

Prova São Paulo: pontos de tensão na avaliação externa em larga escala municipal

MAURICIO DE SOUSA^I

LUANA FERRAROTTO^{II}

<http://dx.doi.org/10.22347/2175-2753v11i33.1892>

Resumo

O artigo insere-se nas discussões sobre os sistemas de avaliação externa em nível municipal. Objetivou-se comparar os resultados das escolas municipais paulistas na Prova São Paulo, avaliação externa em larga escala da própria rede, com os obtidos na Prova Brasil, nos anos de 2007 a 2012. Através do estudo quantitativo comparativo das médias de proficiências dos estudantes dos 5º e 9º anos, em Português e Matemática, de 435 escolas, verificou-se a disparidade entre os resultados na Prova São Paulo comparados com os da Prova Brasil. Conclui-se que as avaliações externas em larga escala em nível local são instrumentos importantes para o estabelecimento de políticas públicas que, por um lado, necessitam de mecanismos de meta-avaliação, para não colocar em xeque o próprio instrumento. Por outro, faz-se necessário não atribuir a responsabilidade por tais resultados apenas às escolas e seus atores, deixando de lado o papel do Estado, enquanto responsável por uma educação pública de qualidade.

Palavras-chave: Prova São Paulo. Avaliação externa em larga escala. Desempenho escolar.

Submetido em: 05/08/2018

Aprovado em: 31/07/2019

^I Unicamp e Secretaria Municipal de Educação de São Paulo (SMESP), São Paulo, (SP), Brasil; <http://orcid.org/0000-0002-4259-6901>; e-mail: sousama72@gmail.com.

^{II} Unicamp e Instituto Federal de São Paulo (IFSP), Bragança Paulista, (SP), Brasil; <http://orcid.org/0000-0002-2861-2127>; e-mail: luanaferrarotto@yahoo.com.br.

Prova São Paulo: tension points in the external evaluation in municipal large-scale

Abstract

This article is part of the discussions about external evaluation systems at the municipal level. The objective was to compare the results of municipal schools in the *Prova São Paulo*, a large-scale external evaluation of the network itself, with those obtained in *Prova Brasil*, between 2007 and 2012. Through a comparative quantitative study of the average proficiency of students in the 5th and 9th grades, in Portuguese and Mathematics, from 435 schools, it was verified a disparity between the results in *Prova Sao Paulo* when compared to *Prova Brasil*. It is concluded that large-scale external evaluations at local level are important instruments for the establishment of public policies that, on the one hand, require meta-evaluation mechanisms, in order not to jeopardize the instrument itself. On the other side, it is necessary not to assign responsibility for such results only to schools and their actors, leaving aside the role of the State as responsible for public education of quality.

Keywords: *Prova São Paulo*. Large-Scale external evaluation. School performance.

Prueba São Paulo: puntos de tensión en la evaluación externa municipal a gran escala

Resumen

Este artículo es parte de las discusiones sobre los sistemas de evaluación externa a nivel municipal. El objetivo fue comparar los resultados de las escuelas municipales de São Paulo, en *Prova São Paulo* (Prueba São Paulo), una evaluación externa municipal a gran escala realizada por esta red de enseñanza, con los obtenidos en *Prova Brasil* (Prueba Brasil), entre 2007 y 2012. A través del estudio cuantitativo comparativo de la competencia promedio de los estudiantes de 5º y 9º años, en portugués y matemáticas, de 435 escuelas, como resultado, verificamos una disparidad entre los resultados de *Prova Sao Paulo* (Prueba São Paulo) en comparación con *Prova Brasil* (Prueba Brasil). Se concluye que las evaluaciones externas a gran escala a nivel local son instrumentos importantes para el establecimiento de políticas públicas que, por un lado, necesitan mecanismos de metaevaluación, para no poner en peligro el instrumento en sí. Por otro lado, es necesario no asignar responsabilidad por tales resultados solo a las escuelas y sus actores, dejando de lado el papel del estado como responsable de la educación pública de calidad.

Palabras clave: Prueba Sao Paulo. Evaluaciones externas a gran escala. Rendimiento escolar.

Introdução

As avaliações externas em larga escala consolidaram-se no Brasil, sobretudo a partir dos anos 1990, com a criação do Sistema de Avaliação da Educação Básica (SAEB)¹, elaborado e coordenado pelo governo federal, através do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Hoje, além do SAEB, há vários sistemas/programas de avaliação externa instituídos pelos estados e municípios (GATTI, 2012; BAUER et al., 2015; SOUSA, 2013). De acordo com Sousa (2013), a incorporação dos resultados das avaliações externas por gestores, nas decisões relativas à educação municipal, intensificou-se com a implementação do Índice de Desenvolvimento da Educação Básica (IDEB). Bauer et al. (2015), em uma pesquisa realizada com a participação de 4.309 municípios, identificou que 37% destes têm avaliações externas próprias e 21% desejam ter um sistema próprio de avaliação. Essa pesquisa demonstrou que há uma forte tendência dos municípios em incorporar a lógica das avaliações externas em larga escala como um dos mecanismos indutores da política pública educacional.

São Paulo está entre os municípios que possuem seu próprio sistema de avaliação em larga escala. Entre 2007 e 2012, houve, na referida rede, a Prova São Paulo², a qual abrangia os componentes curriculares de Língua Portuguesa (foco na leitura) e Matemática (foco na resolução de problemas), e que, nos dois últimos anos de sua existência, contemplou, também, Ciências. Em 2013, com a mudança da gestão municipal³, a Prova São Paulo avaliação foi encerrada, segundo justificativa do Secretário de Educação naquele momento:

Avaliando a série histórica da aplicação da Prova São Paulo, não conseguimos observar tendência alguma, nem para cima, nem para baixo, nem para nada. Os resultados são totalmente discrepantes

1 A implantação do sistema de avaliação externa em larga escala tem início nos anos de 1990 com o Sistema de Avaliação do Ensino Público de 1º Grau (SAEP), em 1990. A partir de 1992, adotou-se o nome de Sistema de Avaliação da Educação Básica (SAEB) que foi oficialmente instituído através da Portaria nº 1.795 de 27 de dezembro de 1994. Em, 2005, o SAEB foi reestruturado e passou a ser composto por duas avaliações: Avaliação Nacional da Educação Básica (Aneb) e a Avaliação Nacional do Rendimento Escolar (Anresc), conhecida como Prova Brasil. Em 2013, foi criada uma nova avaliação para aferir os níveis de alfabetização chamada de Avaliação Nacional da Alfabetização (ANA). Dessa forma, o SAEB atualmente é composto por três avaliações: Aneb, Anresc/Prova Brasil e ANA.

2 Existem algumas pesquisas sobre a Prova São Paulo que analisam o impacto dessa avaliação externa em larga escala no currículo, na gestão escolar e suas possibilidades de usos pedagógicos. Entre essas pesquisas destacamos: ALMEIDA; NHOQUE (2012); BORN; CHAPPAZ (2013, 2014); CHAPPAZ; ALVARSE (2018); GIMENES (2013); MACHADO; FREITAS (2014); MACHADO; ALAVARSE; OLIVEIRA (2015); SAVIOLI (2013; SIQUEIRA (2018).

3 O candidato Fernando Haddad venceu José Serra nas eleições de 2012. Seu mandato abrangeu o período de 2013 a 2016.

entre uma edição e outra. Nos últimos anos foram desperdiçados R\$ 14 milhões, porque os dados não puderam ser aproveitados. É um desperdício para a cidade. (ESTADÃO, 2013, não paginado).

Complementando essa afirmação, o então prefeito da cidade afirmou que “os dados das avaliações externas promovidas pelo governo federal já eram suficientes para a adoção de políticas públicas na educação municipal” (ESTADÃO, 2013, não paginado).

Frente ao exposto, o objetivo deste artigo é fazer uma análise dos resultados da Prova São Paulo e verificar se ocorreu tal discrepância. Cabe ressaltar que essa avaliação em larga escala foi retomada pela atual gestão municipal⁴. Assim, é de extrema importância analisar os dados das edições anteriores da Prova São Paulo, no sentido de qualificar o debate acerca da utilização de instrumentos de avaliação externa em larga escala pelas redes municipais.

Para uma melhor compreensão deste artigo, primeiramente realizou-se a análise documental dos atos normativos e do processo de operacionalização da Prova São Paulo. Logo depois, são apresentados os dados das médias de proficiências dos estudantes da 4ª série/5º ano e da 8ª série/9º ano, em Língua Portuguesa e Matemática, das escolas participantes de todas as edições dessa prova. Essas médias foram comparadas com os resultados obtidos na Prova Brasil, a fim de identificar se há variabilidade entre os resultados. Conclui-se o artigo destacando que as avaliações externas em larga escala em nível municipal são instrumentos importantes para o estabelecimento de políticas públicas, mas que necessitam de mecanismos de meta-avaliação, pois nem sempre são esclarecidas as decisões técnicas e políticas que orientaram sua construção, bem como procedimentos analíticos adotados. Ademais, a responsabilidade por tais resultados, por vezes, é atribuída apenas às escolas e seus atores, deixando de lado o papel do Estado, enquanto responsável por uma educação pública de qualidade social.

