
HOMOGENEIZAÇÃO CURRICULAR E O SISTEMA DE AVALIAÇÃO NACIONAL BRASILEIRO

O caso do estado do Rio de Janeiro*

Maria Isabel Ortigão** & Talita Vidal Pereira**

Resumo: Este artigo apresenta os resultados de uma investigação que buscou refletir sobre os impactos das avaliações externas na produção curricular. Em especial procurou-se compreender diferenças nas ênfases curriculares, em matemática, das redes municipais do estado do Rio de Janeiro, por meio da análise dos dados da Prova Brasil, nas edições de 2007 e 2011. Foram analisadas as respostas dos estudantes do quinto e do nono anos do ensino fundamental ao teste de matemática, utilizando-se como metodologia a análise do funcionamento diferencial do item (DIF), que busca detectar itens cuja probabilidade de acertos difere entre grupos distintos, mas de mesma habilidade cognitiva. No total foram analisados 356 itens de matemática. Os resultados evidenciaram que os itens não apresentaram DIF entre os grupos considerados. Argumenta-se que tais resultados podem estar associados, tanto à melhoria nos processos teórico-metodológicos da avaliação em larga escala, como por uma certa «homogeneização» curricular das escolas públicas brasileiras, que sistematicamente vem evidenciando pouco avanço no desenvolvimento de habilidades básicas.

Palavras-chave: homogeneização curricular, Prova Brasil, DIF

HOMOGENIZATION CURRICULUM AND BRAZILIAN NATIONAL ASSESSMENT SYSTEM: THE CASE OF RIO DE JANEIRO STATE

Abstract: This article presents the results of research that sought to reflect on the impact of external evaluations in curricular production. In particular it tried to understand differences in curricular emphases, in mathematics, between different municipal schools in the state of Rio de Janeiro, through the analysis of the *Prova* [Exam] Brazil 2007 and 2011 data. The responses of the students, of the fifth and the ninth years of elementary school, to the mathematics test were analyzed using the differential item functioning (DIF) methodology of analysis, which seeks to detect items whose probability of cor-

* O texto relaciona-se com a pesquisa desenvolvida com apoio da FAPERJ e da CAPES.

** Faculdade de Educação da Baixada Fluminense da Universidade do Estado do Rio de Janeiro (FEBF/UERJ) (Rio de Janeiro/Brasil).

rect answers differs among different groups with equivalent ability. A total of 356 mathematics items were analyzed, 178 of which were applied in the edition of 2007. The results show that the items did not present DIF between the groups considered. It is argued that these findings may be related both to the improvement of the theoretical and methodological processes of the large-scale assessment, as well as to a certain curriculum «homogenization» of the Brazilian public schools, which have been systematically showing little progress in the development of basic skills.

Keywords: curriculum homogenization, *Prova* Brazil, DIF

HOMOGÉNÉISATION DES PROGRAMMES ET LE SYSTÈME D'ÉVALUATION NATIONAL BRÉSILIEN: LE CAS DE L'ÉTAT DE RIO DE JANEIRO

Résumé: Cet article présente les résultats d'une investigation dont le but est de réfléchir sur les impacts des évaluations externes dans la production des programmes. Nous avons surtout cherché à comprendre les différences en ce qui concerne les emphases des programmes de mathématique des réseaux des municipalités de l'état de Rio de Janeiro, à partir de l'analyse des données de la Prova [l'Épreuve] au Brésil des entre 2007 et 2011. Pour ce faire, nous avons analysé les réponses aux épreuves de mathématique de collégiens de la sixième et de la troisième année du 1er cycle, en utilisant comme méthodologie l'analyse du fonctionnement différentiel de l'item (DIF), une méthodologie qui cherche à identifier des items dont la probabilité de parvenir à la bonne réponse diverge entre deux groupes différents dotés de même habilité cognitive. Nous avons analysé 356 items de mathématiques au total. Les résultats nous ont montré que les items n'ont pas présenté de DIF entre les groupes analysés. De tels résultats peuvent être associés soit à l'amélioration des processus théorico-méthodologiques de l'évaluation à une grande échelle, soit à une certaine «homogénéisation» des programmes des écoles publiques brésiliennes qui n'ont toujours pas démontré d'amélioration au niveau du développement des habilités de base.

Mots-clés: homogénéisation des programmes, *Prova* Brésil, DIF

Introdução

Em diversos países, desde o final dos anos 1980, crescem ações que envolvem a avaliação externa de sistemas ou redes educativas, de escolas e seus atores. Como afirma Nóvoa (2009: 13), «deixa de ser possível imaginar processos educativos que não conduzam a modalidades de julgamentos». Para o autor,

Nenhum tema deu origem a tantos estudos e pesquisas. Os investigadores e os especialistas procuraram compreender essa realidade, elaborando instrumentos cada vez mais sofisticados. Os professores basearam grande parte de sua autoridade pedagógica e mesmo de sua identidade profissional, no exercício da avaliação. Os decisores políticos colocaram-na no centro de suas preocupações. As famílias encararam sempre a avaliação como elemento central da sua ligação à escola. (Nóvoa, 2009: 14)

De modo geral, a avaliação é apontada como peça central das reformas educacionais ocorridas a partir do final da década de 1980 e início dos anos 1990, em muitos países. Diversos estudos, conduzidos a partir desse período, evidenciam fortes similaridades entre tais reformas, «como se estivessem seguindo um receituário de políticas educacionais, algum tipo de orquestração ou, no mínimo, uma história de origens comuns» (Brooke, 2012: 325). Brooke (2012), contudo, alerta para a pluralidade de sentidos que o termo «reforma educacional» carrega. Para ele, quando entendido com o sentido de uma mudança na política educacional para corrigir rumos, observa-se uma multiplicidade de enfoques e de casos tão amplos que não haveria como criar uma organização que dê conta de tal variedade.

