obtêm estes resultados. Com isso, se banaliza a informação que aportam os estudos nacionais e internacionais de medida do rendimento escolar e, o que é mais grave, se distorce a opinião pública e se limitam seus efeitos.

No que se referem às contribuições dos resultados das avaliações em larga escala para as escolas e sistemas educativos, alguns aspectos positivos têm sido destacados na literatura. Nessa perspectiva, as avaliações padronizadas:

1. Definem padrões e expectativas para o aprendizado dos alunos, podendo servir para orientar o trabalho das escolas (Cf. BOMENY, 1997). Servem, assim, de guia ao planejamento dos professores, dando-lhes pistas sobre o que e quando ensinar aos alunos, apoiando o gerenciamento de sala de aula e evitando, assim, perda de tempo didático (Cf. EVERS; WALBERG, 2002);

2. Mediante o emprego de testes e seus resultados, geram informações que ficam disponibilizadas para os professores e as

escolas, podendo apoiar a tomada de decisões em seu interior (Cf. EVERS; WALBERG, 2002; MADAUS; RUSSELL; HIGGINS, 2009);

3. Produzem, para comunidades, informações sobre a qualidade do ensino nas escolas, em relação aos componentes curriculares avaliados, auxiliando os pais a tomar decisões bem fundamentadas sobre onde desejam que seus filhos estudem (Cf. EVERS; WALBERG, 2002);

4. Fariam com que professores e alunos buscassem melhorar seu desempenho, vinculadas com políticas de premiação a elas associadas, como motivação extrínseca (Cf. MADAUS; RUSSELL; HIGGINS, 2009); e

5. Teriam potencial para manter professores e escolas responsáveis pelo aprendizado de todos os alunos, à medida que os informam acerca dos alunos que não estão aprendendo conforme o esperado (Cf. EVERS; WALBERG, 2002).

Há ainda autores que discutem que algumas implicações das avaliações sobre as redes e escolas não são, em essência, boas ou ruins, mas que podem ter influências diferenciadas, a depender de cada contexto escolar. Segundo Madaus; Russell, Higgins (2009, p. 3), por exemplo, nem sempre os testes têm uma influência sobre escolas, professores e alunos, pois “alguns alunos e professores simplesmente ignoram os testes e continuam a fazer o que sempre fizeram”, corroborando apontamentos de Stecher (2002). Para Mons (2009), problemas podem começar a surgir quando se utiliza o resultado de avaliações locais (regionais, estaduais ou municipais) para serem tanto ferramenta de gerenciamento quanto medida de efetividade dos resultados, visto que o uso gerencial pode gerar consequências que deturpam o próprio dado obtido nas avaliações locais.

No entanto, as críticas aos efeitos das avaliações externas nas escolas e nas redes de ensino superam os argumentos favoráveis, pela observação de que se concentram no uso dos resultados associado a políticas de alto impacto. Diversos são os autores que se detêm na análise

de consequências negativas dos testes como, por exemplo, a tomada de decisões sobre a continuidade de estudos dos alunos, os recursos financeiros a serem disponibilizados para as escolas, a bonificação de professores, bem como outros usos relativos ao gerenciamento das redes escolares, contratação e demissão de gestores escolares, envio de recursos extras para escolas de bom desempenho etc. (Cf. CASASSUS, 2013; MADAUS; RUSSELL; HIGGINS, 2009; MONS, 2009; RAVITCH, 2010). Ainda que se considere que muitos desses resultados negativos não sejam intencionais, é importante salientar que não se pode deixar de considerá-los no debate.