A Prova São Paulo: características e operacionalização

A cidade de São Paulo está entre os municípios que implementaram um sistema próprio de avaliação externa em larga escala. Esse sistema teve início na gestão de

⁴ O prefeito Fernando Haddad tentou a reeleição, em 2016, mas foi derrotado em primeiro turno por João Dória Júnior.

José Serra⁵, que deu um novo rumo para avaliação externa em larga escala na rede municipal de ensino. O primeiro documento desta gestão foi lançado, em 24 de fevereiro de 2005, no Diário Oficial do Município (DOM), intitulado “Educação no Município de São Paulo: uma proposta de discussão”. Já na introdução, esse documento apresentava como um dos problemas mais relevantes da educação municipal a qualidade no ensino fundamental. De acordo com o texto, “já existe universalização, mas a qualidade é precária” (SÃO PAULO, 2005a, p. 5). São reconhecidos, também, os esforços dos professores e das gestões passadas, mas destaca-se que a qualidade era muito ruim, utilizando como fundamento para essa afirmação os resultados das avaliações externas “em que pese não haver consenso na literatura educacional a respeito dos dados coletados nas avaliações realizadas pelo SAEB e PISA (Programa Internacional de Avaliação de Estudantes), tais resultados são indicativos importantes sobre a questão da qualidade de ensino” (SÃO PAULO, 2005a, p. 5).

Frente ao cenário educacional apresentado nesse documento, um dos mecanismos adotados pela gestão municipal foi a instituição de um sistema próprio de avaliação externa, cujos resultados pudessem ser acompanhados pelos gestores educacionais e pela sociedade. Essa avaliação foi normatizada em 14 de outubro de 2005 pela Lei nº 14.063 (SÃO PAULO, 2005b).

De acordo com este ato normativo, além de verificar o desempenho dos estudantes, os resultados da avaliação externa subsidiariam a formação de professores e a reorientação da proposta pedagógica das escolas, fornecendo dados para a implementação de uma política de recuperação paralela para estudantes com dificuldades de aprendizagem. Para tanto, no ano de 2006, o Decreto nº 47.683 (SÃO PAULO, 2006) determinou a criação do Núcleo de Avaliação Educacional (NAE), responsável por: a) coordenar e supervisionar as ações de avaliação educacional, internas e externas; b) estruturar e coordenar trabalhos para a criação e implementação de um banco de dados para armazenar os resultados das avaliações e elaborar estatísticas institucionais acerca da educação municipal; c) propor e gerenciar trabalhos vinculados aos processos de avaliações educacionais (SÃO PAULO, 2006).

⁵ José Serra não concluiu o seu mandato, uma vez que se afastou para disputar as eleições estaduais. O vice, Gilberto Kassab, assumiu como prefeito. Logo depois, Kassab foi eleito como prefeito da cidade para o período de 2009 a 2011 e deu continuidade à mesma política educacional.

Inicialmente, no ano de 2007, o NAE – em conjunto com um grupo de professores e gestores da rede municipal de ensino e com o apoio de uma assessoria externa (composta por professores universitários) – organizou a “Matriz de Referência para a Avaliação do Rendimento Escolar” (SÃO PAULO, 2007), que continha as “competências e habilidades básicas que se esperam dos estudantes ao final de cada período avaliado” (SÃO PAULO, 2011, p. 19-20).

Cabe ressaltar que o processo de elaboração dessa Matriz de Referência mostrou

A necessidade de adequação às **Matrizes de Referência do SAEB/Prova Brasil**, para quarta (5º ano) e oitava séries (9º ano), também foi considerada como um argumento para a manutenção de habilidades não validadas dentro dessas séries, uma vez que a matriz e a própria avaliação da Prefeitura Municipal de São Paulo devem representar os conhecimentos consensualmente aceitos e avaliados também em nível nacional. (SÃO PAULO, 2007, p. 4, grifo nosso).

Isto é, a Matriz de Referência para Avaliação do Rendimento Escolar da Prova São Paulo adotou as mesmas habilidades para os anos finais do ensino fundamental I (4ª série/5º ano) e do ensino fundamental II (8ª série/9º ano) que a Matriz de Referência do SAEB/Prova Brasil. Em relação à Língua Portuguesa, ainda se encontra uma ou outra habilidade que não havia na Matriz de Referência do SAEB/Prova Brasil. Entretanto, em relação à Matemática, são as mesmas habilidades nas duas matrizes. Além disso, a escala utilizada pela Prova São Paulo é a mesma escala do SAEB⁶, ou seja, os itens utilizados também foram calibrados na mesma métrica. Esses elementos comuns demonstram um alinhamento entre a avaliação de nível federal e a de nível local (município).

Assim, no período de 2007 a 2012, a Prova São Paulo foi aplicada anualmente, e sua organização ficou a cargo de empresas contratadas mediante licitação, realizada pela Secretaria Municipal de Educação de São Paulo (SMESP). É importante ressaltar que esse processo de contratação de empresas especializadas em

6 A escala de proficiência do SAEB é construída com base na TRI e “[...] atribui a cada item do teste uma posição que reflete o seu grau empírico de dificuldade, ou seja, o grau de dificuldade observado de acordo com o comportamento do item quando apresentado a participantes de diversos níveis de proficiência. A proficiência de cada participante do teste, estimada a partir de suas respostas aos itens, é posicionada nessa mesma escala, permitindo a comparação entre estudantes e entre ciclos da avaliação. Os resultados de desempenho nos testes da Aneb e Anresc são expressos por números na escala de proficiência que variam de 0 a 500 pontos, com média de 250 e desvio-padrão de 50” (INEP, 2018, 45-46).

avaliação externa em larga escala envolve um custo muito alto para uma rede de ensino municipal, especialmente na dimensão da cidade de São Paulo. Na consulta feita ao DOM, constatou-se um gasto de mais de 35 milhões de reais para a realização das seis edições da Prova São Paulo. Essas empresas⁷ eram responsáveis pela organização da prova, distribuição, análise dos resultados e elaboração dos relatórios pedagógicos. Quanto aos itens da prova, não existe uma clareza de quem eram os elaboradores. Os relatórios mencionam que foi composto um banco de itens para a Prova São Paulo, sendo alguns fornecidos pelo INEP e outros que seriam elaborados por profissionais da rede orientados pelo Núcleo de Avaliação da SMESP.

A aplicação era feita pelos professores da própria rede de ensino, com a orientação e a supervisão de coordenadores contratados pela empresa vencedora.

Em seu primeiro ano, a Prova São Paulo avaliou, censitariamente, os estudantes da 2ª, 4ª, 6ª e 8ª série (3º, 5º, 7º e 9º anos) e, além dos testes, continha questões dissertativas e uma produção de texto. A partir de 2008, passou a ser aplicada de forma censitária para os estudantes do ensino fundamental I (2ª série/3º ano e 4ª série/5º ano) e ensino fundamental II (6ª série/7º ano e 8ª série/9º ano) e de forma amostral nas demais séries/ano do ensino fundamental. Até 2010, os estudantes realizavam as provas de Língua Portuguesa e Matemática em dois dias. Nos anos de 2011 e 2012, as provas passaram a ser aplicadas em três dias em virtude da inserção de um novo componente curricular: Ciência e Natureza.

Na primeira aplicação da Prova São Paulo, houve uma junção da Teoria Clássica de Testes (TCT)⁸ e da Teoria de Resposta ao Item (TRI)⁹. A partir de 2009, a metodologia passou a contemplar apenas a TRI, com os resultados expressos na escala SAEB, variando entre 100 a 500 pontos (ALVES; SOARES, 2007). De acordo com o documento "Matriz de Referência para a Avaliação do Rendimento Escolar",

7 As empresas vencedoras das licitações foram as seguintes: Instituto de Tecnologia e Desenvolvimento de Minas Gerais (nome fantasia Herkenhoff & Prates), em 2007 e 2008; Fundação Cesgranrio, em 2009; Fundação Universidade de Brasília, em 2010; Avalia Qualidade Educacional Ltda, em 2011; Consultoria e Planejamento em Administração Pública (Consuplan), em 2012.

8 A TCT preocupa-se em explicar o resultado final total, isto é, a soma das respostas dadas a uma série de itens, expressa no chamado escore total (T). Por exemplo, o T em um teste de 30 itens de aptidão seria a soma dos itens corretamente acertados. Se for dado 1 para um item acertado e 0 para um errado e o sujeito acertou 20 itens e errou 10, seu escore T seria de 20 (PASQUALI, 2009, p. 993).

