

DOSSIÊ

Qualidade, aprendizagem e avaliação sistêmica: discursos dos organismos internacionais para os países latino-americanos

Avaliação em larga escala e estabelecimento de metas e indicadores educacionais: usos e abusos da quantofrenia

Large scale assessment and establishment of educational goals and indicators: uses and abuses of quantophrenia

Adriana Bauer^a
adbauer@fcc.org.br

RESUMO

Este artigo problematiza a ênfase dada em municípios brasileiros a aspectos quantitativos do desempenho dos alunos, mensurados por indicadores e médias de proficiência, em detrimento à discussão sobre a aprendizagem e o significado pedagógico dos dados e à produção de indicadores de processos que se associam ao atingimento (ou não) das metas. Inicialmente, contextualiza-se a consolidação e expansão das testagens em larga escala no Brasil e no mundo para, em um segundo momento, discutir iniciativas desenvolvidas nacionalmente, em âmbito municipal. Conclui-se que é necessário retomar o debate técnico e pedagógico, aliado às considerações políticas sobre o uso dos resultados das avaliações e de indicadores educacionais para a proposição de políticas públicas voltadas às escolas.

Palavras-chave: Avaliação em Larga Escala. Índices Educacionais. Municípios. Desempenho. Quantofrenia.

ABSTRACT

This article seeks to problematize the emphasis given by Brazilian municipalities to quantitative aspects of student performance, measured by large-scale assessments, instead of the discussion about the learning process and the pedagogical meaning of data and the production of process indicators associated (or not) to achievement of learning goals. For this, it showed the consolidation and expansion of the large-scale tests, with national and international examples, to focus on initiatives developed in Brazil at the district level. The conclusions highlight the need to retake the technical debate, along with political considerations, on the use of assessment results and educational indicators for the proposal of public policies directed to schools.

Keywords: Large-Scale Assessment. Educational Indexes. Districts. Performance. Quantophrenia.

Introdução

A expansão e consolidação de avaliações externas e em larga escala de redes de ensino e sistemas educativos, fenômeno observado em diversos países e também no Brasil, nos âmbitos

^a Universidade de São Paulo (USP) e Fundação Carlos Chagas (FCC), São Paulo, São Paulo, Brasil.

federal, estadual e municipal (Madaus; Russell; Higgins, 2009; Bauer, 2010; Rey, 2010; Tobin *et al.*, 2015), tem sido bastante discutida nos últimos anos, não constituindo um fenômeno novo. As explicações para tal expansão variam, incorporando desde a necessidade de se acompanhar a evolução do acesso dos alunos aos níveis básicos de educação e do seu nível de aprendizagem, até a influência de novas formas de gerenciamento educacional, propostas por organismos multilaterais, que teriam nessas avaliações uma de suas ferramentas de monitoramento e controle (Rey, 2010; Mons, 2009). Conexões entre a expansão das testagens em larga escala e as mudanças nas lógicas de organização e gerenciamento educacional, ensejadas pela globalização e pela submissão da educação às forças de mercado (Benett, 1998; Afonso, 2000; Ball, 2004), também estão presentes na literatura da área, recheada de argumentos contrários e favoráveis a esse tipo de iniciativa (Bauer; Alavarse; Oliveira, 2015).

Apesar de muito centrada nos contextos estadunidense e europeu, a crescente presença da testagem em larga escala abrange países de todos os continentes. Um levantamento efetuado pela autora em 2010 já apontava que dos 35 países do continente americano, 21 tinham seus próprios sistemas ou iniciativas de avaliação externa (Bauer, 2010).

Na mesma época, Rey (2010) também verificou a ampliação desse tipo de iniciativa no continente europeu, asseverando que, até o início da década passada, a maioria dos países¹ já utilizava algum tipo de avaliação externa padronizada, quer no nível federal, quer regionalmente:

Uma grande maioria dos países europeus agora usa avaliações padronizadas externas em nível regional ou nacional. Até a década de 1990, apenas um pequeno número de países usava testes nacionais na educação obrigatória (ensino fundamental e médio inferior), seja para fins de transição para o próximo ano (Islândia, Portugal, Escócia, Irlanda do Norte, Holanda, Luxemburgo, Dinamarca e Malta) ou para fornecer informações de diagnóstico sobre o sistema educacional como um todo (Irlanda, França, Hungria, Suécia e Reino Unido). Dez outros países e regiões seguiram o exemplo na década de 1990, incluindo Espanha, Comunidade Francesa da Bélgica, Letônia, Estônia e Romênia. Desde o início dos anos 2000, testes nacionais foram introduzidos na Comunidade Flamenga da Bélgica, Lituânia, Polônia, Noruega, Eslováquia, Áustria, Alemanha, Bulgária, Chipre, Dinamarca e Itália (Rey, 2010, p. 140-141)².

O autor aponta que, a partir dos anos 1990, as testagens passam a servir não só para o controle dos resultados em uma perspectiva de Estado Avaliador – que estaria ausente dos processos educativos, controlando apenas os resultados obtidos –, como grande parte da literatura atinente

¹ Baseado no relatório Eurydice (2009), o autor afirma que apenas cinco países não possuem uma proposta de avaliação externa e em larga escala, a saber: República Tcheca, Grécia, País de Gales, Liechtenstein e a comunidade germanófona da Bélgica.

² A large majority of European countries now use external standardized assessments at a regional or national level. Until the 1990s, only a small number of countries used national tests in compulsory education (primary and lower secondary education), either for the purposes of transition to the next year (Iceland, Portugal, Scotland, Northern Ireland, the Netherlands, Luxemburg, Denmark and Malta) or for providing diagnostic information about the education system as a whole (Ireland, France, Hungary, Sweden and the UK). Ten other countries and regions followed suit in the 1990s including Spain, the French Community of Belgium, Latvia, Estonia and Romania. Since the early 2000s, national tests have been introduced in the Flemish Community of Belgium, Lithuania, Poland, Norway, Slovakia, Austria, Germany, Bulgaria, Cyprus, Denmark and Italy (Rey, 2010, p. 140-141).

discute, mas para o direcionamento do que é feito nesses sistemas educativos, permitindo questionar até que ponto há, de fato, uma descentralização da gestão nos diversos âmbitos administrativos que compõem os sistemas de ensino dos países analisados. Tal questionamento é válido, como se irá argumentar posteriormente, para o caso brasileiro, a partir da realidade das avaliações municipais.

De todo modo, observa-se que a lógica de gerenciamento que apela à avaliação externa dos resultados dos alunos como uma ferramenta de controle sobre o que é feito nos sistemas educativos não é privilégio do mundo ocidental. O relatório produzido por Clarke, Liberman e Ramirez (2012), especialistas do *World Bank*, indica que, na Ásia Oriental, nove países contavam, à época do estudo, com sistemas nacionais de avaliação: Camboja, Indonésia, Japão, Coreia do Sul, Laos, Malásia, Mongólia, Tailândia e Vietnã. Além desses países, Hong Kong e Shanghai, na China, também implantaram sistemas de avaliação externa dos alunos, inclusive com orçamento destinado à consecução dessas avaliações. Tobin *et al.* (2015), além de reforçar as informações sobre a maioria dos países asiáticos já citados, informam existir avaliações em larga escala na Nova Zelândia e na Austrália.