Além das críticas realizadas ao uso dos resultados das avaliações em políticas de alto impacto, a leitura dos textos de referência para este artigo permite afirmar que, em geral, as críticas referem-se a influências das avaliações sobre o currículo, o trabalho realizado pela escola e, inclusive, sobre as abordagens metodológicas adotadas pelos professores em sala de aula. Uma consequência seria, por exemplo, que conteúdos e disciplinas não avaliados receberiam menor atenção nas escolas, inclusive com a diminuição do tempo dedicado à sua aprendizagem, fenômeno conhecido como afinilamento curricular, em que há um direcionamento das práticas de sala de aula para preparar os alunos para obter bons resultados (Cf. BROOKE, 2013; CASASSUS, 2013; MADAUS; RUSSELL; HIGGINS, 2009). A valorização dos resultados educacionais, em detrimento dos processos escolares pode, também, incentivar professores e escolas a procurar estratégias para aumentar os resultados obtidos sem, necessariamente, melhorar o aprendizado dos alunos (Cf. BROADFOOT, 1996).

Apesar disso, é possível localizar limites dos usos dos resultados das avaliações relacionados às dimensões técnica, política e ideológica, conforme apontado anteriormente.

Fica claro que o debate a respeito das avaliações em larga escala deve ser encarado em sua complexidade, a fim de que os sistemas já consolidados possam se desenvolver e

produzir informações que permitam superar os usos políticos e ideológicos que vêm sendo feitos dos resultados. Este debate contribuiria efetivamente para iluminar a problemática educacional, possibilitando a realização de ações que sejam direcionadas para a melhoria da educação, o que exige uma análise que ultrapasse a comparação de resultados quantitativos sobre níveis de aprendizagem, mas que considere também aspectos curriculares, de infraestrutura e de formação docente, dentre outros (Cf. ESQUINSANI, 2010; IAIES, 2003).

É importante observar que apesar de muitos pesquisadores se dedicarem ao estudo das avaliações de sistema e aos problemas delas decorrentes, estas reflexões parecem ter pouco impacto entre os gestores das políticas, visto que a incorporação da crítica e uso dos conhecimentos elaborados para a reorientação das avaliações parece estar fora do rol de intenções dos técnicos responsáveis pelo gerenciamento das avaliações. Broadfoot (1996) ilumina essa questão identificando uma desarticulação, um distanciamento, entre os acadêmicos da área de avaliação e os responsáveis pela definição das políticas, que se mantêm isolados em seus respectivos campos de atuação.

De fato, a maioria das avaliações externas em larga escala pretende medir o nível de aprendizagem dos alunos brasileiros em determinadas disciplinas – normalmente, português e matemática – e correlacioná-lo a determinadas condições extra e intraescolares que poderiam favorecer ou não a aprendizagem nessas disciplinas. Mas, mensurar níveis de aprendizagem, a partir de determinados critérios, é o mesmo que medir qualidade de ensino?

Conclusão: um convite à reflexão

Apresentados os argumentos favoráveis e contrários, é necessária uma breve reflexão para julgar as avaliações em larga escala. Certamente, as críticas não são desprezíveis. Levantam problemas reais. Entretanto,

majoritariamente, elas incidem sobre questões relativas ao uso inadequado de seus resultados.

A preocupação com o conjunto do sistema educacional coloca-nos questões substantivas, certamente não respondidas pelo discurso da “autonomia da escola”. Se aceitarmos esse argumento estamos fazendo vistas grossas às desigualdades já existentes. Ou seja, as experiências escolares pregressas da população são diferenciadas, de modo que aquelas que tiveram menos ou nenhuma experiência anterior com a escola (em geral a parcela mais pobre e discriminada da sociedade) têm menos condições de formular uma crítica que induza a escola a melhorar. Ao contrário, as evidências indicam que essas populações tendem a ver a escola mais positivamente, afinal, qualquer escola é melhor que nenhuma. Assim sendo, cabe ao centro do sistema, seus administradores, se preocupar com a redução da desigualdade e a garantia do direito ao aprendizado a todos.