9 A TRI é uma "teoria do traço latente" aplicada primariamente a testes de habilidade ou de desempenho. O termo "teoria do traço latente" refere-se a uma família de modelos matemáticos que relaciona variáveis observáveis (itens de um teste, por exemplo) e traços hipotéticos não observáveis ou aptidões, estes responsáveis pelo aparecimento das variáveis observáveis, ou melhor, das respostas ou comportamentos emitidos pelo sujeito, que são as variáveis observáveis. Assim, temos um estímulo (item) que é apresentado ao sujeito e este responde a ele. A resposta que o sujeito dá ao item depende do nível que o sujeito possui no traço latente ou aptidão. Dessa forma, o traço latente é a causa e a resposta do sujeito é o efeito (PASQUALI; PRIMI, 2003, p. 101-102).

Nas provas do quarto ano dos Ciclos I e II (4ª série/5º ano e 8ª série/9º ano) **serão incluídos blocos de itens já calibrados do SAEB, para que os resultados possam ter comparabilidade nacional**. Os resultados de todas as escolas serão colocados na escala nacional do SAEB". (SÃO PAULO, 2007, p. 5, grifo nosso).

Novamente, observa-se o alinhamento da Prova São Paulo à Prova Brasil, pois os itens (questões) foram calibrados de acordo com a escala SAEB. Cabe destacar que os relatórios pedagógicos desta prova informam que foram utilizados itens cedidos pelo INEP e outros foram elaborados por profissionais da rede, sendo estes pré-testados e calibrados conforme a escala SAEB. Entretanto, não informam claramente quantos itens foram cedidos ou elaborados por profissionais desta rede de ensino.

Dessa forma, a pontuação era convertida em níveis de proficiência denominados: Abaixo do Básico, Básico, Adequado e Avançado. Conforme a série/ano avaliada, os níveis de proficiência atingiam diferentes valores na escala numérica, ou seja, conforme o aluno avançava em sua escolarização, tais valores também aumentavam.

Os resultados da Prova São Paulo eram disponibilizados para as escolas no formato de um relatório pedagógico com o desempenho individualizado de seus estudantes. Residia aí a principal diferença em relação às avaliações externas em larga escala promovidas pelo governo federal, que encaminhavam para as escolas um relatório com a média de desempenho das séries/anos avaliados, e não dos estudantes individualmente. De posse dos resultados da Prova São Paulo, as empresas contratadas elaboravam boletins com o desempenho individual para ser entregues aos estudantes. Em 2010, os boletins passaram a ser encaminhados, via correio, para a residência das famílias. Naquele mesmo ano, os resultados começaram a ser divulgados de forma comparativa entre escola, Diretoria de Ensino (DRE) e município, conforme observa-se na Figura 1.

Figura 1 – Boletim eletrônico

MÉDIA DA PROFICIÊNCIA DOS ALUNOS DA ESCOLA, DA DRE E DA REDE PARA OS ANOS DO CICLO II						
	1º Ano <150	1º Ano	Total 1º Ano	2º Ano	3º Ano	4º Ano
LÍNGUA PORTUGUESA						
Escola	■ 152,7	■ 196,2	■ 192,3	■ 199,5	■ 207,2	■ 223,1
DRE	■ 151,5	■ 198,9	■ 184,7	■ 198,6	■ 213,3	■ 219,7
Município	■ 144,4	■ 197,9	■ 175,7	■ 192,4	■ 202,7	■ 213,1
MATEMÁTICA						
Escola	■ 179,5	■ 201,1	■ 199,8	■ 198,4	■ 218,7	■ 232,3
DRE	■ 162,1	■ 209,1	■ 195,5	■ 204,4	■ 216,7	■ 226,0
Município	■ 158,3	■ 203,4	■ 185,5	■ 199,1	■ 208,5	■ 220,3

Fonte: UNB; CESPE (2011, p. 67).

O boletim era um mecanismo de transparência dos resultados das aprendizagens dos estudantes junto à comunidade escolar que poderia pressionar as escolas para melhoria dos resultados, ou seja, uma de responsabilização branda (BONAMINO; SOUSA, 2012). Não tardou muito tempo para a SMESP propor uma política de bonificação para os profissionais da educação vinculado aos resultados da Prova São Paulo. Assim, em 2011, deu-se a criação do Índice da Qualidade da Educação (Indique). Este indicador era obtido a partir dos resultados da Prova São Paulo e o “perfil socioeconômico com o qual cada unidade escolar trabalha”.¹⁰ O perfil socioeconômico era identificado via questionário respondido pela família do estudante. Vale ressaltar que havia, ainda, um questionário voltado ao aluno acerca de hábitos de estudo e outro para os professores, gestores, coordenadores pedagógicos e supervisores de ensino sobre o trabalho pedagógico relativo aos componentes curriculares avaliados.

De acordo com o Secretário de Educação da época, o Indique apresentava inovações, se comparado a outros indicadores educacionais, uma vez que utilizava o nível socioeconômico dos estudantes e o esforço da escola para melhoria dos resultados da Prova São Paulo. Em suas palavras “A gente passa a levar em consideração também as condições nas quais a escola está inserida: qual o público que ela atende, quais são as crianças. Todo esse conjunto de informações pode nos ajudar a compreender os resultados das crianças nas provas” (MACEDO, 2011).

O processo de bonificação, atrelado à avaliação externa, não avançou na rede municipal, pois, em 2013, Fernando Haddad assumiu o governo, extinguindo a Prova São Paulo, utilizada para o cálculo deste indicador.

Prova São Paulo: quais são e o que dizem seus resultados?

Descritas as características da Prova São Paulo, segue a análise dos seus resultados a partir dos dados divulgados no site da SMESP¹¹.

¹⁰ Disponível em: <http://www2.prefeitura.sp.gov.br/indique>. Acesso em: 26 maio 2016.

¹¹ Esses dados estavam disponíveis no site da SMESP do Núcleo de Avaliação. Devido à extinção da Prova São Paulo, em 2013, os dados foram retirados do site. Portanto, os dados apresentados foram todos acessados em 2014 pelo link:

<http://portalsme.prefeitura.sp.gov.br/Projetos/nucleo/AnonimoSistema/MenuTexto.aspx?MenuID=14>. Acesso em: 20 jan. 2014, do qual possuímos uma cópia com todos os resultados das escolas que participaram da avaliação.

Durante as seis edições da Prova São Paulo, foram avaliadas 552 escolas municipais. Desse total, 435 (78,8%) participaram de todas as edições, nos anos de 2007, 2009 e 2011, anos em que também ocorreu a aplicação da Prova Brasil. Assim, todos os dados apresentados são estabelecidos a partir da média de proficiência em Língua Portuguesa e Matemática apresentada pelos estudantes dessas escolas¹².

No Quadro 1, estão os níveis de proficiência da Prova São Paulo para os estudantes da 4ª série/5º ano, a partir da escala SAEB.

Quadro 1 – Níveis de proficiência da Prova São Paulo para a 4ª série/5º ano do ensino fundamental

Disciplinas	Abaixo do básico	Básico	Adequado	Avançado
Língua Portuguesa	< 150	150 a < 200	200 a < 250	≥ 250
Matemática	< 175	175 a < 225	225 a < 275	≥ 275

Fonte: PESQUISA & AVALIAÇÃO (2011, p. 67).

Com base nessa classificação, na Tabela 1 seguem os resultados das escolas que participaram de todas as edições da Prova São Paulo dos estudantes da 4ª série/ 5º ano, em Língua Portuguesa.

Tabela 1 - Número de escolas de 4ª série/5º ano do ensino fundamental, de acordo com os níveis de proficiências da Prova São Paulo, em Português

	2007	2008	2009	2010	2011	2012
Abaixo do básico	24	26	4	30	11	10
Básico	407	399	413	396	406	398
Adequado	4	9	18	8	18	26
Avançado	0	0	0	0	0	0
Não divulgado*	0	1	0	1	0	1
Total	435	435	435	435	435	435

* Algumas escolas não tiveram os dados divulgados pela SMESP nos anos de 2008, 2010 e 2012. Não foi apresentada nenhuma justificativa pela não divulgação dos dados. Este fato pode colocá-los em dúvida, pois em algumas situações foi divulgado o desempenho da escola em Língua Portuguesa e não em Matemática, ou o contrário. Contudo, nestes anos não ocorreu a Prova Brasil, a qual, consequentemente, não interferiu na comparabilidade do desempenho das escolas entre ambas as avaliações.

Fonte: Os autores (2018), adaptado de SÃO PAULO (2014).

De acordo com a Tabela 1, pode-se perceber uma variação muito pequena no nível de proficiência básico, que fica em torno de 400 entre as 435 escolas. Os níveis abaixo do básico e adequado apresentam uma variação maior. Assim, em 2007, 24 escolas apresentavam o nível abaixo do básico; em 2012, ocorreu uma redução para

¹² Neste texto, buscamos observar as variações das médias entre as duas avaliações externas em larga escala: Prova São Paulo e Prova Brasil. Entretanto, sabemos da importância de se realizar um estudo demonstrando as variações no desempenho individual de cada estudante e em cada escola. Para isso, seria necessário ter acesso aos microdados dos resultados da Prova São Paulo.