Finalmente, têm-se notícias de avaliações em larga escala desenvolvidas em Moçambique (World Bank, 2009), na África do Sul (Hoadley; Muller, 2016), em Uganda (Allen *et al.*, 2016) e em diversos países do continente africano, impulsionadas por três programas avaliativos: o *Monitoring Learning Achievement* (MLA) (Chinapah, 2003), o *Southern Africa Consortium for Monitoring Educational Quality* (SACMEQ) e o *Programme d'Analyse des Systèmes Educatifs des Pays de la CONFEMEN* (PASEC). Kellaghan e Greaney (2001) já haviam identificado 47 países participando do MLA, 15 do SACMEQ e 9 envolvidos no PASEC³. Os autores apontam a vinculação existente entre as diversas avaliações e a proposição de políticas educacionais, não necessariamente propostas ou influenciadas por organismos multilaterais, como é comum aparecer nos estudos críticos do fenômeno. Percebem ainda diversos problemas nos desenhos avaliativos propostos, cujo aprofundamento não cabe no desenvolvimento deste texto.

Ou seja: obter informações de diferentes naturezas acerca do sistema educacional, do fluxo e da aprendizagem dos estudantes passou a ser, desde o final do século XX, uma prática em diversos países. Inicialmente, os mais abastados e centrais e, posteriormente, os que enfrentam mais desafios socioeconômicos, sejam eles semiperiféricos ou periféricos (Barfield, 1997). O que se quer destacar é que a valorização do uso de informações para o gerenciamento educacional, seja por meio de incentivo de organismos multilaterais ou não, é uma realidade independentemente do nível de poder econômico do país ou de seu regime político.

Tanto as avaliações elaboradas localmente nos diversos países quanto os programas avaliativos internacionais têm como meta fundamental a melhoria da qualidade da educação, ao menos no plano discursivo, o que consiste em seu aspecto comum, ainda que a concepção de qualidade varie de local para local. Há, por exemplo, países em que a qualidade é compreendida como acesso à escolaridade básica, enquanto outros já resolveram as questões de acesso e buscam

³ Em 2019, na segunda edição do Pasec, o número de países já havia subido para 15: Benin, Burquina Fasso, Burundi, Camarões, Congo, Costa do Marfim, Gabão, Guiné, Madagascar, Mali, Nigéria, República Democrática do Congo, Senegal, Chade e Togo.

bons parâmetros de aprendizagem para os estudantes, sendo a qualidade, então, entendida como alcance a padrões de desempenho em disciplinas escolares.

O que parece diferir os países é a cultura que se cria em torno dessas avaliações, os âmbitos das redes de ensino e dos sistemas educativos sobre os quais elas incidem, o nível de coleta de dados a que chegam e o uso feito dos resultados, ora sendo vistas como ferramentas utilizadas no gerenciamento de políticas e programas educacionais no nível do sistema educativo, ora com propostas que buscam reverberar no nível das escolas. Segundo Brink (2020), diferentes tipos de desenhos avaliativos servem a diferentes propósitos. Avaliações censitárias, por exemplo, adequadas a sistemas menores, serviriam para avaliar a “saúde” educacional de cada país e analisar diferenças entre grupos de indivíduos, enquanto as amostras poderiam ser usadas para o monitoramento do desenvolvimento das aprendizagens, para traçar caminhos para o futuro.

Ainda que se observe a expansão do uso dessas avaliações, com foco no desempenho dos estudantes, na determinação de políticas públicas educacionais, o que parece diferir, no Brasil, em relação a algumas iniciativas internacionais, é o uso majoritário de desenhos censitários e o crescente apelo aos resultados obtidos para a gestão educacional em nível meso e micro, dentro e fora das salas de aula.

Cabe lembrar que a proposição de avaliações de rendimento de alunos em território nacional começou a se desenvolver a partir do início dos anos 1990, primeiramente em nível federal (Gatti, 2013) e, posteriormente, expandindo-se para os estados, sendo possível observar diversas alterações quanto aos seus objetivos e desenhos metodológicos (Bonamino, 2013) ao longo do tempo. Inicialmente, a expansão dessas avaliações nos estados se deu de forma irregular. No entanto, após a proposição do Índice de Desenvolvimento da Educação Básica (IDEB⁴) (Bauer; Horta Neto, 2018), nota-se a expansão de avaliações próprias no âmbito estadual: até o final da década passada, 21 dos 27 estados subnacionais tinham sistemas próprios de avaliação. Mais recentemente, além das propostas estaduais, passou-se a observar o surgimento de propostas municipais de avaliação, como será discutido posteriormente. Além disso, houve o movimento de incorporação de avaliações de novas etapas de ensino ao Sistema de Avaliação da Educação Básica (Saeb) e do componente curricular de Ciências, além das já existentes aferições para os alunos do 5º e 9º anos do ensino fundamental e do 3º ano do ensino médio.

Inicialmente, no Brasil, as avaliações externas justificaram-se pela necessidade de obtenção de informações que pudessem ser úteis para subsidiar a tomada de decisões nas esferas centrais (Pestana, 2013; Gatti, 2013), assumindo uma função diagnóstica, sem que houvesse consequências diretas para as escolas, seus atores e as práticas pedagógicas nelas realizadas. Entretanto, desde a primeira década do século XXI, com a expansão das propostas e as mudanças técnicas introduzidas nos desenhos avaliativos das principais iniciativas, observam-se modificações também nos objetivos e funções subjacentes às avaliações, com a tendência de se atrelar consequências fortes ou brandas (Brooke, 2005, 2006; Andrade, 2008), “simbólicas” ou “materiais” (Bonamino; Sousa,

⁴ O IDEB é calculado a partir de dois componentes: taxa de rendimento escolar (aprovação) e médias de desempenho nos exames padronizados aplicados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Os índices de aprovação são obtidos a partir do Censo Escolar e as médias de desempenho utilizadas são as da Prova Brasil (para escolas e municípios) e do SAEB (no caso dos estados e nacional), todos produzidos pelo INEP.

2012), aos resultados obtidos, bem como um forte apelo ao seu caráter indutor, seja por meio do estabelecimento de padrões e níveis desejáveis de desempenho, seja pelo fortalecimento de sistema de metas e indicadores de desempenho que permitem o acompanhamento e a evolução dos resultados obtidos.

O estabelecimento desse *modus operandi* tem levado a várias críticas sobre a excessiva valoração de informações quantitativas em detrimento das informações sobre os processos educativos que ocorrem nos contextos que geram tais resultados.

No que diz respeito aos resultados de desempenho em provas e ao IDEB, tem-se tratado pouco dos aspectos técnicos propriamente ditos e observa-se uma discussão ainda tímida sobre a validade e a fidedignidade dos dados coletados, os erros de medida e os limites e potencialidades do estabelecimento de metas, seus significados e outros fatores que seriam importantes para a tomada de decisões bem-informada. Percebe-se, inclusive, que os aspectos técnicos são pouco reportados nos relatórios produzidos. Além disso, na divulgação e uso dos resultados no cotidiano da gestão pública das secretarias de educação e das escolas, é comum ver ações que parecem desconsiderar que os indicadores obtidos a partir das avaliações são um retrato de uma realidade mais complexa e, portanto, não devem ser confundidos com essa mesma realidade nem assumir o *status* de verdade absoluta sobre o desenvolvimento dos alunos e dos sistemas educativos, como ensina Paulo Jannuzzi (2001).