Por um lado, as avaliações e os seus resultados tiveram o mérito de revelar os processos de desigualdades que permeiam os sistemas educativos. Muitos estudos foram conduzidos com a intenção de buscar compreender a distribuição desigual da educação escolar entre os diversos grupos sociais, culturais e económicos. Por meio de sofisticados procedimentos metodológicos, esses estudos evidenciam características dos estudantes e das suas escolas associadas à melhoria da qualidade educacional. Destacam-se, nesse sentido, os estudos desenvolvidos por Carnoy, Gove, e Marshall (2007), Franco, Sztajn, e Ortigão (2007), Scheerens (2005), Lee e Smith (2002), Lee e Bryk (1989), Anyon (1980), entre outros. Por outro lado, têm surgido críticas bastante fortes às formas como as avaliações têm sido conduzidas, às métricas utilizadas para monitorizar o desempenho dos estudantes e para avaliar os sistemas educativos. Dentre tais críticas, destacam-se os estudos que evidenciam que as avaliações externas estão pautadas por uma lógica de mercado que justificam práticas meritocráticas, conduzem a uma compreensão estreita e reduzida do sentido de «qualidade», induzem a uma padronização da produção curricular, silenciando as diferenças e conduzindo a um processo de homogeneização nos sistemas educativos (Tura & Pereira, 2013; Freitas, 2012; Fernandes, 2009; Ravitch, 2011; Lubienski & Lubienski, 2004, entre outros).

Há ainda estudos que evidenciam a avaliação externa como um processo social, que precisa de ser entendido em toda a sua complexidade, em especial, por possibilitarem às escolas «reflectir sobre as suas práticas numa perspectiva mais alargada» (Veloso, Abrantes, & Craveiro, 2011: 85). Para os autores,

A valorização da heterogeneidade dos contextos de aprendizagem e os impactos diferenciados da avaliação nas práticas das escolas poderão implicar uma aferição do modelo de avaliação externa atendendo: à necessidade de uma objectivação de determinados indicadores à permanência mais prolongada nas escolas por parte da equipa de avaliadores, à necessidade de o modelo de avaliação incorporar uma análise do contexto territorial, social e económico das escolas. (*ibidem*)

O objetivo deste texto é o de refletir sobre o impacto da avaliação externa na produção curricular brasileira, a partir de uma pesquisa que buscou analisar o comportamento diferencial

dos itens de matemática utilizados na avaliação nacional denominada Prova Brasil. Especificamente, o estudo envolveu um determinado tipo de análise estatística dos itens de matemática, respondidos por estudantes do ensino fundamental das diferentes redes municipais de educação do estado do Rio de Janeiro. A escolha por matemática relaciona-se ao caráter tipicamente escolar do conhecimento matemático, em comparação com o conhecimento da língua nativa, que mais frequentemente, desenvolve-se nos diversos ambientes frequentados pelos jovens. Já o recorte do estudo a um único estado brasileiro teve duas motivações: primeiro a limitação da capacidade do software estatístico usado na análise, que não comporta uma quantidade tão grande como os dados gerados por todo o universo brasileiro; segundo, devido ao fato da pesquisadora envolvida no estudo ser originária do estado do Rio de Janeiro. Na sequência, situamos o leitor em relação ao sistema brasileiro de avaliação, com foco especial à Prova Brasil e, na continuidade, apresentamos a abordagem metodológica utilizada na análise dos dados. Por fim, discutimos os resultados obtidos e conduzimos as nossas considerações finais.

Sistema nacional de avaliação: a Prova Brasil

A constituição de um sistema de avaliação da educação brasileira tem início no final dos anos 1980, a partir do reconhecimento da inexistência de estudos que mostrassem com clareza o atendimento educacional ofertado à população. O objetivo geral era o de mapear como, quando e quem tem acesso a um ensino de qualidade.

Para Gentili (1995), a mudança de eixo – da democratização da educação para a qualidade da educação – foi acompanhada, por «um duplo processo de transposição» (p. 116), caracterizado, tanto pelo deslocamento do problema da democratização ao da qualidade, como pela transferência dos conteúdos que caracterizam a discussão sobre qualidade no campo produtivo-empresarial para o campo das políticas educativas e para a análise dos processos pedagógicos.

Segundo o autor, essa transposição trouxe para o campo da educação a necessidade da mensurabilidade, pois «[este] sempre foi o aspecto (central) capaz de materializar qualquer aspiração empresarial que tenda a gerar melhorias nos níveis de qualidade» (p. 140). Para Gentili «o motivo de tal apego aos aspectos mensuráveis da qualidade é orientado pela maximização de utilidades e pela necessária adaptabilidade ao mercado que persegue toda empresa competitiva» (*ibidem*). Para o autor, essa lógica de ajustar educação ao mercado, num contexto de globalização, não mais vincula qualidade a desenvolvimento, mas à competitividade, na qual eficiência, adequabilidade e produtividade são conceitos centrais. De acordo com Lopes (2015: 455), uma lógica em que a qualidade da educação desejável se concretizaria a partir de um

sistema capaz de «medir, comparar e avaliar conhecimentos/padrões de aprendizagem para alcançar a medida/comparação/avaliação de pessoas».

A noção de qualidade, sem dúvida excludente, como uma nova estratégia competitiva que invade o campo educacional, marca a construção de sistemas avaliativos em diversos países, inclusive o Brasil. A primeira experiência tem início em 1990¹ com a criação do Sistema de Avaliação do Ensino Público (SAEP). Por dois momentos consecutivos (1990 e 1992) o SAEP aplicou testes² a uma amostra de estudantes do ensino fundamental de escolas públicas. Em 1995, tem-se a primeira grande reformulação do sistema avaliativo, que passa a denominar-se Sistema de Avaliação da Educação Básica (SAEB)³, abarcando em seu plano amostral, além das redes públicas, a educação privada e envolvendo os níveis fundamental e médio de ensino, em todas as unidades federativas. Com isso, passa-se a avaliar estudantes dos anos terminais em cada fase escolar – 5º e 9º anos do ensino fundamental e 3º ano do ensino médio – aplicando-se a estes testes de matemática e de leitura. É neste ano, também, que a avaliação brasileira incorpora a teoria da resposta ao item (TRI) como abordagem metodológica na construção e na análise dos instrumentos de coleta de informação – os testes e os questionários.