Da mesma forma, alimentar ilusões a respeito do poder indutor das avaliações em larga escala é desconhecer que a resposta das escolas

e professores a estímulos por melhoria a partir de indicadores ancorados nessas avaliações pode ter um sentido diverso do pretendido. Podem estimular comportamentos perversos, dos quais os mais conhecidos são a exclusão das populações que se supõe terão os piores resultados, já nos processos de matrícula, sua ocultação nos dias de aplicação das provas, convidando-os a não comparecerem, ou simplesmente uma amplificação das desigualdades no interior da escola, investindo-se mais nos alunos que apresentem potencialmente perspectivas de melhores resultados.

Assim, diante do se poderia denominar de visão ingênua das virtudes das avaliações em larga escala, inclusive porque constatamos características que deveriam ser suprimidas, consideramos que essas avaliações têm potencialidades para produzir avanços no conhecimento do universo educacional. Entendemos que podem, garantidas determinadas condições e adequada utilização, estabelecer pontos de apoio para políticas de melhoria dos sistemas educacionais no sentido da garantia do direito à educação a todos.

Referências

AFONSO, Almerindo. **Avaliação educacional: regulação e emancipação: para uma sociologia das políticas educativas contemporâneas**. São Paulo: Cortez, 2000.

AFONSO, Almerindo. **Políticas educativas e avaliação educacional**. Braga: Universidade do Minho, 1998.

ALAVARSE, Ocimar Munhoz. Desafios da avaliação educacional: ensino e aprendizagem como objetos de avaliação para a igualdade de resultados. **Cadernos Cenpec**, São Paulo, v. 3, n. 1, p. 135-153, jun. 2013.

ARELARO, Lisete Regina Gomes. Direitos sociais e política educacional: alguns ainda são mais iguais que outros. In: SILVA, Shirley; VIZIM, Marli (Org.). **Políticas públicas: educação, tecnologias e pessoas com deficiências**. Campinas: Mercado de Letras, 2003. p. 13-36. (Leituras no Brasil).

BARRERA, Susana. Políticas educacionais, qualidade da educação na América Latina e aporte da avaliação como ferramenta de gerenciamento social. In: CONFERÊNCIA INTEGRAÇÃO E CONVERGÊNCIA ENTRE O MERCOSUR E A COMUNIDADE ANDINA DE NAÇÕES, 2000, La Paz. **Anais...** La Paz: Instituto Internacional de Integração (I.I.I.), 2000.

BAUER, Adriana. É possível relacionar avaliação discente e formação de professores? **Educação em Revista**, Belo Horizonte, v. 28, n. 2, p. 61-82, jun. 2012.

BAUER, Adriana. Limites, desafios e possibilidades das avaliações de sistemas educacionais: contribuições do ciclo de debates para as políticas de avaliação. In: BAUER, Adriana; GATTI, Bernadete A.; TAVARES, Marialva R. (Org.). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: implicações nas redes de ensino, no currículo e na formação de professores**. Florianópolis: Insular, 2013. p. 281-294.

BAUER, Adriana. Usos dos resultados das avaliações de sistemas educacionais: iniciativas em curso em alguns países da América. **Revista Brasileira de Estudos Pedagógicos**, Brasília, v. 91, n. 228, p. 315-344, maio/ago. 2010.

BOMENY, Helena (Org.) **Avaliação e determinação de padrões na educação latinoamericana**: realidades e desafios. São Paulo: FGV, 1997.

BONAMINO, Alicia C. Avaliação educacional no Brasil 25 anos depois: onde estamos? In: BAUER, Adriana; GATTI, Bernardete A. (Orgs.). **Vinte e cinco de avaliação de sistemas educacionais no Brasil**: origens e pressupostos. Florianópolis: Insular, 2013. p. 43-60.

BONAMINO, Alicia C. **Tempos de avaliação educacional**: o SAEB, seus agentes, referências e tendências. Rio de Janeiro: Quartet, 2001.