Nas próximas seções será ilustrada a ênfase em aspectos mensuráveis dos resultados educacionais que vem sendo dada nos municípios brasileiros a partir de dados obtidos em pesquisa que buscou mapear características das avaliações existentes nesses entes federados. O estudo realizado e suas trilhas metodológicas serão apresentados a seguir.

Procedimentos do estudo sobre a avaliação educacional nos municípios brasileiros⁵

O mapeamento das iniciativas municipais de avaliação existentes no território brasileiro ocorreu por meio de *survey*, aplicado com o auxílio da ferramenta *on-line Survey Monkey*, com os objetivos de: (1) obter informações sobre a existência ou não de avaliações externas próprias dos municípios, (2) identificar as razões que motivaram sua criação e (3) caracterizar seus delineamentos metodológicos e os usos feitos dos resultados para a gestão educacional. O questionário, estruturado em quatro dimensões e com 44 questões (abertas e fechadas), foi enviado para 5.532 dos 5.570 municípios brasileiros, ou seja, 99% do total⁶.

⁵ Os dados utilizados neste artigo foram obtidos por meio das pesquisas intituladas “Avaliação e gestão educacional em municípios brasileiros: mapeamento e caracterização das iniciativas em curso”, desenvolvida por meio de uma parceria entre a Fundação Carlos Chagas (FCC) e o Inep, sob a coordenação de Adriana Bauer e João Luis Horta Neto, com apoio financeiro do INEP e da FCC; e “Relações entre Avaliação e gestão educacional em municípios brasileiros: estudo em dez municípios da federação”, que recebeu apoio financeiro da FCC, do INEP e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). As considerações expressas nesse artigo são de exclusiva responsabilidade da autora, não expressando necessariamente as opiniões de todos aqueles que dela participaram.

⁶ No momento de realização da pesquisa não foi possível obter o endereço eletrônico de 38 municípios, motivo pelo qual não foram contatados para participar da pesquisa.

As dimensões estruturantes do instrumento foram: identificação do município; participação do município em avaliações propostas pelos âmbitos federal e estadual; avaliações existentes na rede municipal (podendo ser de alunos, de profissionais de ensino e gestores ou avaliação institucional) e caracterização dos usos das avaliações. A coleta dos dados ocorreu entre abril e julho de 2014 (Bauer; Horta Neto, 2018).

Dos 5.532 municípios contatados, 4.309 responderam à pesquisa, o que representa 77% do total de municípios brasileiros. Dentre esses, 1.573 afirmaram ter uma avaliação própria. Esse texto utiliza as respostas obtidas com esse subconjunto de municípios, que abrangem todas as cinco regiões geográficas do país e todos os estados brasileiros.

Destaca-se, primeiramente, a aparente não vinculação entre a existência de uma avaliação própria e as características do município – tamanho, número de habitantes e quantidade de escolas existentes nas redes de ensino –, uma vez que foram relatadas iniciativas de avaliação gestadas pelas secretarias de educação em municípios pequenos, médios e grandes, com redes de escolas maiores e menores, o que permite afirmar que esse tipo de avaliação exerce um aparente fascínio sobre os secretários de educação, que as percebem, em geral, como uma estratégia de gerenciamento da educação, mas sem que necessariamente haja um planejamento gerencial lógico previamente definido.

A ampliação da expansão das avaliações externas municipais a partir de 2010 pode ser analisada pela Tabela 1. Os dados evidenciam que seu processo de expansão começa, com raras exceções, nos anos 2000, consoante ao movimento de expansão que ocorreu internacionalmente. Apenas 34 municípios declararam ter avaliações próprias antes dos anos 2000 (até 1999); destes, 15 afirmam que o ano de início desse tipo de avaliação coincide com o ano do início do processo de municipalização das escolas no município⁷.

Tabela 1: Distribuição dos municípios respondentes

Ano em que teve início a iniciativa municipal de avaliação	Nº	%
Anterior a 1990	2	0,1
Entre 1990 e 1999	32	2,0
Entre 2000 e 2009	449	28,5
Entre 2010 e 2013	794	50,5
Janeiro a julho de 2014	106	6,7
Respostas ausentes	190	12,1
Total	1.573	100

Fonte: Elaboração da autora com base em dados da pesquisa.

Cabe comentar que esse processo de expansão não parece estar ligado necessariamente à busca de alternativas, no âmbito dos municípios, à lógica que preside as avaliações já existentes nos âmbitos federal e estadual, visto que a maioria dos municípios que declara utilizar avaliação

⁷ Buscar compreender a natureza dessas iniciativas e sua relação com o processo de gestão da rede educativa fugiu aos limites da pesquisa, ainda que fosse uma questão bastante pertinente para auxiliar na compreensão sobre como as redes de ensino compreendem e se apropriam da avaliação externa.

externa própria participa também de outras iniciativas avaliativas, propostas pelo governo federal ou estadual, de forma concomitante à proposição local, em um movimento acumulativo. Tal fato é evidenciado quando, dos 1.573 municípios que afirmam ter uma sistemática própria de avaliação, 96,8% atestam participar também da Prova Brasil, 99% da Provinha Brasil, 97,4% da Avaliação Nacional da Alfabetização e 98,3% de avaliações organizadas pelo respectivo governo estadual.

Além disso, dentre os tipos possíveis de avaliação externa conduzidas pela gestão municipal de educação, sobressaem-se as iniciativas de avaliação em larga escala com foco no rendimento dos alunos, ressaltada por 1.302 municípios, ao passo que propostas de avaliação institucional são declaradas por 977 respondentes, de avaliação de profissionais por 624 e outros tipos de avaliação por 99 representantes das municipalidades consultadas. Observa-se que a proposição de uma avaliação externa e em larga escala própria, com foco nos alunos, em geral é vista como um complemento ao que já existe, e não uma substituição ou uma alternativa, gerando a já mencionada sobreposição de testes nas redes de ensino.

Segundo os depoimentos coletados nas secretarias municipais de educação desses municípios, a proposição de iniciativas próprias de avaliação em larga escala justifica-se pela preocupação com a qualidade dos serviços educacionais ofertados, e pela demanda para o melhor gerenciamento (entendido como eficaz e efetivo) dos recursos disponíveis e de políticas e programas educacionais, considerando-se os dados por elas produzidos como fontes de informação para o seu aprimoramento (Bauer; Horta Neto, 2018).

Soma-se a esses fatores a ideia de diagnóstico e monitoramento do ensino e das aprendizagens, bastante presente nos discursos dos respondentes, bem como a visão da avaliação como instrumento para o planejamento de intervenções pedagógicas, controle e adequação curricular, proposição de atividades de formação continuada, avaliação do trabalho docente e mobilização de professores e alunos para a atividade educacional.

As respostas revelam ainda a crença de que a avaliação de alunos pode induzir à melhoria de índices educacionais (fluxo, evasão e repetência), assim como colaborar com o alcance de resultados e metas estipuladas tanto pelo nível federal quanto pelas próprias administrações municipais. Notam-se, nos depoimentos feitos nas questões abertas, referências explícitas ao aumento do IDEB, ou seja, marca-se uma preocupação com aspectos quantitativos dos resultados da avaliação.

Discussão dos resultados

Como dito anteriormente, dos 1.573 municípios que afirmavam ter uma avaliação própria na metade da década passada, 1.302 atestam que o foco da proposta avaliativa são os alunos, por meio da aferição de seu desempenho escolar. A esses municípios foi apresentada uma questão fechada, com uma listagem de oito objetivos⁸ que se procurava atingir por meio da iniciativa própria de avaliação, sendo que os municípios poderiam escolher as alternativas que melhor condiziam com a sua realidade.