Diferente da teoria clássica das medidas (Vianna, 1987), que leva em conta o *score* dos indivíduos na realização dos testes como um todo, na TRI a proficiência dos respondentes é avaliada tendo como referência as respostas dadas em cada item. Dessa forma, a TRI possibilita comparações, tanto entre populações diferenciadas, desde que submetidas a testes que tenham itens em comum, como entre indivíduos que integram uma mesma população submetidos a provas totalmente diferentes. Com isso, é possível avaliar os resultados de testes aplicados entre diferentes tipos de instituições escolares, assim como o desempenho dos estudantes de um ano escolar para o outro. São vantagens que explicam como a avaliação nacional tem avançado na direção de uma maior sofisticação teórico-metodológico. Soma-se a elas o estabelecimento de matrizes de referência para a avaliação, cuja função é a de direcionar o que será avaliado em cada área do conhecimento, prevista no sistema avaliativo, organizada de modo a descrever as competências e habilidades, em articulação ao conteúdo que cada aluno deve dominar. A principal justificativa para a elaboração de matrizes está na necessidade de se estabelecerem

¹ A proposta de institucionalização de um sistema de avaliação surgiu, no Brasil, durante o período de redemocratização do país (1985/1986). Ganhavam força o debate sobre a questão federativa, a redefinição dos papéis e atribuições dos diversos níveis de governo, assim como o sentido de cooperação, de articulação, da integração e da parceria.

² As primeiras avaliações envolveram a aplicação de testes (de matemática, leitura, ciências, geografia e história) a estudantes do 3º, 5º e 7º anos do ensino fundamental.

³ A partir desse momento o SAEB passa a ser conduzido pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), uma autarquia vinculada ao Ministério da Educação (MEC).

provas a partir de parâmetros consensuais, sejam estes advindos da reflexão teórica sobre a estrutura da ciência e sua correspondente adequação às estruturas de conhecimento, sejam advindos de uma consulta nacional sobre os conteúdos praticados e indicados pelas escolas brasileiras. (Secretaria de Educação Fundamental, 1997: 9)

Inicia-se, desde então, por meio da avaliação em larga escala, um processo indutor nos currículos brasileiros. A avaliação ganha centralidade e passa a ser concebida como instrumento que vai impulsionar a qualidade da educação.

Em 2005, passados doze anos da primeira experiência avaliativa, o governo brasileiro amplia a atuação da avaliação externa com a criação da Prova Brasil, um sistema censitário para avaliar as escolas públicas brasileiras. Com isso, expande o alcance dos resultados porque oferece informações, não apenas para o Brasil e unidades da Federação, mas também para cada município e escola participante. Cabe observar que a Prova Brasil possibilitou a criação do índice de desenvolvimento da educação básica (IDEB), um indicador que mede a relação entre resultados médios dos estudantes no exame nacional e a taxa de aprovação da escola.

O SAEB passa, então, a ser composto por dois sistemas avaliativos: Avaliação Nacional da Educação Básica (Aneb) e a Avaliação Nacional do Rendimento Escolar (Anresc), conhecida como Prova Brasil.

A Anresc (Prova Brasil) foi idealizada para atender a demanda dos gestores públicos, educadores, pesquisadores e da sociedade em geral por informações sobre o ensino oferecido em cada município e escola. O objetivo da avaliação é auxiliar os governantes nas decisões e no direcionamento de recursos técnicos e financeiros, assim como a comunidade escolar, no estabelecimento de metas e na implantação de ações pedagógicas e administrativas, visando à melhoria da qualidade do ensino. (Ministério da Educação, 2015: 1)

Mais recentemente, uma nova reestruturação do SAEB é implementada no sentido de a avaliação abarcar também a alfabetização. Isso ocorre com a criação da avaliação nacional da alfabetização (ANA), ampliando a atuação do SAEB a toda a educação básica.

Paralelamente às avaliações nacionais e à participação do Brasil no Programa Internacional de Avaliação dos Estudantes (PISA), muitos estados e ou municípios brasileiros constituíram seus próprios sistemas de avaliação, em geral conduzidos por empresas privadas, que se responsabilizam pela organização e condução da avaliação, bem como pela elaboração de relatórios contendo os resultados obtidos por cada escola participante.

Podemos perceber que, nesses vinte cinco anos de convivência com a avaliação em larga escala, ela tem-se destacado como elemento central da regulação e da administração gerencial e competitiva do «Estado-avaliador» (Coelho, 2008: 231). Tal centralização tem sido apontada como produtora de um processo de apagamento das diferenças – um processo que, segundo Ball (2005), provoca o aumento da individualização, que inclui a destruição das solidariedades baseadas numa identidade profissional comum. Mais ainda, é apontada como indutora de uma

homogeneização que, em nome da igualdade de oportunidade, cerceia professores e estudantes de insubordinarem-se criativamente (D'Ambrósio & Lopes, 2015).

Como anunciado inicialmente, este texto objetivou refletir sobre homogeneização curricular a partir dos resultados de uma pesquisa, conduzida por uma de nós, que analisou os resultados da Prova Brasil. Especificamente, investigou ênfases curriculares diferenciadas em matemática entre escolas das redes municipais do estado do Rio de Janeiro. Na sequência, apresentamos brevemente um panorama das discussões emanadas por pesquisas em avaliação e ensino de matemática e, em seguida, discutimos os procedimentos metodológicos utilizados.

Avaliação externa e ensino de matemática: evidências das pesquisas

A década de 1960 foi marcante no cenário educacional, em especial, pelas publicações que, recorrentemente, apresentavam evidências sobre a associação entre condições escolares e características socioeconômicas e culturais dos estudantes. Dentre essas, destaca-se o estudo realizado por Jean Anyon na década de 1970 (Anyon, 1980). Nele, a autora analisou o trabalho pedagógico, por meio de observações de aulas de Matemática (ocorridas num período de um ano em turmas de quinta série) e de entrevistas a alunos e professores em cinco escolas americanas – escolhidas com base no perfil socioeconômico dos estudantes matriculados. Anyon concluiu que os currículos, as práticas e as habilidades desenvolvidas estavam associados aos diferentes perfis socioeconômicos observados. Os estudantes com baixo perfil socioeconômico tinham acesso a menos conteúdos matemáticos, quando comparados àqueles com perfis mais elevados. As práticas pedagógicas também se diferenciavam significativamente: estudantes mais pobres eram expostos a atividades que envolviam repetição e treino, enquanto os outros, os mais favorecidos socialmente, eram conduzidos a se envolverem com resolução de problemas matemáticos, com a reflexão e a discussão dos resultados e processos utilizados. Para a autora, ficou evidente a reprodução e a manutenção, pela escola, das desigualdades existentes na sociedade.