10 escolas. Em relação ao nível adequado, tem-se quatro escolas em 2007, número elevado para 26 em 2012. Já no nível avançado, não se tem nenhuma escola em todas as edições da prova São Paulo.

Quanto à proficiência em Matemática, a Tabela 2 apresenta os resultados das escolas de 4ª série/5º ano.

Tabela 2 – Número de escolas de 4ª série/5º ano do ensino fundamental, de acordo com os níveis de proficiências da Prova São Paulo, em Matemática

	2007	2008	2009	2010	2011	2012
Abaixo do básico	234	90	116	153	155	154
Básico	201	344	319	277	279	280
Adequado	0	0	0	1	1	0
Avançado	0	0	0	2	0	0
Não divulgado	0	1	0	2	0	1
Total	435	435	435	435	435	435

Fonte: Os autores (2018), adaptado de SÃO PAULO (2014).

Nessa disciplina, observa-se que houve uma variação maior entre os anos de 2007 e 2009, nos níveis de proficiência abaixo do básico. Em 2007, existiam 234 escolas, que foram reduzidas a 90 em 2008 e subiram para 116 em 2009. Da mesma forma, o nível de proficiência básico apresentou grande variação. Em 2007, existiam 201 escolas neste nível de proficiência, número que subiu para 344 em 2008 e caiu para 319 escolas em 2009.

De 2010 a 2012, houve estabilidade nos níveis de proficiência abaixo do básico e básico. Quanto ao nível adequado, os anos de 2010 e 2011 apresentaram uma escola cada. Por fim, no ano de 2010, duas escolas estavam no nível de proficiência avançado.

A julgar por esses dados, os estudantes da 4ª série/5º ano do ensino fundamental apresentaram proficiência maior em Língua Portuguesa, índice que se manteve estável em todas as edições da Prova São Paulo. Já no componente curricular de Matemática, houve uma variação maior, mas o que chama a atenção é o aumento do número de escolas no nível abaixo do básico, em comparação aos anos de 2008 e 2009.

Em relação aos estudantes da 8ª série/9º ano do ensino fundamental, a classificação dos níveis de proficiência, de acordo com a Prova São Paulo, é a que pode ser vista no Quadro 2.

Quadro 2 - Níveis de proficiência da Prova São Paulo para a 8ª série/9º ano do ensino fundamental

Disciplinas	Abaixo do básico	Básico	Adequado	Avançado
Língua Portuguesa	< 200	200 a < 275	275 a < 325	≥ 325
Matemática	< 225	225 a < 300	300 a < 325	≥ 350

Fonte: PESQUISA & AVALIAÇÃO (2011, p. 67).

Na Tabela 3 são apresentados os dados das escolas em relação à Língua Portuguesa.

Tabela 3 - Número de escolas de 8ª série/9º ano do ensino fundamental, de acordo com os níveis de proficiências da Prova São Paulo, em Língua Portuguesa

	2007	2008	2009	2010	2011	2012
Abaixo do básico	3	53	2	28	9	6
Básico	432	369	433	407	426	429
Adequado	0	2	0	0	0	0
Avançado	0	0	0	0	0	0
Não divulgado	0	11	0	0	0	0
Total	435	435	435	435	435	435

Fonte: Os autores (2018), adaptado de SÃO PAULO (2014).

Pelos dados apresentados, percebe-se estabilidade quanto ao número de escolas no nível de proficiência básico, em média 416 escolas (desvio-padrão - 22,7 escolas), com exceção do ano de 2008, com 369 escolas. Já em relação às escolas abaixo do básico, o número de instituições não chegou a dez, excetuados os anos 2008 e 2010 com, respectivamente, 53 e 28 escolas. Em contrapartida, no nível adequado, o ano de 2008 apresenta duas escolas, nos demais, nenhuma esteve nesse nível de proficiência. Da mesma forma, nenhuma escola atingiu o nível avançado em Língua Portuguesa nesta série/ano.

Já em relação à Matemática, na 8ª série/9º ano, a Tabela 4 apresenta a seguinte situação.

Tabela 4 - Número de escolas de 8ª série/9º ano, de acordo com os níveis de proficiências da Prova São Paulo, em Matemática

	2007	2008	2009	2010	2011	2012
Abaixo do básico	1	49	72	107	148	177
Básico	434	383	363	328	287	258
Adequado	0	0	0	0	0	0
Avançado	0	0	0	0	0	0
Não divulgado	0	3	0	0	0	0
Total	435	435	435	435	435	435

Fonte: Os autores (2018) adaptado de SÃO PAULO (2014); INEP (2015).

Como se pode observar, ocorreu uma queda no número de escolas no nível básico, no decorrer dos anos. Em 2007, eram 434, passando para 258 escolas em 2012. Em movimento oposto, a quantidade de escolas com nível abaixo do básico cresceu a

cada ano. Em 2007, tem-se apenas uma, constatando-se o número de 177 escolas em 2012. Nenhuma escola esteve no nível adequado durante toda a série histórica.

Comparando as duas séries/anos, de acordo com os níveis de proficiência estabelecidos pela Prova São Paulo, há uma estabilidade entre os níveis de proficiências em Língua Portuguesa nas duas séries/anos do ensino fundamental. No entanto, em Matemática, ocorreu uma queda acentuada no que se refere à proficiência da 8ª série/9º ano do ensino fundamental.

A Prova Brasil e a Prova São Paulo caminhos de intersecção

A Prova Brasil, organizada pelo INEP, faz parte do SAEB e tem como objetivo avaliar a educação básica brasileira. Trata-se de uma avaliação censitária envolvendo os estudantes da 4ª série/5º ano e 8ª série/9º ano do ensino fundamental. São avaliadas as escolas públicas municipais, estaduais e federais que possuem mais de 20 estudantes matriculados nas séries/anos citadas. O objetivo dessa avaliação é contribuir para a melhoria da qualidade da educação pública do país.

Essa avaliação externa em larga escala assumiu o *status* de um dos principais exames do país, pois é um dos componentes para a elaboração do IDEB, indicador que adquiriu legitimidade entre os estudos acadêmicos, com muitos apoiadores e críticos (ALVES; SOARES, 2013, p. 181). Além disso, o resultado da Prova Brasil, como descreve Machado, Alvarase e Oliveira (2015, p. 340), “pode indicar os contornos da qualidade da educação básica no país, bem como fornecer elementos para o direcionamento de políticas com vistas a equacionar dificuldades apresentadas pelos sistemas educacionais”.

Diante desse contexto, os resultados da Prova Brasil, em especial o IDEB, têm fomentado políticas públicas educacionais nos níveis federal, estadual e municipal. Desta forma, mesmo com a criação de uma avaliação externa em larga escala própria, os dados da Prova Brasil não deixaram de ser utilizados pela rede de ensino do município de São Paulo.

A comparação dos resultados dessas duas avaliações externas em larga escala, no período de 2007 a 2012, pode ser feita pelos seguintes motivos:

1- Foram aplicadas de forma censitária para os estudantes da 4ª série/5º ano e 8ª série/9º ano e sempre no final do ano letivo¹³;

2- As matrizes de referência de avaliação nos componentes avaliados abordavam as mesmas habilidades, conforme destacado anteriormente;

3- Os itens da Prova São Paulo foram calibrados conforme a escala SAEB¹⁴, definida pelo INEP;

4- Foi utilizada a mesma metodologia de análise dos itens, a TRI, aliada à organização da prova em Blocos Incompletos Balanceados (BIB)¹⁵. Com a utilização dos BIB, possibilitam que sejam ofertados itens da prova anterior como também itens comuns as séries anteriores, estabelecendo a comparação entre as edições da prova.

Mediante essas similitudes, o desempenho dos estudantes teria proximidade nas duas provas. Além do que, a Prova São Paulo se propôs, desde a sua criação, a ser um instrumento de comparação com os resultados em âmbito nacional. Isto é, a Prova São Paulo foi planejada para ser alinhada à avaliação externa elaborada pelo governo federal (Prova Brasil) quanto à elaboração, metodologia e análise dos resultados.

Para comparação dos resultados dessas duas provas, a rede municipal de ensino de São Paulo realizou um levantamento das médias de proficiências das 435 escolas que participaram de todas as edições da Prova São Paulo, nos componentes de Língua Portuguesa e Matemática, nos mesmos anos em que foi aplicada Prova Brasil: 2007, 2009 e 2011. Para essa comparabilidade, utilizou-se duas classificações: i- Prova São Paulo nos níveis de proficiência abaixo do básico, básico, adequado e avançado; ii- a escala SAEB elaborada pelo INEP para a Prova Brasil. Com base nessas comparações, observou-se que há uma variação no resultado alcançado pelas escolas.