⁸ Os objetivos que foram propostos como alternativas da questão foram inspirados em discussões presentes na literatura atinente, em especial Bauer (2010) e Brooke e Cunha (2014).

A Tabela 2 apresenta as maiores incidências de resposta, em ordem decrescente. Para além da afirmativa da preocupação com a aprendizagem dos alunos, apontada por 71,8% dos respondentes, os principais objetivos destacados foram: melhorar o IDEB e definir as prioridades de gestão, ou seja, uma preocupação com a elevação de uma medida.

Tabela 2: Objetivos das avaliações próprias com foco no desempenho dos alunos, propostas pelos municípios brasileiros

Objetivos das avaliações com foco nos alunos	Nº	%
Melhorar a aprendizagem dos alunos	1.130	71,8
Melhorar o IDEB	528	33,6
Definir prioridades para a gestão da educação municipal	524	33,3
Obter informações para a formação continuada	454	28,9
Reduzir a taxa de repetência e/ou de abandono	431	27,4
Aprimorar os processos de gestão da rede municipal	239	15,2
Divulgar os resultados das escolas para os pais e a comunidade	73	4,6
Obter informações para a premiação/bonificação de escolas	45	2,9

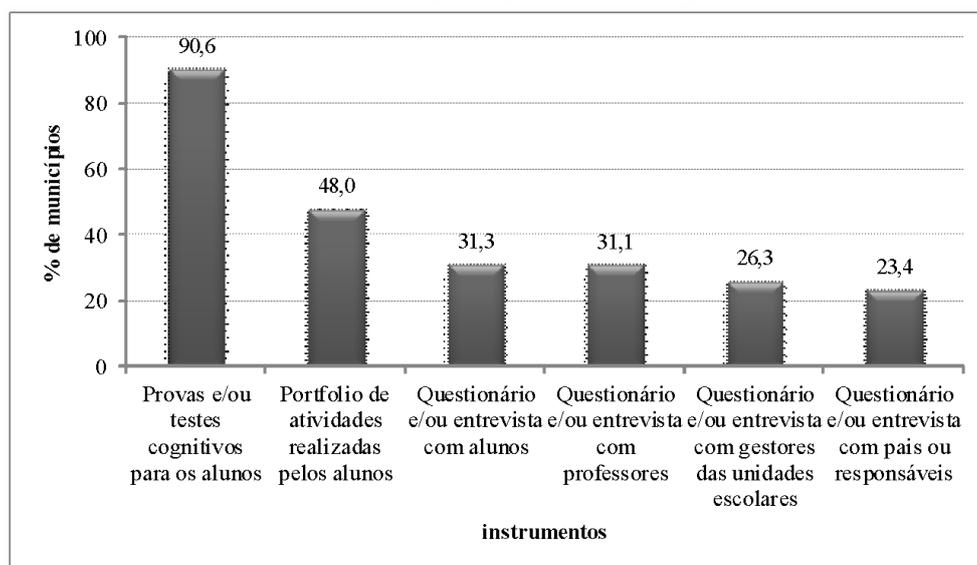
Fonte: Elaborado pela autora com base nos dados do survey.

A fim de compreender os desenhos metodológicos propostos nas avaliações municipais voltadas aos alunos, foi apresentada uma série de questões que procuravam captar informações sobre as disciplinas e séries avaliadas, as formas de análise dos resultados, a periodicidade da avaliação, os tipos de itens produzidos, dentre outras informações.

Ainda que, mais recentemente, um movimento de ampliação das disciplinas e séries avaliadas seja observado, inclusive no âmbito municipal, normalmente o foco das diversas avaliações recai sobre os mesmos aspectos: conteúdos cognitivos, especialmente em Língua Portuguesa e Matemática, séries terminais de cada ciclo ou etapa de ensino. Resultados encontrados em estudo realizado por Bauer e Horta Neto (2018), cujo foco foi o mapeamento de desenhos usuais em propostas municipais de avaliação em larga escala, mostram que 48,1% dos municípios já buscavam aferir o conhecimento dos alunos em Ciências (antes mesmo da introdução desses componentes em nível federal), 45,1% contemplavam o acompanhamento dos resultados obtidos em História e 44,7% observavam o desempenho em Geografia. Outros componentes curriculares também considerados pelos municípios em suas aferições próprias foram: Artes (32,6%), Educação Física (32,0%) e Língua Estrangeira (29%), ainda que pensar uma avaliação externa para esses componentes curriculares possa ser um desafio e até causar estranhamento. Em relação às disciplinas comumente avaliadas, 93,8% dos municípios afirmaram avaliar Leitura e Interpretação de texto, 76,9% aferem conhecimentos de Gramática e 72,8% buscam acompanhar o desenvolvimento de habilidades de escrita, por meio de provas de Redação. 91,7% avaliam Matemática.

Em relação aos instrumentos de avaliação utilizados, os dados coletados estão organizados no Gráfico 1. Pela sua análise, pode-se dizer que as avaliações de alunos propostas nas redes municipais tendem a utilizar provas e/ou testes cognitivos para aferir seu desempenho, podendo ou não estar associadas a questionários ou entrevistas com diversos atores (gestores, professores, alunos, pais ou responsáveis).

Gráfico 1: Proporção de municípios que utilizam cada um dos instrumentos citados, dentre as redes municipais que têm avaliação de alunos



Fonte: Bauer e Horta Neto, 2018.

A pergunta direcionada a captar as características dos modelos de processamento e análise dos dados, obteve resposta de 1.096 dos 1.302 municípios que declararam ter uma avaliação externa voltada aos alunos. As indicações obtidas permitem afirmar que, tendencialmente, os resultados são analisados por meio de tabulações e/ou elaboração de gráficos, organização de sínteses de resultados e das informações em bancos de dados. Em menor número estão os respondentes que declararam que os resultados são analisados coletivamente, por meio de reuniões, encontros e seminários (112 indicações). Há, ainda, registros da realização de comparação dos dados com as matrizes ou outras metas definidas (90 registros). Duas capitais informaram a utilização da Teoria da Resposta ao Item (TRI) como metodologia de análise.

A impressão que se tem, pela leitura dos depoimentos, é de que muitas das avaliações não contam com requisitos técnicos básicos que garantam sua validade e fidedignidade. A análise e divulgação dos dados não parece incorporar metodologias estatísticas mais aprimoradas, sendo mais recorrente a menção à produção de gráficos e tabelas como recurso de organização e interpretação dos dados, conforme ilustram os depoimentos a seguir.

Um caderno contendo 20 itens de Língua Portuguesa e 20 itens de Matemática. A escola recebe o kit contendo os cadernos de questões, um documento com orientações de aplicação e devolução. A correção é feita pelos técnicos da Secretaria Municipal de Educação e os resultados são catalogados e representados em gráficos (Bauer; Horta Neto, 2018, p. 86).

A avaliação diagnóstica é realizada semestralmente com os alunos do Ensino Fundamental. Os professores recebem da Secretaria Municipal uma chave de correção das provas que são aplicadas para os alunos. Os resultados são encaminhados pela coordenação da unidade escolar e os dados tabulados na Secretaria para análise e posteriormente, para divulgação (Bauer; Horta Neto, 2018, p. 87).