No Brasil, principalmente a partir da disponibilização dos dados das avaliações nacionais da educação básica, tem sido possível investigar a relação entre características escolares, fatores socioeconômicos e culturais e os resultados acadêmicos dos estudantes. Franco, Sztajn e Ortigão (2007), com base nos dados do SAEB 2001, mostraram, por meio de análise multinível, que quando professores enfatizam resolução de problemas em suas aulas de Matemática, os estudantes tendem a apresentar desempenhos melhores nesta disciplina, o que resulta em uma apropriação melhor do conhecimento matemático.

Ortigão, Franco e Carvalho (2007) também apresentam evidências de desigualdades na distribuição de conteúdos matemáticos selecionados por professores para o ensino de matemática

no ensino fundamental da cidade do Rio de Janeiro⁴. O estudo envolveu conhecer os conteúdos selecionados para o ensino e, em que medida esta seleção era impactada por características sociais e econômicas médias dos estudantes das escolas. Por meio da aplicação de um modelo de regressão logística, os autores concluem que em escolas cujos estudantes apresentam baixo poder aquisitivo, a probabilidade de professores selecionarem temas de estatística – uma das subáreas do currículo escolar – diminui quando comparado com escolas que atendem a estudantes cujas famílias apresentam melhores condições socioeconômicas.

Mais recentemente, tem sido possível encontrar investigações que buscam compreender a distribuição dos estudantes ao longo da escala de proficiência da avaliação nacional Prova Brasil. É o caso do estudo exploratório desenvolvido por Oliveira (2015), que analisou o percentual de acerto em itens de matemática aplicados na Prova Brasil 2011. No estudo evidenciou-se um baixo percentual de acertos (abaixo de 40%) em itens vinculados a descritores que avaliam habilidades em resolver problemas, envolvendo noções de geometria e medidas e grandezas. E mais, em uma análise longitudinal, envolvendo dados da Prova Brasil 2007 a 2011, a autora evidenciou pouco avanço nas habilidades de matemática de estudantes do 5º e 9º anos do ensino fundamental na escala de proficiência em matemática.

Resultados análogos foram observados no estudo desenvolvido por Câmara, Ortigão e Aguiar (2014), envolvendo os dados da avaliação do estado de Pernambuco – SAEPE. Nele, os autores analisaram os itens de matemática utilizados na avaliação externa do estado de Pernambuco e aplicados aos estudantes do quinto ano do ensino fundamental, com o intuito de identificar, por meio do percentual de acertos nos itens, avanços nas aprendizagens dos estudantes. O estudo concluiu o pouco avanço na aprendizagem de matemática neste segmento escolar no estado em questão.

O breve panorama acima foi o mote para a proposição da investigação, referenciada neste texto, que objetivou compreender os resultados em matemática da Prova Brasil, por meio da aplicação de um modelo estatístico denominado análise diferencial do item (DIF)⁵. A ideia foi a de verificar se haveria itens da prova de Matemática mais fáceis para um grupo de alunos em detrimento a outro, por alguma razão que extrapola a dificuldade do item. Os dois grupos escolhidos para a verificação do DIF foram: (a) estudantes de escolas públicas (municipais), urbanas, situadas na capital e (b) estudantes de escolas públicas (municipais), urbanas, situadas nas outras regiões do estado do Rio de Janeiro. A seguir, apresentamos os procedimentos metodológicos desenvolvidos, cujos resultados apoiam a discussão do artigo proposto.

⁴ A pesquisa envolveu uma amostra probabilística de 48 escolas, sendo 25 da rede municipal do Rio de Janeiro e 23 escolas particulares da mesma cidade.

⁵ *Differential item functioning.*

Procedimentos metodológicos

Estudos envolvendo DIF vêm ganhando força no campo da avaliação da educação, em especial por se constituir uma boa estratégia para a compreensão do currículo escolar. Essa abordagem metodológica possibilita identificar itens que violam um dos principais pressupostos da TRI, segundo o qual a probabilidade de acertar um item é função da proficiência do aluno (Andriola, 2006; Soares, Genovez, & Galvão, 2005; Soares, 2005; Aguiar & Ortigão, 2012; Ortigão, 2014). Ou seja, alunos de grupos distintos com igual proficiência têm a mesma probabilidade de acertar um item. Caso isso não ocorra, podemos afirmar que algum fator que extrapola a habilidade cognitiva do aluno está a tornar um item mais fácil para um dos grupos. A análise de DIF permite investigar se um item apresenta ou não grau de dificuldade diferente para subgrupos dentro de uma população que têm, ou deveria ter, o mesmo nível de conhecimento.

Historicamente, a preocupação com o DIF está associada ao desejo de que se construíssem questões de teste que não fossem afetadas por características étnico-culturais dos grupos submetidos aos testes de avaliação educacional – ligada, portanto, às campanhas em prol da melhoria dos direitos civis dos cidadãos comuns, nos anos de 1960, nos Estados Unidos da América. Esses anos foram marcados por uma enorme preocupação com a igualdade de oportunidades, pelas críticas aos sistemas educacionais discriminadores, pelo desenvolvimento de um conceito popular e legal de ações afirmativas e pela consciência racial/étnica. Diferenças educacionais, resultantes de sistemas educacionais com muita iniquidade, passaram a ser vistas como vestígios de uma velha ordem segregadora. Assim, *scores* de testes, refletindo essas diferenças, foram considerados, da mesma forma, discriminadores, e passou-se a usar o termo *viés*, ao referenciá-los. Os estudos para identificação de *viés* tinham por objetivo provar que os testes ou instrumentos de medida não possuíam nenhum tipo de *viés*.

Muitos pesquisadores começaram a dedicar-se ao estudo sistemático das diferenças entre os grupos étnicos, com o objetivo de encontrar explicações convincentes para as grandes diferenças de rendimento, observados entre os diversos grupos étnicos e socioeconômicos, que refletiam disparidades nas oportunidades educacionais e se mostravam injustos (Angoff, 1993; Linn, Levine, Nicholas, & Wardrop, 1981; Shepard, Camilli, & Averill, 1981). Tais estudos adotaram a seguinte definição de *viés*: um item é enviesado se sujeitos de habilidades iguais, mas de culturas diferentes, não têm a mesma probabilidade de acertar o item.