¹³ Esclarecemos que a Prova São Paulo tinha uma abrangência maior, pois era aplicada da 2ª série/3º ano até a 8ª série/9º ano (sendo algumas séries de forma censitária e outras amostral).

¹⁴ Os blocos incompletos balanceados são "um esquema otimizado para o rodízio de blocos com aplicações em diversas áreas, inclusive educação e agricultura. A necessidade do rodízio se justifica se pressupusermos que possuímos b blocos e só podemos utilizar k deles em cada conjunto. Isto é especialmente útil nos sistemas de avaliação quando desejamos obter informações amplas sobre o ensino, utilizando um grande número de itens, ao passo que precisamos limitar a quantidade de itens submetido a cada aluno num valor aceitável e adequado ao tempo de prova" (BEKMAN, 2001, p. 121).

¹⁵ Os blocos incompletos balanceados são "um esquema otimizado para o rodízio de blocos com aplicações em diversas áreas, inclusive educação e agricultura. A necessidade do rodízio se justifica se pressupusermos que possuímos b blocos e só podemos utilizar k deles em cada conjunto. Isto é especialmente útil nos sistemas de avaliação quando desejamos obter informações amplas sobre o ensino, utilizando um grande número de itens, ao passo que precisamos limitar a quantidade de itens submetido a cada aluno num valor aceitável e adequado ao tempo de prova" (BEKMAN, 2001, p. 121).

Em um primeiro momento, tem-se o número de escolas de acordo com a classificação dos níveis de proficiências na Prova São Paulo para a 4ª série/5º ano.

Tabela 5 – Comparação do número de escolas da 4ª série/5º ano do ensino fundamental, em Língua Portuguesa – Prova Brasil e Prova São Paulo, conforme níveis de proficiência da Prova São Paulo

	2007		2009		2011	
	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo
Abaixo do básico	19	24	2	4	2	11
Básico	405	407	406	413	394	406
Adequado	7	4	27	18	39	18
Avançado	0	0	0	0	0	0
Não participaram	4	0	0	0	0	0
Total	435	435	435	435	435	435

Fonte: Os autores (2018), adaptado de SÃO PAULO (2014).

Os dados da Tabela 5 demonstram que, na Prova Brasil, o desempenho em Língua Portuguesa das escolas é melhor que na Prova São Paulo. O número de escolas no nível abaixo do básico diminuiu em ambas as avaliações, mas, na Prova Brasil, a redução foi de quase 90%, enquanto na Prova São Paulo este índice ficou em torno de 50%. Já em relação ao nível adequado, na Prova Brasil houve um aumento constante em cada edição: de sete escolas, em 2007, para 39 escolas em 2011. Na Prova São Paulo, havia quatro escolas em 2007 e 18 em 2009, números que se mantiveram estáveis em 2011.

Em relação à Matemática, nesta mesma série/ano, ocorreu a seguinte situação exposta na Tabela 6.

Tabela 6 – Comparação do número de escolas da 4ª série/5º ano do ensino fundamental, em Matemática – Prova Brasil e Prova São Paulo, conforme níveis de proficiência da Prova São Paulo

	2007		2009		2011	
	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo
Abaixo do básico	77	234	8	116	10	155
Básico	354	201	415	319	414	279
Adequado	0	0	12	0	11	1
Avançado	0	0	0	0	0	0
Não participaram	4	0	0	0	0	0
Total	435	435	435	435	435	435

Fonte: Os autores (2018), adaptado de SÃO PAULO (2014); INEP (2015).

Verifica-se uma diferença acentuada no desempenho das escolas em ambas as avaliações. No nível abaixo do básico, houve uma redução quanto ao desempenho das escolas na Prova Brasil: em 2007, eram 77 e, em 2011, 10 escolas. Na Prova São Paulo, o número de escolas nesse nível de proficiência foi maior do que na Prova Brasil. Em 2007, havia 234 escolas, número reduzido para 155 em 2011. Mesmo com essa redução, a diferença entre ambas as avaliações, no que se refere ao número de escolas com este nível de proficiência, é grande.

Em relação ao nível básico, a situação é semelhante: o número de escolas nesse nível de proficiência foi maior na Prova Brasil do que na Prova São Paulo. Por fim, no nível adequado, na Prova São Paulo, houve apenas uma escola, no ano de 2011. Na Prova Brasil, havia 12 escolas em 2009 e 11 escolas em 2011.

Essa mesma situação se repetiu na 8ª série/9º ano do ensino fundamental, em Língua Portuguesa, de acordo com os dados da Tabela 7.

Tabela 7 - Comparação do número de escolas da 8ª série/9º ano do ensino fundamental, em Língua Portuguesa – Prova Brasil e Prova São Paulo, conforme níveis de proficiência da Prova São Paulo

	2007		2009		2011	
	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo
Abaixo do básico	3	3	4	2	0	9
Básico	426	432	428	433	434	426
Adequado	0	0	1	0	1	0
Avançado	0	0	0	0	0	0
Não participaram	6	0	2	0	0	0
Total	435	435	435	435	435	435

Fonte: Os autores (2018) adaptado de SÃO PAULO (2014); INEP (2015).

Em Língua Portuguesa, o desempenho das escolas foi semelhante. Observando os dados da Tabela 7, percebe-se que o desempenho das escolas na Prova Brasil, no nível abaixo do básico, foi melhor do que aquele apresentado na Prova São Paulo (em 2011, existiam nove escolas, enquanto nenhuma escola se apresentou na primeira). Da mesma forma, na Prova São Paulo, nenhuma escola alcançou o nível de proficiência adequado e, na Prova Brasil, houve apenas uma escola em 2009 e uma em 2011.

Por fim, utilizando os critérios de classificação dos níveis de proficiência da Prova São Paulo relativos à Matemática, na 8ª série/9º ano do ensino fundamental, os dados foram os da Tabela 8.

Tabela 8 - Comparação do número de escolas da 8ª série/9º ano ensino fundamental, em Matemática – Prova Brasil e Prova São Paulo, conforme níveis de proficiência da Prova São Paulo

	2007		2009		2011	
	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo
Abaixo do básico	77	1	76	72	50	148
Básico	352	434	357	363	385	287
Adequado	0	0	0	0	0	0
Avançado	0	0	0	0	0	0
Não participaram	6	0	2	0	0	0
Total	435	435	435	435	435	435

Fonte: Os autores (2018) adaptado de SÃO PAULO (2014); INEP (2015).

Novamente, na Tabela 8, observa-se uma diferença nos resultados de ambas as avaliações externas em larga escala em relação a Matemática na 8ª série/9º ano. Em 2007, apenas uma escola estava no nível abaixo do básico na Prova São Paulo, enquanto 77 estavam nesse nível, de acordo com os resultados da Prova Brasil. No entanto, na Prova Brasil, houve um movimento de queda das escolas nesse nível de proficiência, ao passo que, na Prova São Paulo, o movimento foi de ascensão. Já em relação ao nível básico, na Prova Brasil, o número de escolas variou de 350 a 385 na série histórica. Na Prova São Paulo, ocorreu uma queda acentuada da quantidade de escolas nesse nível, de 434 escolas, em 2007, para 287 escolas, em 2011.

Apresentados esses dados, observa-se que as escolas da rede municipal de São Paulo apresentam melhor resultado na Prova Brasil do que na Prova São Paulo, excetuando os dados relativos à Língua Portuguesa na 8ª série/9º ano do ensino fundamental. Situação semelhante ocorre ao adotar os níveis de proficiência utilizado na Prova Brasil, como consta no Quadro 3.

Quadro 3 - Níveis de proficiências adotado na Prova Brasil para a 8ª série/9º ano do ensino fundamental

	nível 1	nível 2	nível 3	nível 4	nível 5	nível 6	nível 7	nível 8	nível 9
Língua Portuguesa	200 a 225	225 a 250	250 a 275	275 a 300	300 a 325	325 a 350	350 a 375	375 a 400	
Matemática	200 a 225	225 a 250	250 a 275	275 a 300	300 a 325	325 a 350	350 a 375	375 a 400	400 A 425

Fonte: INEP (2015).

Verifica-se, na Tabela 9, o desempenho das 435 escolas na Prova Brasil e na Prova São Paulo, em Língua Portuguesa. Os resultados na Prova Brasil, com exceção do ano de 2007, foram melhores do que na Prova São Paulo.

Tabela 9 - Comparação do número de escolas da 8ª série/9º ano do ensino fundamental, em Língua portuguesa – Prova Brasil e Prova São Paulo, conforme níveis de proficiência da Prova Brasil

	2007		2009		2011	
	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo
Abaixo 200	3	0	4	2	0	9
Nível 1 - 200 a 225	199	26	95	145	66	164
Nível 2 - 225 a 250	202	299	261	253	257	230
Nível 3 - 250 a 275	25	107	72	35	111	32
Nível 4 - 275 a 300	0	3	1	0	1	0
Não participaram	6	0	2	0	0	0
Total	435	435	435	435	435	435

Fonte: Os autores (2018), adaptado de SÃO PAULO (2014); INEP (2015).