Não se quer defender, aqui, que as redes municipais de ensino recorram aos especialistas em medida educacional para produzir dados a fim de embasar o gerenciamento educacional. A questão da qualidade técnica é polêmica, e há acúmulo de uma vasta literatura crítica à ênfase nos aspectos técnicos da avaliação em detrimento dos políticos, ideológicos e pedagógicos, subjacentes aos modelos de aferição e análise de dados de rendimento dos alunos.

O que se questiona é a ênfase que se tem sido dada no monitoramento das medidas de proficiência como forma de melhoria da qualidade educacional, que induz inclusive municípios pequenos, muitas vezes com uma estrutura financeira, administrativa e técnica reduzida, a adotar a avaliação em larga escala de rendimento dos alunos como ferramenta de gestão educacional, mesmo sem ter a capacidade técnica para aprofundar as análises dos dados produzidos. Ao que parece, assume-se a ideia de que monitorar as aprendizagens é importante, sem que se discuta, com o devido aprofundamento, a natureza dos instrumentos utilizados, sua validade e a fidedignidade dos resultados obtidos. Aliás, conceitos como validade, fidedignidade, estatística do teste, entre outros, que ilustram uma preocupação com a qualidade do que se está “medindo”, ou da informação que se está coletando por meio das iniciativas, são ausentes nas respostas obtidas no *survey*.

Tal fato não impede, contudo, que sejam tomadas decisões importantes sobre o sistema educacional: os dados obtidos são usados, segundo os depoimentos, para aprimorar o currículo, incidir sobre a formação dos professores, etc. Ainda que os usos “com consequências fortes” (Carnoy; Loeb, 2002; Brooke, 2006) sejam residuais, e muitos depoimentos atestem que a lógica que preside a proposição dessas avaliações é formativa e diagnóstica, seria importante um olhar mais crítico acerca das possibilidades e limites dos dados produzidos.

Além das propostas de testar os conhecimentos dos alunos, algumas das iniciativas municipais de avaliação originam indicadores quantitativos, construídos à semelhança do IDEB, indicador nacional já comentado anteriormente. Dos 1.573 municípios que afirmaram elaborar suas avaliações, 231 declararam ter, há cerca de dez anos, um indicador educacional próprio, o que corresponde a 14,7% do total de respostas válidas, enquanto outros afirmam utilizar sistemática de produção de indicadores de qualidade da educação próxima à proposta em âmbito federal.

Ao citar os componentes considerados como indicadores para a gestão local, a maioria dos gestores faz referência aos indicadores de resultados, como medidas de fluxo (taxas de aprovação, reprovação, abandono e evasão escolar) e proficiência (médias obtidas nas diversas testagens/avaliações externas), consideradas pelos depoentes preocupações centrais na área de educação. Alguns gestores afirmam estar atentos à necessidade de produção de indicadores de insumo e de processo, exemplificados a seguir, com aspectos coletados dentre as respostas ao *survey*:

- Indicadores de insumos: infraestrutura, gasto por aluno, investimento em formação docente, salário docente;
- Indicadores de processo: informações sobre a organização didático-pedagógica, planejamento institucional, informações sobre atendimento aos educandos, processo de ensino e aprendizagem, organização do ambiente físico escolar, prática pedagógica, gestão escolar democrática, condições de trabalho dos profissionais da escola e taxas de acesso e permanência dos alunos na escola;

- Indicadores de resultados: fluxo (taxas de aprovação, reprovação, abandono e evasão escolar), taxa de distorção idade-série, taxa de alfabetização e desempenho de estudantes em exames padronizados (resultados obtidos nas diversas testagens/avaliações externas), etc.

No entanto, ainda que os depoimentos apontem para vários tipos de indicadores, a ênfase recai sobre os de resultados: das 199 respostas válidas obtidas na questão aberta que solicitava aos respondentes explicarem os componentes do indicador proposto pelo município, 75,4% referiram-se a indicadores de resultados e 14,6% a indicadores de processos. Apenas duas respostas mencionaram indicadores de insumos e diversas respostas não se referiam ao objeto da questão, sendo desconsideradas da análise. Dentre as respostas que fizeram referência a aspectos considerados pela rede municipal de ensino para avaliação da qualidade do ensino, a maioria das menções versa sobre aspectos relacionados aos alunos; em segundo lugar, menciona-se itens relativos às escolas; sendo que os relativos aos professores são praticamente desconsiderados (apenas cinco respostas).

Nos municípios estudados, os indicadores parecem ser realmente utilizados como subsídio para tomada de decisão e como orientadores dos processos de gestão. Acompanhando essa apropriação da ideia de ter indicadores educacionais para monitoramento da educação, que provavelmente sofreu influência do IDEB, observa-se a tendência de se definir metas – entendidas como propulsoras de qualidade – para balizar a atuação das escolas, conforme ilustrado pelos depoimentos acerca dos propósitos da elaboração dos indicadores educacionais municipais:

“Atingir as metas propostas pela rede municipal de ensino”.

“Com base nos resultados do IDEB e melhoria na alfabetização foram criados indicadores e metas a serem alcançadas”.

“Desempenho e participação mediante metas estipuladas em portaria pela Secretaria da Educação”.

“Estabelecer metas para alcançar o nível de proficiência em todos os alunos da rede municipal em 100% até o ano de 2016. Projeção de: 30% no ano de 2014; 50% no ano de 2015; 20% no ano de 2016”.

“Foram criadas/estabelecidas sete metas de aprendizagens a partir do mapa e planejamento estratégico da Secretaria Municipal de Educação”.

“Traçar as metas da qualidade educacional da Rede Municipal, ampliar a Formação Continuada, instituir a avaliação interna no município obedecendo aos padrões, normas e critérios das Avaliações do Governo Federal e Estadual, para que os estudantes se apropriem dessas normas estabelecidas na realização das avaliações” (Bauer; Horta Neto, 2018, p. 118).

Ao passo que se observa muita preocupação e discussão em torno das metas numéricas às quais os municípios e as escolas devem chegar, pouca atenção parece ser dada ao que significam os resultados obtidos e quais pistas eles podem ou não fornecer para o alcance de níveis desejáveis de aprendizagem dos alunos. As próprias escalas de desempenho têm sido pouco debatidas no âmbito dos órgãos propositores de políticas, ainda que, nos depoimentos dos gestores municipais, encontrem-se referências às matrizes e escalas do Saeb, referendando-as.

Se pouco debate tem sido observado, no meio acadêmico, sobre o que significa, em termos de aprendizagem, a elevação do IDEB – indicador educacional mais conhecido –, não é de se espantar que essa ausência esteja também presente no discurso dos gestores. Afinal, além da diminuição de

índices de reprovação, não seria desejável que os alunos desenvolvessem habilidades e conhecimentos mais complexos nos componentes curriculares avaliados? Os sistemas de ensino em geral e, mais especificamente, as escolas, têm discutido estratégias para que os alunos aprendam mais e melhor, objetivo crucial ao processo educativo? Ou têm apenas tentado responder irrefletidamente às demandas de aumento dos índices e indicadores valorizados pelo discurso das políticas? Tomando como base a pesquisa, a impressão que se tem é que a tendência é responder positivamente à última questão aqui levantada: há muita preocupação com a busca por resultados (o que não é um fato ruim por si só), mas sem que se busque entender o real significado do resultado em termos de aprendizagem e de processos ensejados no âmbito escolar. É essa busca desenfreada e irrefletida por números que se considera, neste texto, um surto quantofrênico que não necessariamente contribui para o aprimoramento da qualidade das aprendizagens e, conseqüentemente, da educação.