BRITZ, Pablo H. Indicadores y sus usos en educación: una tensión no resuelta. In: Unesco. **Evaluar las evaluaciones**: una mirada política acerca de las evaluaciones de la calidad educativa. Buenos Aires: IPE: Unesco, 2003. p. 85-95.

BROADFOOT, Patrícia. **Education, assessment and society**: a sociological analysis. Buckingham: Open University Press, 1996.

BROOKE, Nigel. Políticas estaduais de bonificação: buscando o diálogo. In: BAUER, Adriana; GATTI, Bernardete A. (Org.). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil**: origens e pressupostos. Florianópolis: Insular, 2013. p. 119-146.

BRUNNER, José Joaquín. Límites de la lectura periodística de resultados educacionales. In: **Evaluar las evaluaciones**: una mirada política acerca de las evaluaciones de la calidad educativa. Buenos Aires: IPE: Unesco, 2003. p. 67-84.

CASASSUS, Juan. Política y metáforas: un análisis de la evaluación estandarizada en el contexto de la política educativa. In: BAUER, Adriana; GATTI, Bernardete A.; TAVARES, Marialva R. (Org.). **Vinte e cinco de avaliação de sistemas educacionais no Brasil**: origens e pressupostos. Florianópolis: Insular, 2013. p. 21-46.

CASTRO, Maria H. G. A consolidação da política de avaliação da educação básica no Brasil. **Meta: avaliação**, Rio de Janeiro, v. 1, n. 3, p. 271-296, set./dez. 2009a.

CASTRO, Maria H. G. O desafio da qualidade. In: ITUASSU, Arthur; ALMEIDA, Rodrigo de (Org.). **O Brasil tem jeito?**: educação, saúde, justiça e segurança. v. 2. Rio de Janeiro: Jorge Zahar, 2007. p. 35-72.

CASTRO, Maria H. G. Sistemas de avaliação da educação no Brasil: avanços e novos desafios. **São Paulo em Perspectiva**, São Paulo, v. 23, n. 1, p. 5-18, jan./jun. 2009b.

COSTA, Marta E. A avaliação de desempenho na área de língua na Argentina: a utilização de provas objetivas. In: CONHOLATO, Maria Conceição; FERREIRA, Maria José do Amaral (Coord.). **Sistemas de avaliação educacional**. São Paulo: FDE, 1998. p. 39-88. (Ideias, n. 30).

ESQUINSANI, Rosemar Siqueira S. Performatividade e educação: a política das avaliações em larga escala e a apropriação da mídia. **Práxis Educativa**, Ponta Grossa, v. 5, n. 2, p. 131-137, jul./dez. 2010.

ESTEBAN, Maria Teresa. Considerações sobre a política de avaliação da alfabetização: pensando a partir do cotidiano escolar. **Revista Brasileira de Educação**, Rio de Janeiro, v. 17, n. 51, p. 573-592, set./dez. 2012.

EVERS, Williamson M.; WALBERG, Herbert J. **School accountability**. Stanford: Hoover Institution Press, 2002.

FARBER, Katy. **Why great teachers quit**: and how we might stop the exodus. Thousand Oaks: Corwin, 2010.

FERNANDES, Reynaldo; GREMAUD, Amaury Patrick. Qualidade da educação: avaliação, indicadores e metas. In: VELOSO, Fernando et al. (Orgs.). **Educação básica no Brasil**: construindo o país do futuro. Rio de Janeiro: Elsevier, 2009. p. 213-238.

FERRER, Alejandro Tiana. Avaliação e mudança de sistemas educacionais: a interação que falta. In: SEMINÁRIO INTERNACIONAL DE AVALIAÇÃO DA EDUCAÇÃO, 1996, Brasília, **Anais...** Brasília: MEC, 1996. p. 33-54.

FERRER, J. G.; ARREGUI, Patrícia. **Provas internacionais de aprendizado aplicadas na América Latina e seu impacto na qualidade da educação**: critérios para futuras aplicações. Rio de Janeiro: Preal, 2003.