Esse mesmo movimento ocorreu no componente curricular de Matemática (Tabela 10), mesmo existindo uma proximidade no número de escolas no nível 2, em ambas as avaliações externas. Também se observa uma queda acentuada no nível 1, quanto ao número de escolas na Prova Brasil, enquanto na Prova São Paulo houve um aumento, passando de uma escola em 2007 para 148 escolas em 2011. Já em relação ao nível 3 de proficiência, tem-se um aumento do número de escolas na Prova Brasil, de 74 em 2007 para 117 escolas em 2011; e queda na Prova São Paulo, de 159 para 49 escolas.

Tabela 10 - Comparação do número de escolas da 8ª série/9º ano do ensino fundamental, em Matemática – Prova Brasil e Prova São Paulo, conforme dos níveis de proficiência da Prova Brasil

	2007		2009		2011	
	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo	Prova Brasil	Prova São Paulo
Nível 1 - 200 a 225	77	1	72	76	50	148
Nível 2 - 225 a 250	276	269	281	288	264	237
Nível 3 - 250 a 275	74	159	80	69	117	49
Nível 4 - 275 a 300	2	6	2	0	4	1
Não participaram	6	0	0	2	0	0
Total	435	435	435	435	435	435

Fonte: Os autores (2018), adaptado de SÃO PAULO (2014); INEP (2015).

Portanto, seja qual for a forma de análise, o desempenho das escolas na Prova Brasil foi melhor do que aquele apresentado na Prova São Paulo. Ademais, mesmo com todas as críticas a esse sistema de avaliação instituído pelo governo federal, através do SAEB/Prova Brasil, seus dados construíram uma série histórica que fundamenta diversas pesquisas acadêmicas nas duas últimas décadas.

O processo de implementação da Prova São Paulo, na maior rede de ensino municipal do país, mobilizou um grande aporte financeiro na contratação de empresas responsáveis pela elaboração, distribuição, coordenação e sistematização dos resultados. Entretanto, não evitou a variação dos resultados na série histórica apresentada¹⁶.

Avaliações externas em larga escala: pontos de tensões

Essas variações entre as duas avaliações podem estar relacionadas às fragilidades das avaliações externas em larga escala. Tais avaliações muitas vezes são difundidas sem que se mencionem eventuais erros de medida¹⁷ e, portanto, são apresentadas como reveladoras do desempenho dos estudantes e da aprendizagem proporcionada (ou não) pela escola (SUPOVITZ, 1997, 2009; KORETZ, 2010; HOLCOMBE; JENNINGS; KORETZ, 2013). Não é pretensão deste texto questionar a validade e fidedignidade dos instrumentos, entretanto, cabe destacar que existem vários fatores, de acordo com a literatura sobre a temática, que podem interferir nos resultados das avaliações externas em larga escala.

Para Holcombe, Jennings e Koretz (2013), entre os principais problemas das variações de desempenho, estão: a escolha e amplitude das habilidades selecionadas para o exame, a apresentação dos itens várias vezes nos exames, a forma como são definidas as habilidades (mais ampla ou restrita), como são apresentados os itens, com uso de imagem, graficamente, algebricamente etc. Para os autores, uma das consequências desses problemas é o treinamento dos estudantes especificamente para os testes, o que pode “inflar” os resultados e não significar aprendizagem, e sim o treino para responder aos testes de múltiplas escolhas.

Koretz (2010) também destaca, em outro artigo, o erro de medida dos testes, pois não existe nenhum teste ou método capaz de medir toda a aprendizagem de um aluno ou a eficácia de um programa desenvolvido por uma escola. Na mesma direção, Supovitz (1997) entende que um único teste também é incapaz de dar conta

¹⁶ Essa situação foi observada pela própria SMESP em 2012: “Os resultados da Prova São Paulo de 2011 e de 2012 não foram oficialmente divulgados. Segundo a gestão Kassab, houve dúvidas em relação à consistência dos dados. Em 2011, houve uma queda acentuada nas médias dos estudantes do terceiro e quarto ano do ensino fundamental em relação a 2010, enquanto no nono ano houve um forte aumento. Esses saltos destoavam da avaliação federal” (TAKAHASHI, 2013).

¹⁷ Pasquali (2013, p. 48) assim descreve o erro de medida: “todos os esforços para controlar o erro através de procedimentos experimentais são necessários, mas nem por isso o erro vai desaparecer, dado que a ocorrência dele é imprevisível, isto é, não é nunca possível se determinarem as causas de todos os erros possíveis numa média”.

de toda a aprendizagem dos estudantes e que seria necessária uma gama variada de exames (testes de múltipla escolha, itens abertos, portfólios etc.) para tentar conseguir medir o desempenho de um aluno.

Em uma crítica às avaliações em larga escala, Supovitz (1997) faz uma comparação com outros profissionais, como o médico, que precisa de diferentes instrumentos para chegar a um diagnóstico, ou seja, por que na educação, apenas um exame, pode decidir o rumo de uma política educacional ou responsabilizar os professores ou a escola, pelos resultados? Além disso, para Supovitz (2009), a experiência norte-americana demonstra que esses exames não têm uma conexão direta com o que acontece na sala de aula, principalmente em virtude da demora na entrega dos resultados, contribuindo muito pouco para o trabalho do professor (exceto se este deseja treinar os seus estudantes para o teste).

Assim como nos EUA, alguns autores também têm demonstrado os problemas com as avaliações externas em larga escala realizadas no Brasil. Em relação à Prova Brasil, Soares e Xavier (2013) apontam como um dos problemas a "inflação" na proficiência de matemática. Outro fator apontado por Viana (2003), ao estudar o SAEB é "[...] a questão da validade é extremamente delicada em qualquer contexto educacional, e, no nosso caso particular, precisamos considerar a formação da nossa nacionalidade, a grande diversidade social, econômica e cultural, demonstrada em todo território brasileiro". Dessa forma, um item pode promover diferentes leituras dependendo do contexto cultural em que se apresenta.

Pasquali (2013), ao descrever o processo de construção e validade dos itens de um exame, destaca a importância da escolha metodológica, pois "[...] na TCT se diz que o item é válido se tiver alta correlação com o critério; na TRI, se ele for uma representação adequada do traço latente. Assim você pode prever que há maneiras diferentes de calcular a validade do item". (p. 140).

Na mesma direção, Rodrigues (2006), ao analisar os itens das provas de matemática do SAEB, de 1997 e 1999, observou problemas relacionados à concentração de itens sobre determinado tema, como exemplo, no exame de 1999, da 8ª série, em que 49% dos itens versavam sobre geometria. A pesquisadora observou, ainda, que o percentual de itens de dificuldade média abaixo do que o recomendado pela literatura sobre a construção de itens em avaliação externa de larga escala. Por fim, a autora afirma que "[...] as provas têm, portanto, o dever de comprovar a sua objetividade, confiabilidade e qualidade" (RODRIGUES, 2006, p. 74).

Para Klein (2013, p. 55), uma forma de reduzir os erros de medida, em especial através do uso da metodologia da TRI é “[...] a construção de cadernos equivalentes e adequadas à população, é necessário um banco de itens calibrados com muitos itens e satisfazendo os critérios desejados”.

A partir dos problemas apresentados pelos autores e descritos acima, as diferenças entre as duas avaliações, Prova São Paulo e Prova Brasil, podem resultar de vários fatores, desde a escolha dos itens até a metodologia de análise do instrumento, cabendo um estudo mais aprofundado, principalmente a partir dos microdados da Prova São Paulo.

Além das questões de caráter técnico, outro problema, a saber: a adoção de políticas sustentadas pelos resultados das avaliações externas, pode acarretar sua validade consequential, descrita por Viana como

O impacto da avaliação sobre o sistema, determinando mudanças de pensamento, gerando novos comportamentos, formando novas atitudes e promovendo novas ações. A validade consequential reflete em que medida a avaliação faz realmente alguma diferença para a comunidade [...] A sociedade, por intermédio da publicação dos resultados em jornais, com inúmeros e bem construídos gráficos e tabelas, que procuram ser auto-explicativos, assiste a tudo sem entender bem o que se passa e, acreditamos, muitos pais se indagarão: - a escola do meu filho se saiu bem? O meu filho teve uma boa nota na avaliação? O meu filho foi melhor ou pior que os seus companheiros de classe? E os seus colegas de série se saíram melhor ou pior do que ele? São grandes incógnitas em uma situação pouco compreensível para a grande massa. (2003, p. 54).

Olhando por esse prisma e através da experiência norte-americana, muitos autores (CASASSUS, 2009; RAVITCH, 2011; MADDAUS; RUSSEL; HIGGINS, 2009) têm destacado que uma educação voltada para as avaliações externas em larga escala provoca entre outras medidas: o estreitamento do currículo, o treinamento dos estudantes e a responsabilização dos professores e gestores.