A relevância que hoje se atribui ao IDEB – em especial ao seu componente de avaliação de rendimento discente – no âmbito municipal, levando, inclusive, a propostas de avaliação próprias e indicadores e à adoção de simulados e currículos relacionados com os conteúdos cobrados nos testes nacionais, parece residir no entendimento comum de que o que se avalia é a própria qualidade do ensino ou da educação, visto que é assim que o indicador é compreendido e foi, equivocadamente, propagandeado para a sociedade.

No entanto, não se encontram evidências, nos depoimentos dos respondentes, de que a preocupação com a métrica é acompanhada pelo entendimento dos desafios existentes para a melhoria das aprendizagens, e das situações concretas para superá-los, dado o contexto em que os resultados são produzidos, ainda que a preocupação com esse aprimoramento faça parte do discurso.

Medidas de qualidade do ensino: surto quantofrênico⁹ ou necessidade de gestão? Proposta de uma agenda de pesquisas

No início dos anos 2000, Iaies (2003) já alertava para a necessidade de se recuperar o sentido de avaliar os sistemas educacionais e os vínculos existentes entre esse tipo de avaliação e as decisões político-filosóficas dos gestores. Para o autor, as avaliações estavam se configurando em modelos cuja prioridade era medir o rendimento dos alunos e não avaliar¹⁰ o sistema como um todo.

A inexistência de clareza sobre os objetivos da avaliação de sistema e a dificuldade de definir e consensuar padrões de qualidade claros, que permitissem comparar longitudinalmente os resultados obtidos e que pudessem ser utilizados para analisar as possíveis mudanças operadas a partir das políticas e programas implantados, fazia, segundo o Iaies, com que as preocupações recaíssem sobre os resultados das provas e sua dimensão técnica (quando muito), desconsiderando-

⁹ Alusão ao termo utilizado pelo sociólogo russo Sorokin, em meados do século XX, para explicar o excesso de apelo aos aspectos quantitativos e mensuráveis dos fenômenos, determinado pelas ciências psicológicas e sociais, o que reduzia o conhecimento ao que podia ser observado e medido.

¹⁰ A diferenciação entre avaliação e medida tem sido alvo de atenção de inúmeros pesquisadores da área e pode ser encontrada nos trabalhos de Vianna (2000).

se as análises contextuais que permitiriam um melhor entendimento da situação educacional e uma intervenção mais efetiva. Como explica o autor, “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” (Iaies, 2003, p. 18).

Haveria, assim, 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). Segundo o autor:

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 (Iaies, 2003, p. 20-21).

A análise da configuração de propostas de avaliação de redes, ensejadas por diversos municípios brasileiros, permite afirmar a atualidade do debate proposto por Iaies há vinte anos, quando os modelos de avaliação de redes estavam apenas se expandindo. Além disso, permite questionar se os desenhos regionais de avaliação estão realmente orientados por organismos multilaterais ou se há uma lógica própria de gerenciamento sendo ressignificada e construída nesse nível federado, que é o argumento central neste artigo: o Brasil teria desenvolvido uma cultura própria na implementação de avaliações externas e em larga escala, não sendo refém necessariamente das indicações de organismos multilaterais. Não se quer negar que também exista um surto quantofrênico que tem atingido diversos países, estados e distritos educacionais, muitas vezes gerando uma sobreposição de avaliações em um mesmo contexto, desenhadas para aferir, inclusive, os mesmos aspectos.

Faz-se, assim, necessário retomar a discussão acerca do papel que as avaliações em larga escala têm assumido na sociedade (Broadfoot, 1996), e mais especificamente, na gestão da educação e da qualidade do ensino (Pestana, 2013), principalmente porque tais iniciativas se encontram, cada vez mais, relacionadas com políticas e programas educacionais.

A falta de clareza acerca das finalidades, possibilidades e limites de uma avaliação em larga escala, como apoio ao gerenciamento do sistema educativo, talvez possa estar relacionada à aparente falta de efetividade das propostas em si. Afinal, se avaliar e fornecer medidas precisas sobre o desempenho dos alunos fosse estratégia suficiente para superar as dificuldades de ensino das escolas, não seria necessário o apelo às políticas de alto impacto, baseadas quase sempre nos resultados obtidos pelos alunos.

O que a quantofrenia que se observa atualmente no Brasil parece desconsiderar é que a medida, mesmo quando de qualidade técnica inegável – como é o caso das provas federais e de algumas estaduais e municipais –, não substitui as ações que precisam ser realizadas para que se proceda à uma reforma educativa mais ampla, cuja aposta seria o aprimoramento da educação em seus diversos aspectos, e não somente em aspectos quantificáveis do desempenho cognitivo dos alunos.

Outro aspecto que pode ser observado nas iniciativas municipais analisadas, relacionado ao anteriormente mencionado, é que a preocupação central das equipes técnicas parece ser a de produzir informação, obtida por meio dos instrumentos utilizados, em detrimento de explorar o potencial das informações para a mudança educativa. Mesmo nas propostas municipais em que há maior preocupação com uma análise refletida dos resultados e em seu uso para a tomada de decisão, a elaboração e divulgação desse tipo de análise parece ainda rudimentar, sendo pouco aproveitada para a reelaboração de estratégias pedagógicas ou reorientação de políticas educacionais. Ou seja, o forte apelo aos resultados, indicadores e estabelecimento de metas de aprendizagem parece assumir características próprias no contexto brasileiro: há um discurso pedagógico como justificativa para a valorização das provas e indicadores e inclusive para um trabalho de intervenção mais pontual da gestão nas salas de aula diretamente, ainda que não se encontre evidências de que as informações quantitativas sejam compreendidas em seu significado pedagógico.

Esse descompasso entre a produção da informação e sua análise e utilização para a melhoria da educação já havia sido apontada por Iaies, nos anos 2000 (Iaies, 2003) e também é identificado em diversos outros trabalhos (Bauer, 2010).

Por outro lado, nota-se a preocupação, por parte dos gestores, com a obtenção de informações que possam ser utilizadas em nível local, ou seja, nas escolas. Poucas têm sido as pesquisas que se preocupam em analisar como os resultados obtidos por meio de avaliações externas próprias podem ser utilizados positivamente no âmbito local, quer pelas secretarias de educação, quer pelas escolas. Nesse sentido, ainda que as avaliações produzidas localmente possam estar mais comprometidas em aquilatar a realidade do trabalho realizado pelas escolas, sabe-se pouco ainda sobre as características dessas avaliações locais e os mecanismos que ensejam nas escolas.

Faz-se necessário discutir também os modelos estatísticos utilizados nas avaliações (desenho, escalas, matrizes de referência), as metodologias de análise dos dados (TRI *versus* Teoria Clássica) e de procedimentos para análise e utilização dos dados obtidos e, em que medida, os resultados, como apresentados, fazem sentido para os gestores e professores, que deles devem se apropriar. Há pesquisadores, inclusive, que apontam para os limites dos dados quantitativos que vêm sendo produzidos pelas várias testagens de alunos existentes, principalmente no que se refere ao seu uso pedagógico e ao seu limite para apoiar o trabalho dos professores, que não se apropriam de seus resultados (Bauer, 2010).