Tradicionalmente, os itens de uma avaliação externa são pré-testados e analisados estatisticamente. Portanto, em princípio, não se espera encontrar itens com DIF, no entanto, uma série de fatores pode contribuir para que ele se manifeste, como a não familiaridade com o conteúdo do item ou o grau de interesse pessoal em um dado conteúdo. No relatório de análise de DIF, entre as diferentes regiões do Brasil, produzido a partir dos resultados do SAEB de 1999 (Minis-

tério da Educação e do Desporto, 2000; Barbosa et al., 2000) são elencados alguns fatores que poderiam ter influenciado o comportamento de itens, tais como: regionalismos, especificidades linguísticas ou culturais, dado as dimensões continentais do país. Mas, ainda assim, segundo Fidalgo (2006), existindo poucos itens com DIF, a evidência estatística da diferença entre o desempenho de grupos não deve ser persuasiva quanto à decisão de excluir ou não esses itens do teste. Pois, além de interferirem pouco no resultado da proficiência, podem trazer informações adicionais relevantes para entender algumas possíveis diferenças pedagógicas e/ou características específicas dos grupos.

A pesquisa aqui apresentada teve por objetivo investigar a presença DIF nos itens de matemática da Prova Brasil, edições de 2007 e 2011⁶, considerando-se apenas os dados dos estudantes das redes municipais do estado do Rio de Janeiro. Como anunciado anteriormente, problemas operacionais com o *software* usado para fazer as análises de DIF não nos permitiu realizá-la com a totalidade das informações, o que nos conduziu a considerar apenas os dados do estado do Rio de Janeiro. No desenvolvimento da pesquisa, as seguintes questões foram conduzidas:

- (a) há diferenças de competência cognitiva entre estudantes de escolas municipais situadas na capital e aqueles que estudam em escolas das outras regiões do estado do Rio de Janeiro?
- (b) é possível identificar competências matemáticas que são exploradas de modo diferenciado nos currículos dessas escolas?

O grupo de referência para o estudo foi o correspondente à rede municipal da cidade do Rio de Janeiro, que congrega cerca de 1200 escolas. Já o grupo focal engloba as redes municipais das demais cidades do estado do Rio de Janeiro. Desta forma, a análise envolveu dois grupos: capital e interior. No total foram analisados 356 itens de matemática, sendo 87 e 88 aplicados aos estudantes do 5º ano, respectivamente em 2007 e 2011, e 91 e 90 aos estudantes do 9º ano, respectivamente em 2007 e 2011.

A análise de DIF foi realizada segundo o modelo desenvolvido por Mantel e Haenszel (1959) que tem por base a comparação de frequências observadas nos subgrupos constituídos por indivíduos com habilidades semelhantes. A sua avaliação é feita pelo nível de significância das diferenças. O programa estatístico usado para fazer as análises foi o SisAnI, desenvolvido pelo Centro de Avaliação e Políticas Públicas – CAED (UFJF).

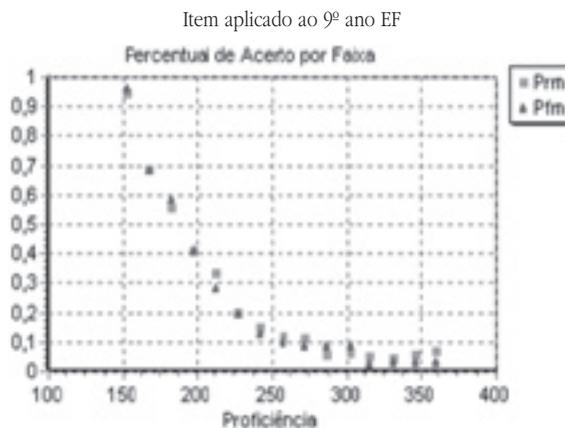
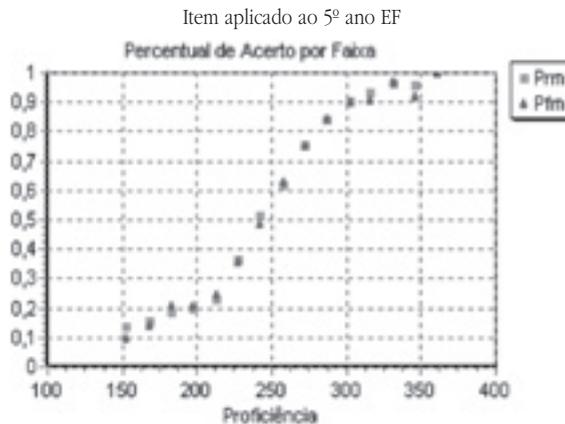
⁶ Devido a problemas técnicos com a edição da Prova Brasil 2009, não foi possível considerar esta base no estudo.

Resultados e discussão

Os resultados da análise estatística realizada neste estudo evidenciaram a não presença de DIF nos itens da Prova Brasil nas duas edições consideradas – 2007 e 2011. Ou seja, não se detetou diferenças nas probabilidades de acerto do item quando se comparou grupos de estudantes da rede municipal da capital fluminense com os estudantes das demais redes municipais do estado. As figuras abaixo ilustram os resultados das análises envolvendo DIF. São apresentados os resultados de dois itens da Prova Brasil 2007 (Figura 1) e dois da prova Brasil 2011 (Figura 2), sendo um item aplicado aos estudantes do 5º ano e outro aos estudantes do 9º ano do Ensino Fundamental.

FIGURA 1

Resultado da análise de DIF aos itens de matemática da Prova Brasil 2007: 5º e 9º anos do ensino fundamental

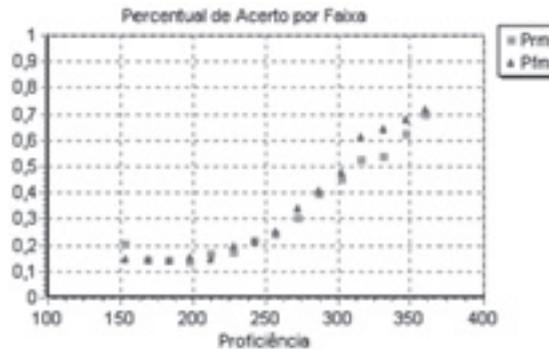


Fonte: INEP. Prova Brasil 2007 (elaboração própria).