Analisando o contexto do ensino da cidade de São Paulo, no período de 2007 a 2012, os resultados da Prova São Paulo geraram algumas políticas públicas educacionais que se aproximam da experiência norte-americana, como exemplo, a adoção de material apostilado, para a melhoria nos resultados (SAVIOLI, 2013) e a quase implantação de uma política de bonificação dos profissionais da educação¹⁸.

¹⁸ Na gestão atual de Bruno Covas foi criado um novo indicador Índice de Desenvolvimento da Educação Paulista (IDEP) que recupera os indicadores do Indique e tem a finalidade de bonificação dos professores e gestores desta rede. Disponível em: <http://portal.sme.prefeitura.sp.gov.br//Main/Noticia/Visualizar/PortalSMESP/Secretaria-Municipal-de-Educacao-cria-indice-para-avaliar-Ensino-Fundamental>. Acesso em: 12 ago. 2019.

Para melhorar esses instrumentos de avaliação externa em larga escala em nível estadual e municipal, Avalarse (2013) defende que é preciso

Estabelecer uma avaliação dessas avaliações. O que compreenderia avançarmos para um código de ética das avaliações no qual deveriam estar inscritas condições, de várias ordens, sobretudo no sentido de haver um maior controle sobre tais avaliações. Atualmente, ninguém tem acesso a dados que pudessem garantir, por exemplo, a validade - a propriedade de seus resultados refletirem o que se trabalha nas escolas - e sua fidedignidade - a propriedade de seus resultados terem um erro de medida aceitável. Isso parece ser, aliás, a condição para que as avaliações em larga escala conquistem a legitimidade política diante da comunidade dos profissionais da educação.

Gatti (2012) também é defensora da meta-avaliação das avaliações externas em larga escala, e aprofundando a sua crítica destaca que

Quando se trata de avaliação que atinge pessoas em sua vida escolar, portanto, em sua vida social, há cuidados imprescindíveis a tomar. Envolve questões de ética, além de envolver conhecimento científico e técnico. Não pode ser um processo construído com ligeireza e adesismo. É preciso refletir sobre os objetivos e os impactos desse processo, que tem consequências pessoais, institucionais, sociais. Os processos avaliativos são processos que implicam necessariamente julgamento de valor, e é preciso que se tenha consciência ética em relação aos objetivos, finalidades, procedimentos empregados, socialização das informações e ações decorrentes e seus consequentes. (p. 37).

Considerações finais

Através da descrição do processo de implementação e operacionalização da Prova São Paulo, avaliação externa em larga escala própria da rede municipal paulistana, observou-se o alinhamento, naquele período (2007 a 2012), com a política de avaliação estabelecida pelo governo federal, em especial após a criação do IDEB, que tinha como um dos seus propósitos ser um indicador de qualidade da educação pública no país. Esse alinhamento seguiu a tendência de proliferação de avaliações externas em larga escala, em nível estadual e municipal, abordada em outros estudos (BAUER; 2015; SOUSA, 2013).

Todavia, ainda se fazem necessários novos estudos sobre o processo de elaboração, análise e uso dos dados dessas avaliações produzidas em nível municipal e estadual. Em nossa análise, há disparidade entre os resultados da Prova São Paulo e da Prova Brasil, utilizando-se das médias dos níveis de proficiência das escolas participantes em todas as edições de ambas as avaliações.

Dessa forma, esses resultados apresentados pela Prova São Paulo dificultam a construção de uma série histórica do desempenho dos estudantes da rede de ensino municipal de São Paulo, nos anos de 2007 a 2012 e comprometem o próprio exame em dois aspectos, como foi apontado nos estudos descritos anteriormente: i- validade consequential (VIANA, 2003) que significa rotular uma escola como “boa” ou “ruim” (mesmo não ocorrendo o ranqueamento das escolas, a SMESP estabeleceu ações de acompanhamento, com base nos dados da Prova São Paulo, para as escolas com baixo desempenho); e ii- a legitimidade política (FREITAS et al., 2009), pois seus dados podem não ser reconhecidos e, conseqüentemente, não são utilizados no planejamento das práticas pedagógicas desenvolvidas nas escolas.

Isso não significa que as avaliações externas em larga escala, elaboradas pelas próprias redes de ensino, devem ser eliminadas, pois podem contribuir para o acompanhamento e monitoramento do processo de aprendizagem dos estudantes de forma mais direta e individualizada, no espaço escolar, ao apresentar o desempenho de cada estudante para a escola. De posse dos seus resultados, as escolas/professores podem utilizá-los para: o trabalho de formação continuada de professores, o planejamento das aulas, a organização de conteúdos, a revisão do projeto pedagógico, a compreensão da metodologia utilizada na avaliação externa em larga escala etc. Estas possibilidades são mais limitadas nos exames nacionais ou internacionais, como Prova Brasil ou PISA, que não disponibilizam o resultado individual do estudante.

Todas essas possibilidades que a avaliação externa em larga escala em nível local apresentada tem que ser compreendida de forma ampliada, isto é, entender que a responsabilidade pela educação pública e a formação dos estudantes deve ser partilhada entre poder público e escola. Realizar as avaliações e expor os resultados não significam, em si só, responsabilidade e transparência do poder público. É preciso analisar e contextualizar os resultados, com a possibilidade de realizar demandas bilaterais entre Estado e instituições. Neste sentido, como defendem Darling-Hammond e Ascher (2006), o Estado deve ser responsável pelas políticas que adota, avaliando seus efeitos e fornecendo os recursos necessários para atender aos interesses dos estudantes e da comunidade escolar. A escola, seus professores e demais profissionais, por sua vez, devem ser responsáveis pela concretização de práticas que favoreçam a formação humana das crianças e dos jovens.

Referências

- ALAVARSE, O. M. Avaliar as avaliações em larga escala: desafios políticos. *Revista Educação*, São Paulo, 12 dez. 2013. Disponível em: <http://www.revistaeducacao.com.br/avaliar-as-avaliacoes-em-larga-escala-desafios-politicos/>. Acesso em: 25 maio 2017.
- ALMEIDA, J. G.; NHOQUE, J. R. Avaliação externa: impactos em uma escola da rede municipal de São Paulo. In. *Revista @mbienteeducação*, [s. l.], v. 5, n. 1, p. 83-91, jan./jun. 2012.
- ALVES, M. T. G.; SOARES, J. F. As pesquisas sobre o efeito das escolas: contribuições metodológicas para a sociologia da educação. *Sociedade e Estado*. [s. l.], v. 22, n. 2, p. 435-473, 2007.
- ALVES, M. T. G.; SOARES, J. F. Contexto escolar e indicadores educacionais: condições desiguais para efetivação de uma política educacional. *Educação e Pesquisa*, [s. l.], v. 39, n. 1, p. 177-194, 2013.
- BAUER, A. et al. Avaliação em larga escala em municípios brasileiros: o que dizem os números?. *Estudos em Avaliação Educacional*, [s. l.], v. 26, n. 62, p. 326-352, 2015.
- BEKMAN, R. M. Aplicação dos blocos incompletos balanceados na Teoria de Resposta ao Item. *Estudos em Avaliação Educacional*, [s. l.], n. 24, p. 119-136, 2001.
- BORN, B. B.; CHAPPAZ, R. de O. Currículo e avaliação externa na rede municipal de ensino de São Paulo: reducionismo ou imbricações?. *Roteiro*, [s. l.], v. 39, n. 2, p. 371-390, 2014.
- BORN, B. B.; CHAPPAZ, R. de O. Influências na constituição do currículo oficial na rede municipal de ensino de São Paulo. *Cadernos Cenpec*, São Paulo, v. 3, n. 2, p. 187-208, 2013.
- BONAMINO, A. C; SOUSA, S. Z. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. *Educação e Pesquisa*, São Paulo, v. 38, n. 2, p. 373-388, 2012.
- CASASSUS, J. Uma nota crítica sobre a avaliação estandardizada: a perda de qualidade e a segmentação social. *Sísifo/ Revista de Ciências da Educação*, [Portugal], n. 9, maio/ago. 2009.
- CHAPPAZ, R. de O.; ALAVARSE, O. M. Avaliação externa na Rede Municipal de Ensino de São Paulo: os desafios da participação docente. *Cadernos Cenpec*, São Paulo, v. 7, n. 2, 2018.
- DARLING-HAMMOND, L.; ASCHER, C. Construindo sistemas de controle em escolas urbanas. *Estudos em Avaliação Educacional*, São Paulo, n. 35, p. 7-48, set./dez. 2006.
- ESTADÃO. A extinção da Prova São Paulo. São Paulo. 30 abr. 2013. *Caderno Opinião*. Disponível em: <http://opinio.estado.com.br/noticias/geral,a-extincao-da-prova-sao-paulo-imp-,1026992>. Acesso: 15 abr. 2015.

FREITAS, L. C. et al. *Avaliação educacional: caminhando na contramão*. 2. ed. Petrópolis: Editora Vozes, 2009.