- FREITAS, Luiz C. Caminhos da avaliação de sistemas educacionais no Brasil: o embate entre a cultura da auditoria e a cultura da avaliação. In: BAUER, Adriana; GATTI, Bernardete A.; TAVARES, Marialva R. (Org.). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origens e pressupostos**. Florianópolis: Insular, 2013. p. 147-176.
- HAGOPIAN, Jesse (Ed.). **More than a score: the new uprising against high-stakes testing**. Chicago: Haymarket Books, 2014.
- HOUSE, Ernest. Arranjos institucionais para avaliação. In: SEMINÁRIO INTERNACIONAL DE AVALIAÇÃO EDUCACIONAL, 1998, Brasília. **Anais...** Brasília: MEC; Inep; Unesco, 1998.
- HYPOLITO, Álvaro M. A necessária meta-avaliação das políticas de avaliação. In: BAUER, Adriana; GATTI, Bernardete A. (Org.). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origens e pressupostos**. Florianópolis: Insular, 2013. P. 211-227.
- IAIES, Gustavo. Evaluar las evaluaciones. In: Unesco. **Evaluar las evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa**. Buenos Aires: ILPE; Unesco, 2003. p. 15-36.
- KELLAGHAN, Thomas; GREANEY, Vicent. **Using examinations to improve education: a study in fourteen African countries**. Washington, DC: World Bank, 1992.
- KLEIN, Ruben; FONTANIVE, Nilma S. Avaliação em larga escala: uma proposta inovadora. **Em Aberto**, Brasília, v. 15, n. 66, p. 29-34, abr./jun. 1995.
- LEVIN, Benjamin. **Reforming education: from origins to outcomes**. New York: Routledge; Falmer, 2001. (Educational Change and Development).
- LIMA, Licínio C. Avaliação, competitividade e hiperburocracia. In: ALVES, Palmira; DE KETELE, Jean-Marie (Org.). **Do currículo à avaliação, da avaliação ao currículo**. Porto: Porto, 2011. p. 71-82.
- MACHADO, Cristiane; ALAVARSE, Ocimar Munhoz. Qualidade das escolas: tensões e potencialidades das avaliações externas. **Educação & Realidade**, Porto Alegre, v. 39, n. 2, p. 413-436, abr./jun. 2014.
- MADAUS, George; RUSSELL, Michael; HIGGINS, Jennifer. **The paradoxes of high stakes testing: how they affect students, their parents, teachers, principals, schools, and society**. Charlotte: Information Age, 2009.
- MONS, Nathalie. **Theoretical and real effects of standardised assessment**. [S.l.]: EACEA; Eurydice, 2009.
- NEVO, David. **School-based evaluation: a dialogue for school improvement**. Oxford: Pergamon, 1995.
- OLIVEIRA, Romualdo P. A utilização de indicadores de qualidade na unidade escolar ou porque o IDEB é insuficiente. In: BAUER, Adriana; GATTI, Bernadete A.; TAVARES, Marialva R. (Org.). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: implicações nas redes de ensino, no currículo e na formação de professores**. Florianópolis: Insular, 2013. p. 87-100.
- OLIVEIRA, Romualdo P. Da universalização do ensino fundamental ao desafio da qualidade: uma análise histórica. **Educação e Sociedade**, Campinas, v. 28, n. 100, p. 661-690, out. 2007.
- OLIVEIRA, Romualdo P. O direito à educação na Constituição de 1988 e seu reestabelecimento pelo sistema de Justiça. **Revista Brasileira de Educação**, Rio de Janeiro, v. 11, p. 61-74, maio/ago, 1999.
- OLIVEIRA, Romualdo P. Reformas educativas no Brasil na década de 90. In: CATANI, Afrânio Mendes; OLIVEIRA, Romualdo P. (Org.). **Reformas educacionais em Portugal e no Brasil**. Belo Horizonte: Autêntica, 2000. p. 77-94.
- OLIVEIRA, Romualdo P. IDEB e trabalho pedagógico da escola: uma articulação possível? **Revista Escola Pública**, São Paulo, v. especial, p. 76-92, 2011.
- OLIVEIRA, Romualdo P. et al. Análise das desigualdades intraescolares no Brasil. **Estudos e Pesquisas Educacionais**, v. 4, p. 19, 2013.
- OZGA, Jenny. **Investigação sobre políticas educacionais: terreno de contestação**. Porto: Porto, 2000.
- POPHAM, W. James. The merits of measurement-driven instruction. **Phi Delta Kappan**, v. 68, n. 9, p. 679-682, 1987.