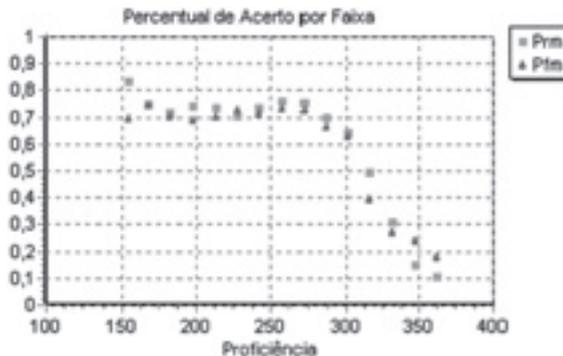
FIGURA 2

Resultado da análise de DIF aos itens de matemática da Prova Brasil 2011: 5º e 9º anos do ensino fundamental

Item aplicado ao 5º ano EF



Item aplicado ao 9º ano EF



Fonte: INEP. Prova Brasil 2011 (elaboração própria).

Observa-se nas figuras acima que os pontos sobre a curva aparecem sobrepostos, indicando a não presença de DIF entre os dois grupos considerados.

Duas hipóteses podem estar associadas a esse resultado: (i) melhoria nos processos teórico-metodológicos da avaliação em larga escala; (ii) certa «homogeneização» no currículo de Matemática de escolas públicas do estado do Rio de Janeiro, que pouco possibilita diferenciar as habilidades dos estudantes que estudam em escolas da capital e do interior.

Com relação à primeira hipótese, é notório o avanço teórico e o aprimoramento metodológico e técnico nos procedimentos que envolvem a avaliação em larga escala. Os debates ocorridos no seminário *Vinte e cinco anos de avaliação de sistemas educacionais no Brasil*, pro-

movido pela Fundação Carlos Chagas, em 2012, evidenciou bem este aspecto (Bauer, Gatti, & Tavares, 2013; Bauer & Gatti, 2013). Para estas autoras, desde a criação do sistema brasileiro de avaliação, no início dos anos 1990, foram detetados quase trezentos estudos acadêmicos, entre teses e dissertações, sobre as avaliações de sistemas educacionais no Brasil. Desses, há estudos teóricos sobre os procedimentos metodológicos adotados nas avaliações em larga escala, que evidenciam o aprimoramento das teorias das medidas e o avanço em relação à comparabilidade de grupos distintos. Há ainda estudos que se dedicaram a compreender as implicações das avaliações de sistemas educacionais em seus diversos aspectos – na gestão da escola e do sistema de ensino, na formação continuada e nas práticas dos professores em sala de aula etc. Destacam-se ainda estudos que utilizam bases de dados das diversas avaliações para fazer análises de fatores associados ao desempenho dos estudantes. Com base nesses estudos, inferimos que os avanços alcançados até o momento no campo da avaliação educacional podem estar por trás de aprimoramentos e ajustes aos modelos teóricos usados na construção e validação de itens.

Em relação à segunda hipótese, estudos anteriores têm evidenciando o pouco avanço dos estudantes das redes públicas municipais e estaduais em relação ao desenvolvimento de habilidades matemáticas (Oliveira, 2015; Câmara et al., 2014). Tais estudos chamam a atenção para o perigo de uma perspectiva homogeneizadora no currículo de Matemática das redes públicas municipais do estado do Rio de Janeiro, que pouco possibilita diferenciar resultados médios de estudantes que estudam em escolas da capital e do interior. Eles alertam para a urgência em rever os currículos desta área de conhecimento nas escolas públicas, na medida em que pouco avanço foi verificado no desenvolvimento das habilidades básicas.

Por outro lado, no campo do currículo em geral, têm sido desenvolvidos estudos que destacam o caráter indutor da avaliação em larga escala na produção curricular. Um contexto em que a avaliação ganha centralidade concebida como instrumento capaz de garantir a qualidade da educação sem que haja uma ampla discussão sobre o ideal educacional desejado (Fernandes & Nazareth, 2011). E em que um Estado regulador promove políticas marcadas por tentativas de normatização e regulamentação do cotidiano das escolas (Pereira & Velloso, 2012).

Trata-se, como afirma Ball (2006), de políticas pautadas por uma lógica de mercado «em que valores como competitividade, empreendedorismo e flexibilidade se impõem no lugar de estabilidade, cooperação e justiça social» (Fernandes & Nazareth, 2011: 64) e justificam as práticas meritocráticas e a produção de rankings entre diferentes instituições sem que as especificidades que as caracterizam sejam necessariamente levadas em conta. A busca pela qualidade que justifica as práticas avaliativas e sustenta os sistemas de avaliação em larga escala tem-se limitado à aferição do desempenho dos alunos no desenvolvimento de habilidades arbitrárias e padronizadas (Fernandes & Nazareth, 2011). Habilidades que acabam por ser privilegiadas nos currículos, por que essa torna-se uma estratégia das escolas para melhorar as suas performances

para se inserir na lógica que organiza esse jogo (Ball, 2004). Trata-se de um processo em que as avaliações externas acabam induzindo o currículo. Ou ainda como afirma Lopes (2015: 461), um processo pautado em um «projeto fixado como o melhor para todos, um conhecimento ou uma cultura comum, (...), em detrimento da heterogeneidade e da negociação contextual capaz de criar a significação de currículo para múltiplos projetos possíveis».

Dessa forma, mesmo reconhecendo que o resultado da pesquisa realizada precisa ser olhado com cautela e que os sofisticados procedimentos utilizados para detetar DIF são apropriados apenas para detetar o *viés* em potencial de um item, mas não possibilitam uma explicação das suas causas, a lógica de funcionamento das avaliações, associada à sofisticação dos instrumentos, procedimentos estatísticos e análises teóricas, podem estar a contribuir para um processo de homogeneização e apagamento das singularidades, sem que isso signifique uma melhoria sensível no nível de aprendizagem dos estudantes.