GATTI, B. A. Políticas de avaliação em larga escala e a questão da inovação educacional. *Revista Série-Estudos*, [s. l.], n. 33, p. 29-37, 2012.

GIMENES, N. et al. Além da Prova Brasil: investimento em sistemas próprios de avaliação externa. *Estudos em Avaliação Educacional*, São Paulo, v. 24, n. 55, p. 12-32, 2013.

HOLCOMBE, R.; JENNINGS, J. L.; KORETZ, D. The roots of score inflation: an examination of opportunities in two states' tests. In: SUNDERMAN, G. (ed.). *Charting reform, achieving equity in a diverse nation*. Greenwich, CT: Information Age Publishing, 2013. Disponível em: <https://dash.harvard.edu/bitstream/handle/1/10880587/roots%20of%20score%20inflation.pdf?sequence=1>. Acesso em: 15 maio 2017.

INEP. *Prova Brasil*. Brasília, DF, 2015. Disponível em: <http://portal.inep.gov.br/web/guest/educacao-basica/ideb/resultados>. Acesso em: 20 mar. 2015.

INEP. *Relatório SAEB (ANEB e ANRESC) 2005-2015: panorama da década*. Brasília, DF, 2018, 154 p.

KLEIN, R. Alguns aspectos da teoria de resposta ao item relativos à estimação das proficiências. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 21, n. 78, p. 35-56, 2013.

KORETZ, D. Implications of current policy for educational measurement. In: Center for K-12 Assessment & Performance Management, *Exploratory Seminar: measurement challenges within the race to the top agenda*. Princeton: NJ, 2010. Disponível em: <http://www.ets.org/Media/Research/pdf/KoretzPresenterSession3.pdf>. Acesso em: 12 maio 2017.

MACEDO, L. Prefeitura de São Paulo cria novo critério para a avaliação da qualidade as escolas. *G1*, São Paulo, 27 jul. 2011. Disponível em: <http://g1.globo.com/sao-paulo/noticia/2011/07/prefeitura-de-sp-cria-novo-criterio-para-avaliacao-de-qualidade-das-escolas.html>. Acesso: 20 maio 2016.

MACHADO, C.; FREITAS, P. F. Gestão escolar e avaliação externa: Experiências de escolas da Rede Municipal de São Paulo. *Revista Eletrônica de Educação*, [S. l.], v. 8, n. 3, p. 113-126, 2014.

MACHADO, C.; ALAVARSE, O. M.; OLIVEIRA, A. S. Avaliação da educação básica e qualidade do ensino: estudo sobre os anos finais do ensino fundamental da rede municipal de ensino de São Paulo. *RBPAE*, [s. l.], v. 31, n. 2, p. 335-353, maio/ago. 2015.

MADAUS, G.; RUSSEL, M.; HIGGINS, J. *The paradoxes of high stakes testing: how they affect students, their parents, teachers, principals, schools, and society*. Charlotte, NC: Information Age Publishing, 2009.

PASQUALI, L. *Psicometria: teoria dos testes na psicologia na educação*. Petrópolis, Rio de Janeiro: Vozes, 2013.

PASQUALI, L. Psychometrics. *Revista da Escola de Enfermagem da USP*, São Paulo, v. 43, n. SPE, p. 992-999, 2009.

PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item: TRI. *Avaliação Psicológica*, Porto Alegre, v. 2, n. 2, p. 99-110, dez. 2003.

RAVITCH, D. *Vida e morte do grande sistema escolar americano: como os testes padronizados e o modelo de mercado ameaçam a educação*. Porto Alegre: Editora Sulina, 2011.

RODRIGUES, M. M. M. Proposta de análise de itens das provas do SAEB sob a perspectiva pedagógica e a psicométrica. *Estudos em Avaliação Educacional*, São Paulo, v. 17, n. 34, p. 43-78, maio/ago. 2006.

SÃO PAULO (Município). *Educação no Município de São Paulo: uma proposta para discussão*. Diário Oficial da Cidade de São Paulo. São Paulo, n. 34, 24 fev. 2005a. Disponível em:

https://www.imprensaoficial.com.br/DO/BuscaDO2001Documento_11_4.aspx?link=/2005/suplemento/diario%2520oficial%2520cidade%2520de%2520sao%2520paulo/fevereiro/24/pag_0001_6E7A5PLQGVHBMENCUCQDIAQPJTK.pdf&pagina=1&data=24/02/2005&caderno=Suplemento%20%20Di%3%A1rio%20Oficial%20Cidade%20de%20S%C3%A3o%20Paulo&paginaordenacao=10001. Acesso em: 23 fev. 2016.

SÃO PAULO (Município). Decreto nº 47.683, de 14 de setembro de 2006.

Regulamenta a Lei nº 14.063, de 14 de outubro de 2005, que institui o Sistema de Avaliação de Aproveitamento Escolar dos Alunos da Rede Municipal de Ensino de São Paulo, sob a responsabilidade da Secretaria Municipal de Educação. *Diário Oficial da Cidade de São Paulo*. São Paulo. 15 set. 2006. Disponível em:

http://www3.prefeitura.sp.gov.br/cadlem/secretarias/negocios_juridicos/cadlem/integra.asp?alt=15092006D%20476830000. Acesso: 11 maio 2015.

SÃO PAULO (Município). Lei nº 14063, de 14 de outubro de 2005. Institui o sistema de avaliação de aproveitamento escolar dos alunos da rede municipal de ensino de São Paulo, sob a responsabilidade da secretaria municipal de educação. *Diário Oficial da Cidade de São Paulo*. São Paulo. 15 out. 2015b. Disponível em: <http://cm-sao-paulo.jusbrasil.com.br/legislacao/806118/lei-14063-05>. Acesso: 20 maio 2015.

SÃO PAULO (Município). Secretaria Municipal de Educação. *Prova de São Paulo*. São Paulo, 2014. Disponível em:

<http://portalsme.prefeitura.sp.gov.br/Projetos/nucleo/AnonimoSistema/MenuTexto.aspx?MenuID=14>. Acesso em: 20 jan. 2014.

SÃO PAULO (Município). Secretaria Municipal de Educação. *Matrizes de referência para a avaliação do rendimento escolar*. São Paulo: SME, 2007. 228p.

SÃO PAULO (Município). Secretaria Municipal de Educação. *Matriz de referência para a avaliação do rendimento escolar no ensino fundamental: ciências naturais*. São Paulo: SME, 2011. 54 p. (Versão Preliminar).

SAVIOLI, M. R. *Das potencialidades de uma proposta curricular às fragilidades de um recurso didático: caderno de apoio e aprendizagem (Língua Portuguesa)*. 2013, 219 f. Tese (Doutorado em Educação) - Programa de Pós-Graduação, Pontifícia Universidade Católica de São Paulo, São Paulo, 2013.

SIQUEIRA, V. A. de S. *Avaliações externas e trabalho docente: o que dizem os registros de professores?*. *Cadernos Cenpec | Nova série, [s. l.]*, v. 7, n. 2, 2018.
SOARES, J. F.; XAVIER, F. P. Pressupostos educacionais e estatísticos do Ideb. *Educação & Sociedade, [s. l.]*, v. 34, n. 124, p. 903-923, jul./set. 2013.

SOUSA, S. Z. Avaliação externa e em larga escala no âmbito do estado brasileiro: interface de experiências estaduais e municipais de avaliação da educação básica com iniciativas do governo federal. In: BAUER, A.; GATTI, B. *Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: implicações nas redes de ensino, no currículo e na formação de processos*. Florianópolis: Insular, 2013, p. 61-86. V. 2.

SUPOVITZ, J. A. From multiple choice to multiple choices. *Education Week Commentary, Pennsylvania*, 1997. Disponível em: http://repository.upenn.edu/gse_pubs/280/. Acesso em: 10 maio 2017.

SUPOVITZ, J. Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change, [s. l.]*, v. 10, n. 2-3, p. 211-227, 2009. Disponível em: http://www.cpre.org/sites/default/files/journal/1151_2009canhighstakes.pdf. Acesso em: 11 maio 2017.

TAKAHASHI, F. Haddad acaba com a prova de Kassab que avaliava escolas de São Paulo. *Folha de São Paulo, São Paulo*, 12 abr. 2013. Educação. Disponível em: <http://www1.folha.uol.com.br/educacao/1261366-haddad-acaba-com-prova-de-kassab-que-avaliava-escolas-de-sao-paulo.shtml> Acesso: 20 maio 2016.

PESQUISA & AVALIAÇÃO: *Revista da Escola: Prova São Paulo 2010*, Brasília, n. 3, 116p. 2011. (UNB/CESP; Coordenadoria e Pesquisa em Avaliação).

VIANNA, H. M. Avaliações nacionais em larga escala: análises e propostas. *Estudos em Avaliação Educacional, São Paulo*, n. 27, p. 41-76, jan./jun. 2003.