Gatti, por exemplo, alerta que as avaliações em larga escala têm produzido uma numerologia vazia, não compreensível para o professor e o gestor escolar:

Minha opinião é que, com a avaliação, caímos numa numerologia que está se tornando vazia. Os resultados da Teoria da Resposta ao Item são traduzíveis numa escala – 125, 250, etc. – que não traz para as escolas e os professores uma informação facilmente compreensível. Por exemplo, um aluno que chega a uma escala 125 (sic) sabe escrever um bilhete, com algumas dificuldades da língua. Mas isso não quer dizer nada para o professor alfabetizador. O que está por trás disso? Qual o processo que está embutido nisso? A Teoria da Resposta ao Item – do jeito que a escala é feita e a matriz é concebida – não dá esse tipo de resposta. Não foi feita para isso. A prova é uma prova “aleatória”. É algo concebido para avaliar, em tese, um traço latente de potencialidade de aprendizagem. É um traço potencial. É muito difícil para as escolas compreenderem e lidarem com esse indicador, mesmo transformando em uma

escala de 0 a 10 ou de 0 a 100, mais próxima da compreensão geral. Não informa sobre os desempenhos específicos e processos de aprendizagem. As informações são genéricas (Gatti, 2011, p. 10).

Ainda que os benefícios para as políticas públicas oriundos do uso de teorias de análise como a TRI não possam ser desconsiderados – visto que ela permite a realização de comparações entre populações ao longo do tempo e que responderam provas diferentes (desde que tenham itens comuns), permitindo analisar o desenvolvimento e evolução dos indivíduos avaliados em determinado traço latente (Valle, 2000) –, ela é pouco compreendida e precisa de muito investimento em formação e materiais de apoio para que os professores possam se apropriar dos seus resultados a favor das aprendizagens.

Outro aspecto que merece reflexão aprofundada refere-se à seleção dos conteúdos e habilidades que embasam as avaliações, bem como a definição de níveis ou padrões de desempenho, tanto em seu caráter técnico (o que medir e como medir) quanto pedagógico. A preocupação com as relações que se estabelecem entre avaliação e currículo e a discussão sobre o que é medido e o currículo mais geral, que constituem o projeto formativo mais amplo a que todos os alunos devem ter acesso, aparecem de forma secundarizada nos discursos dos gestores registrados por meio do *survey*.

Até meados da segunda década deste século, havia no Brasil um movimento que buscava formular e disseminar as Matrizes de Referência da Avaliação, que passaram a ser consideradas como subsídio ao trabalho pedagógico das escolas, sem que tal formulação tenha sido acompanhada de um amplo debate sobre os propósitos da educação, que, em tese, definiriam os currículos e posteriormente a discussão sobre as aprendizagens básicas cuja aquisição se gostaria de monitorar (Bauer; Gatti; Tavares, 2013). Não foram encontradas na literatura evidências de que tal lógica acompanhou também as avaliações que ocorrem em outros países, a partir de orientações de organismos multilaterais. Mesmo em relação ao Programa Internacional de Avaliação de Estudantes (PISA), que tem sido considerado um instrumento de indução de práticas pedagógicas para os países participantes da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), cabe questionar essa tendência crítica, visto que as matrizes curriculares desse programa internacional não são amplamente conhecidas e divulgadas.

De todo modo, até a aprovação da Base Nacional Curricular Comum (BNCC), em 2017, os debates sobre a seleção dos conteúdos e habilidades que embasam as avaliações parecem ter sido relegados a um segundo plano, em detrimento das discussões sobre padrões, metas e indicadores. Cabe, nesse sentido, questionar: que visões de educação, de jovens e de mundo orientam as matrizes e parâmetros de avaliação que têm sido adotados nas redes de ensino do Brasil? Que conteúdos básicos se espera que todos os alunos aprendam, ou quais são os padrões mínimos a que se almeja, considerando-se as reais necessidades sociais e possibilidades de desenvolvimento dos alunos? Mesmo em tempo de implementação da primeira base curricular comum nacional¹¹, parece que tais reflexões ainda carecem de desenvolvimento em âmbito nacional.

¹¹ O Plano Nacional de Educação, de 2014, estabeleceu, dentre diversas estratégias, a necessidade de se estabelecer, até 2016, uma base nacional curricular comum, para servir de referência curricular a todos os entes da federação. Tal

Por fim, cabe observar que apesar de muitos pesquisadores se dedicarem ao estudo das avaliações de sistema ou testagens em larga escala (Bauer, 2010) e aos problemas delas decorrentes, essas reflexões parecem dialogar pouco com as necessidades e interesses de gestores e propositores de políticas educacionais, numa aparente desarticulação, ou distanciamento, entre o meio acadêmico e a realidade da gestão, que se mantêm isolados em seus respectivos campos de atuação.

Apesar da crescente expansão dessas avaliações, e de sua consolidação em nível mundial, cabe continuar a discussão técnica e política acerca desse objeto de estudo, a fim de desnaturalizar as práticas avaliativas vigentes e não deixar submergir os marcos políticos e os princípios gerenciais aos quais tais avaliações estavam atreladas inicialmente.

No Brasil, onde já se passaram mais de trinta anos da implantação dos primeiros sistemas de avaliação, percebe-se a urgência de se retomar os estudos já efetuados e fazer uma ampla revisão dos pressupostos que vêm norteando as ações avaliativas em todos os níveis. Questionamentos elaborados por membros da academia, mais distantes dos órgãos responsáveis pelo desenho e gerenciamento das avaliações, parecem ter pouca influência sobre os gestores e as equipes de avaliação que, salvo raras exceções, mantêm-se, ao menos no nível do discurso, desconhecedores dos limites dos usos dos resultados das avaliações para a consecução da melhoria educacional.

A partir do estudo relatado parcialmente neste texto, fica claro que o debate sobre avaliação 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 de fato para a melhoria da educação, o que exige uma análise que ultrapassa a comparação de resultados quantitativos sobre níveis de aprendizagem, e que considere também aspectos curriculares, de infraestrutura e de formação docente, entre outros.

Referências

AFONSO, Almerindo. *Avaliação educacional: regulação e emancipação*. São Paulo: Cortez, 2000.

ALLEN, Reg; ELKS, Phil; OUTHRED, Rachel; VARLY, Pierre. *Uganda's assessment system: a road map for enhancing assessment in education*. Oxford: Health & Education Advance & Resource Team, 2016.

ANDRADE, Eduardo de Carvalho. "School accountability" no Brasil: experiências e dificuldades. *Revista de Economia Política*, v. 28, n. 3, p. 443-453, 2008. <https://doi.org/10.1590/S0101-31572008000300005>

BALL, Stephen. Performatividade, privatização e o pós Estado. *Educação e Sociedade*, v. 25, n. 89, p. 1105-1126, 2004. <https://doi.org/10.1590/S0101-73302004000400002>

BARFIELD, Thomas. *The dictionary of anthropology*. Hiboken: Wiley-Blackwell, 1997.

base, aprovada em dezembro de 2017, estabelece uma parte comum de conhecimentos e habilidades, denominadas de direitos de aprendizagem, aos quais todos os alunos devem ter acesso, além de princípios norteadores para a elaboração de uma parte diversificada.

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

<https://doi.org/10.24109/2176-6681.rbep.91i228.576>

BAUER, Adriana; GATTI, Bernardete Angelina; TAVARES, Marialva Rossi. *Ciclo de debates*. Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origem e pressupostos. Florianópolis: Insular, 2013.