Considerações finais

O procedimento que desenhou o percurso deste trabalho foi o de responder se Prova Brasil utiliza itens que apresentam DIF. A hipótese inicial da presença de itens com DIF na avaliação da Prova Brasil não se confirmou. Ou seja, não foram verificados itens de matemática com DIF quando comparamos os resultados de estudantes de escolas da rede municipal da capital e das demais redes municipais do estado do Rio de Janeiro. Os primeiros indícios deste resultado conduziram-nos à busca de argumentos que nos ajudassem a explicá-los. A literatura apontou-nos dois caminhos: o avanço teórico e o aprimoramento metodológico e técnico nos procedimentos que envolvem a avaliação em larga escala e, por outro lado, certa homogeneidade no perfil cognitivo dos estudantes da rede pública brasileira. Homogeneidade que pode ser resultado de um processo de indução curricular que as avaliações externas têm produzido, sem que isso se expresse em resultados que promulguem a qualidade desejada. Um fenómeno que Mello (2014) constata ao defender a necessidade de uma base nacional curricular comum para revertê-lo de forma que a avaliação possa ser norteada por um currículo e não vice-versa. Um fenómeno que Lopes (2013: 20) concebe como «tentativa de regular a atividade curricular de professores e alunos». De qualquer forma, a adoção de qualquer das duas hipóteses, certamente, deve ser vista com parcimónia e cautela. É necessário um aprofundamento da investigação para que se possa confirmá-las ou refutá-las e isso demandaria um novo mergulho nos dados e nos modelos de análise.

Por fim, o esforço para aquisição das ferramentas empregadas nesta pesquisa e o rigor metodológico que as acompanha foi decisivo para permitir uma análise relevante aos dados

apresentados. No entanto, é desejável que haja um desdobramento deste trabalho no sentido de complementá-lo, com análises mais sofisticadas, que possam conjugar diferentes abordagens de pesquisa, o que implica a inclusão de resultados de outros estados e a apropriação de aportes teóricos que possibilitem a análise de um conjunto de fatores que podem estar interferindo no DIF.

Correspondência: Faculdade de Educação da Baixada Fluminense, Universidade do Estado do Rio de Janeiro (UERJ), Rua Gen. Manoel Rabelo, s/n, Vila São Luiz – Duque de Caxias, Rio de Janeiro, 25065-050, Brasil
Email: isabelortigao@terra.com.br; p.talitavidal@gmail.com

Referências bibliográficas

- Aguiar, Glauco, & Ortigão, Maria Isabel Ramalho (2012). Letramento em matemática: Um estudo a partir dos dados do PISA 2003. *Boletim de Educação Matemática – Bolema*, 26, 1-21.
- Andriola, Wagner B. (2006). Estudo sobre o viés de itens em testes de rendimento: Uma retrospectiva. *Estudos em Avaliação Educacional*, 17(35), 115-134. Retirado de www.fcc.org.br/pesquisa/publicacoes/eae/1342/1342/pdf
- Angoff, William H. (1993). Perspectives on differential item functioning. In Paul W. Holland & Howard Wainer (Orgs.), *Differential item functioning* (pp. 3-24). Hillsdale, NJ: LEA.
- Anyon, Jean (1980). Social class and the hidden curriculum at work. *Journal of Education*, 162(1), 67-92.
- Ball, Stephen (2004). Performativities and fabrications in the educational economy: Towards the performative society. In Stephen J. Ball (Ed.), *The Routledge Falmer reader in sociology of education* (pp. 143-155). London: Routledge Falmer.
- Ball, Stephen (2005). Profissionalismo, gerencialismo e performatividade. *Cadernos de Pesquisa*, 35(126), 539-564.
- Ball, Stephen (2006). Sociologia das políticas educacionais e pesquisa crítico-social: Uma revisão pessoal das políticas educacionais e da pesquisa em política educacional. *Currículo sem Fronteiras*, 6(2), 10-32. Retirado de <http://www.curriculosemfronteiras.org/vol6iss2articles/ball.pdf>
- Barbosa, Maria Eugênia Ferrão, Fernandes, Cristiano, & Beltrão, Kaizô (2000). *Análise descritiva dos dados do SAEB-99: Relatório técnico INEP/MEC*. Unpublished manuscript.
- Bauer, Adriana, & Gatti, Bernadete A. (Org.). (2013). *Ciclo de debates vinte cinco anos de avaliação de sistemas educacionais no Brasil: Implicações nas redes de ensino, no currículo e na formação de professores*. Florianópolis: Editora Insular.
- Bauer, Adriana, Gatti, Bernadete, & Tavares, Marisa (Org.). (2013). *Ciclo de debates vinte cinco anos de avaliação de sistemas educacionais no Brasil: Origens e pressupostos*. Florianópolis: Editora Insular.
- Brooke, Nigel (Org.). (2012). *Marcos históricos na reforma da educação*. Belo Horizonte: Fino Traço.

- Câmara, Marcelo, Ortigão, Maria Isabel Ramalho, & Aguiar, Glauco (2014). *Saberes docentes versus saberes discentes: Convergências e divergências. Relatório Técnico. Edital Projetos de pesquisa e desenvolvimento de instrumentos de avaliação de competências para a educação básica*. Convênio CAED/PUC-Rio. Unpublished manuscript.
- Carnoy, Martin, Gove, Amber K., & Marshall, Jeffery H. (2007). *Cuba's academic advantage: Why students in Cuba do better in school*. California: Stanford University Press.
- Coelho, Maria Inês de Matos (2008). Vinte anos de avaliação da educação básica no Brasil: Aprendizagens e desafios. *Ensaio*, 16(59), 229-258. Retirado de <http://www.scielo.br/pdf/ensaio/v16n59/v16n59a05.pdf>
- D'Ambrósio, Beatriz, & Lopes, Celi Espasadin (Eds.). (2015). *Creative insubordination in brazilian mathematics education research*. Ohio: Lulu Press, Inc.
- Fernandes, Claudia de Oliveira, & Nazareth, Henrique D. G. (2011). A retórica por uma educação de qualidade e a avaliação de larga escala. *Impulso*, 21(51), 63-71. Retirado de <http://www.metodista.br/revistas/revistas-nimep/index.php/impulso/article/viewFile/526/550>
- Fernandes, Domingos (2009). *Avaliar para aprender: Fundamentos, práticas e políticas*. São Paulo: Editora UNESP.
- Fidalgo, Ángel M. (1996). Funcionamiento diferencial de los ítems. In Jose Muñiz (Org.), *Psicometría* (pp. 371-455). Madrid: Universitat.
- Franco, Crezo, Sztajn, Paola, & Ortigão, Maria Isabel Ramalho (2007). Mathematics teachers, reform and equity: Results from the brazilian national assessment. *Journal for Research in Mathematics Education*, 38(4), 393-419.
- Freitas, Luiz C. (2012). Os reformadores empresariais da educação: Da desmoralização do magistério à destruição do sistema público de educação. *Educação & Sociedade*, 119, 379-404. Retirado de <http://www.scielo.br/pdf/es/v33n119/a04v33n119.pdf>
- Gentili, Pablo (1995). O discurso da qualidade como nova retórica conservadora no campo educacional. In Pablo Gentili & Tomaz Tadeu da Silva (Orgs.), *Neoliberalismo, qualidade total e educação: Visões críticas* (pp. 111-177). Petrópolis: Vozes.
- Lee, Valerie, & Bryk, Anthony (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, 62, 172-192.
- Lee, Valerie, & Smith, Julia B. (2002). Tamanho da escola: Qual é o mais efetivo e para quem?. *Estudos em Avaliação Educacional*, 25, 77-121. Retirado de <http://publicacoes.fcc.org.br/ojs/index.php/ea/article/view/2192>
- Linn, Robert, Levine, Michael, Nicholas, Hastings, & Wardrop, James L. (1981). Item bias in a test of reading Comprehension. *Applied Psychological Measurement*, 5, 159-173.
- Lopes, Alice C. (2013). Teorias pós-críticas, política e Currículo. *Educação, Sociedade & Culturas*, 39, 7-23. Retirado de <http://www.fpce.up.pt/ciie/sites/default/files/02.AliceLopes.pdf>
- Lopes, Alice C. (2015). Por um currículo sem fundamentos. *Linhas Críticas*, 21(45), 445-466. Retirado de <http://periodicos.unb.br/index.php/linhascriticas/article/view/16735>
- Lubienski, Christopher, & Lubienski, Sarah (2004). *The public school advantage: Why public schools outperform private schools*. Chicago: The University of Chicago Press.