RAVITCH, Diane. **The death and life of the great American school system**: how testing and choice are undermining education. New York: Basic Books, 2010.

SANTOS, Lucíola L. A avaliação em debate. In: BAUER, Adriana; GATTI, Bernardete A. (Org.). **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil**: origens e pressupostos. Florianópolis: Insular, 2013. p. 229-245.

SOUSA, Sandra Z. **Ensino fundamental**: avaliando as políticas de avaliação educacional. Brasília: MEC/INEP, 2003. Texto apresentado no Seminário Avaliar para quê? Avaliando as políticas de avaliação educacional. Brasília, 08 abr. 2003.

SOUSA, Sandra Z. Avaliação e carreira do magistério: premiar o mérito? **Revista Retratos de Escola**, Brasília, v. 2, n. 2/3, p. 81-93, jan./dez. 2008.

SOUSA, Sandra Z.; OLIVEIRA, Romualdo P. Sistemas estaduais de avaliação: uso dos resultados, implicações e tendências. **Cadernos de Pesquisa**, São Paulo, v. 40, p. 793-822, 2010.

STECHEER, Brian M. Consequences of large-scale, high-stakes testing on school and classroom practice. In: HAMILTON, Laura; STECHER, Brian M.; KLEIN, Stephen P. (Ed.). **Making sense of test-based accountability in education**. Santa Monica: Rand, 2002. p. 79-100.

TEDESCO, Juan C. Prólogo. In: UNESCO. **Evaluar las evaluaciones**: una mirada política acerca de las evaluaciones de la calidad educativa. Buenos Aires: ILPE/Unesco, 2003. p. 11-14.

VIANNA, Heraldo Marelím. Avaliação: considerações teóricas e posicionamentos. In: VIANNA, Heraldo Marelím. **Avaliação educacional**: teoria, planejamento, modelos. São Paulo: Ibrasa, 2000. p. 21-46.

VIANNA, Heraldo Marelím. **Avaliações nacionais em larga escala**: análises e propostas. São Paulo: Fundação Carlos Chagas / DPE, 2003. 41 p. (Textos FCC, 23).

Recebido em: 07.01.2015.

Aprovado em: 11.03.2015.

Adriana Bauer é pesquisadora do Núcleo de Estudos em Avaliação, do Departamento de Pesquisas Educacionais da Fundação Carlos Chagas e professora do Departamento de Metodologia do Ensino e Educação Comparada da Faculdade de Educação da Universidade de São Paulo. É membro do Conselho Editorial da revista Estudos em Avaliação Educacional.

Ocimar Munhoz Alavarse é professor no Departamento de Administração Escolar e Economia da Educação, da Faculdade de Educação da Universidade de São Paulo, onde coordena o Grupo de Estudos e Pesquisas em Avaliação Educacional (Gepave).

Romualdo Portela de Oliveira é professor Titular no Departamento de Administração Escolar e Economia da Educação, da Faculdade de Educação da Universidade de São Paulo, pesquisador 1D do CNPq, membro da coordenação de Ciências Humanas 3 (Educação e Psicologia) da FAPESP e Coordenador de Educação da Capes.