BAUER, Adriana; ALAVARSE, Ocimar Munhoz; OLIVEIRA, Romualdo Portela de. Avaliações em larga escala: uma sistematização do debate. *Educação e Pesquisa*, v. 41, n. spe, p. 1367-1384, 2015.

<https://doi.org/10.1590/S1517-9702201508144607>

BAUER, Adriana; HORTA NETO, João Luiz. *Avaliação e gestão educacional em municípios brasileiros: mapeamento e caracterização das iniciativas em curso*. Relatório Final – Resultados do survey. Brasília: INEP, 2018. v. 2.

BENETT, Randy Elliot. *Reinventing Assessment. Speculations on the Future of Large-Scale Educational Testing*. A Policy Information Perspective. Princeton: Educational Testing Service, 1998.

BONAMINO, Alicia. Avaliação educacional no Brasil 25 anos depois: onde estamos? In: BAUER, Adriana; GATTI, Bernardete Angelina (Org.). *25 anos de avaliação de sistemas educacionais no Brasil: implicações nas redes de ensino, no currículo e na formação de professores*. Florianópolis: Insular, 2013. p. 43-60.

BONAMINO, Alicia; SOUSA, Sandra Zákia. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. *Educação e Pesquisa*, v. 38, n. 2, p. 373-388, 2012.

<https://doi.org/10.1590/S1517-97022012005000006>

BRINK, Satya. *Effective use of large-scale learning assessments: to improve learning outcomes and education system reform*. Dakar: Unesco, 2020. <https://unesdoc.unesco.org/ark:/48223/pf0000375333>

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

BROOKE, Nigel; CUNHA, Maria Amália de Almeida. *Avaliação externa como instrumento da gestão educacional nos estados*. São Paulo: Fundação Victor Civita, 2014. Volume 4 - série Estudos & Pesquisas Educacionais.

BROOKE, Nigel. *Accountability educativa en Brasil: una visión general*. In: SEMINARIO INTERNACIONAL ACCOUNTABILITY EDUCACIONAL: POSIBILIDADES Y DESAFÍOS PARA AMÉRICA LATINA A PARTIR DE LA EXPERIENCIA INTERNACIONAL, 2005, Santiago do Chile. *Programa de Promoción de la Reforma Educativa en América Latina y el Caribe*. Santiago de Chile: Unesco, 2005.

BROOKE, Nigel. O futuro das políticas de responsabilização no Brasil. *Cadernos de Pesquisa*, v. 36, n. 128, p. 377-401, 2006. <https://doi.org/10.1590/S0100-15742006000200006>

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

<https://doi.org/10.3102/01623737024004305>

CHINAPAH, Vinayagum. Monitoring Learning Achievement (MLA) Project in Africa. In: ADEA BIENNIAL MEETING, 2003, Grand Baie. *Proceedings [...]*. Grand Baie: Association for the Development of Education in Africa, 2003. https://biennale.adeanet.org/2003/papers/2Ac_MLA_ENG_final.pdf

CLARKE, Marguerite; LIBERMAN, Julia; RAMIREZ, Maria-Jose. Student Assessment. In: PATRINOS, Harry (Ed.). *Strengthening education quality in East Asia: SABER System Assessment and Benchmarking for Education Results*. Washington: The World Bank, 2012. p. 27-44.

EURYDICE. *National testing of pupils in Europe: Objectives, Organization and Use of Results* [online], 2009. <https://op.europa.eu/en/publication-detail/-/publication/dac32463-5f05-4782-898b-dd399e66efc6>

GATTI, Bernadette. Números vazios. In: BAUER, Adriana (Org.). *Avaliação Educacional*. Escola Pública (Especial Avaliação Educacional), São Paulo, 2011. p. 45-59.

GATTI, Bernardete. Possibilidades e fundamentos de avaliações em larga escala: primórdios e perspectivas contemporâneas. In: BAUER, Adriana; GATTI, Bernardete Angelina; TAVARES, Marialva Rossi (Orgs.). *Ciclo de debates*. Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origem e pressupostos. Florianópolis: Insular, 2013. p. 47-70.

HOADLEY, Ursula; MULLER, Johan. Visibility and differentiation: systemic testing in a developing country context. *The Curriculum Journal*, v. 27, n. 2, p. 272-290, 2016. <https://doi.org/10.1080/09585176.2015.1129982>

IAIES, Gustavo. Evaluar las evaluaciones. In: IAIES, Gustavo et al. *Evaluar las evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires: IIPE: UNESCO, 2003. p. 15-36.

JANNUZZI, Paulo. *Indicadores sociais no Brasil*. Campinas: Alínea, 2001.

KELLANGHAN, Thomas; GREANEY, Vicent. *Using assessment to improve the quality of education*. Paris: UNESCO, 2001. p. 19-20.

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 standardized assessment policies. *Revue Française de Pédagogie*, v. 169, p. 139-140, 2009. <https://doi.org/10.4000/rfp.1531>

PESTANA, Maria Inês de Sá. A experiência em avaliação de sistemas educacionais: em que avançamos? In: BAUER, Adriana; GATTI, Bernardete Angelina; TAVARES, Marialva Rossi. *Ciclo de debates*. Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origem e pressupostos. Florianópolis: Insular, 2013. p. 117-136.

REY, Olivier. The use of external assessments and the impact on education systems. In: STONEY, Sheila (Ed.). *Beyond Lisbon 2010: perspectives from research and development for education policy in Europe*. Slough: National Foundation for Educational Research, 2010. p. 137-157.

TOBIN, Mollie; LIETZ, Petra; NUGROHO, Dita (Anindita); VIVEKANANDAN, Ramya; NYAMKHUU, Tserennadmid. *Using large-scale assessments of students' learning to inform education policy: insights from the Asia-Pacific region*. Melbourne: ACER; Bangkok: Unesco, 2015.

VALLE, Raquel da Cunha. Teoria da resposta ao item. *Estudos em Avaliação Educacional*, n. 21, p. 7-92, 2000. <https://doi.org/10.18222/eae02120002225>

VIANNA, Heraldo Marelím. *Avaliação educacional: teoria, planejamento, modelos*. São Paulo: IBRASA, 2000.

WORLD BANK. *Mozambique: Student Assessment. Systems Approach for Better Education Results (SABER) country report*. Washington, 2009. <https://documents1.worldbank.org/curated/en/246021468321879999/pdf/>

ADRIANA BAUER

Doutora em Educação, Universidade de São Paulo (USP); Pesquisadora, Fundação Carlos Chagas (FCC), São Paulo, São Paulo, Brasil; professora, Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil.

APOIO/FINANCIAMENTO

FAPESP. Processo n. 14/17940-9.

DISPONIBILIDADE DE DADOS DE PESQUISA

Todos os dados foram gerados/analísados no presente artigo.

COMO CITAR ESTE ARTIGO

BAUER, Adriana. Avaliação em larga escala e estabelecimento de metas e indicadores educacionais: usos e abusos da quantofrenia. *Educar em Revista*, Curitiba, v. 40, e94095, 2024. <https://doi.org/10.1590/1984-0411.94095>

O presente artigo foi revisado por Andreza Silva. Após ter sido diagramado, foi submetido para validação do(s) autor(es) antes da publicação.

Recebido: 25/01/2024

Aprovado: 12/08/2024

Este é um artigo de acesso aberto distribuído nos termos de licença Creative Commons.