- Mantel, Nathan, & Haenszel, William (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Mello, Guiomar N. (2014). *Currículo da educação básica no Brasil: Concepções e políticas*. São Paulo: EESP. Retirado de <http://www.ceesp.sp.gov.br/comunicado.php?id=321>
- Ministério da Educação (2015). *SAEB/Prova Brasil 2011: Primeiros resultados*. Retirado de http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/resultados/2012/Saeb_2011_primeiros_resultados_site_Inep.pdf
- Ministério da Educação e do Desporto (2000). *SAEB 99: Relatório da análise de comportamento diferencial dos itens (Differential Item Functioning – DIF) entre regiões*. Rio de Janeiro: Fundação Cesgranrio.
- Nóvoa, António (2009). Prefácio. In Domingos Fernandes, *Avaliar para aprender: Fundamentos, práticas e políticas* (pp. 13-17). São Paulo: Editora UNESP.
- Oliveira, Adriana Souza (2015). *Um olhar sobre a Prova Brasil: Análise dos resultados em matemática* (Dissertação de mestrado). Faculdade de Educação da Baixada Fluminense, Universidade do Estado do Rio de Janeiro, Duque de Caxias, Brasil.
- Ortigão, Maria Isabel Ramalho, Franco, Creso, & Carvalho, João Bosco Pitombeira (2007). A distribuição social do currículo de matemática: Quem tem acesso a tratamento da informação?. *Educação Matemática Pesquisa*, 9(2), 249-273. Retirado de <http://revistas.pucsp.br/index.php/emp/article/viewFile/904/597>
- Ortigão, Maria Isabel Ramalho (2014). *Estudo comparativo entre escolas situadas em periferias e em capitais brasileiras sobre ênfases curriculares em matemática, a partir da análise do comportamento diferencial do item (DIF) na Prova Brasil*. Rio de Janeiro: Universidade do Estado do Rio de Janeiro.
- Pereira, Talita Vidal, & Velloso, Luciana (2012). Um salto para a performatividade: Sentidos atribuídos à qualidade da educação. *Ensaio: Avaliação e Políticas Públicas em Educação*, 20(74), 73-88. Retirado de <http://www.redalyc.org/articulo.oa?id=399538140005>
- Ravitch, Diane (2011). *Vida e morte do grande sistema escolar americano: Como os testes padronizados e o modelo de mercado ameaçam a educação*. Porto Alegre: Sulina.
- Scheerens, Jaap (2005). *Review and meta-analyses of school and teaching effectiveness*. The Netherlands: Department of Educational Organization and Management of the University of Twente. Retirado de <https://www.utwente.nl/bms/omd/Medewerkers/artikelen/scheerens/rev-meta-analysis2007.PDF>
- Secretaria de Educação Fundamental (1997). *Parâmetros curriculares nacionais: Matemática/1ª e 4ª séries*. Brasília: MEC/Secretaria de Ensino Fundamental.
- Shepard, Lorrie A., Camilli, Gregory, & Averill, M. (1981). Comparison of procedures for detecting test bias with both international and external ability criteria. In Paul W. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 137-159). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Soares, Francisco (2005). Qualidade e equidade na educação básica brasileira: Fatos e possibilidades. In Nigel Brooke & Simon Schwartzman. *Os desafios da educação no Brasil* (pp. 91-118). Rio de Janeiro: Nova Fronteira.
- Soares, Tufi M., Genovez, Silene F. M., & Galvão, Ailton F. (2005). Análise do comportamento diferencial dos itens de Geografia: Estudo da 4ª série avaliada no Proeb/Simave 2001. *Estudos em Avaliação Educacional*, 16(32), 481-493. Retirado de <http://www.din.uem.br/sbpo/sbpo2005/pdf/arq0197.pdf>

- Tura, Maria Lourdes Rangel, & Pereira, Talita Vidal (2013). Políticas curriculares, sistemas de avaliação e conhecimentos escolares. In Maria de Lourdes R. Tura & Maria Malta A. Garcia (Org.), *Currículo, políticas e ação docente* (pp. 111-125). Rio de Janeiro: EdUERJ.
- Veloso, Luísa, Abrantes, Pedro, & Craveiro, Daniela (2011). A avaliação externa de escolas como processo social. *Educação, Sociedade & Culturas*, 33, 69-88. Retirado de http://www.fpce.up.pt/ciie/revistaesc/ESC33/ESC33_Artigos_Veloso.pdf
- Vianna, Heraldo M. (1987). *Testes em educação*. São Paulo: IBRASA.