

O papel dos testes padronizados na política educacional para o ensino básico nos Estados Unidos

Remo Moreira Brito Bastos ^a

Resumo

Com o advento do programa federal *No Child Left Behind*, um aumento desproporcional das horas destinadas a testes, o estreitamento de conteúdos curriculares e as avaliações verticalizantes, focando responsabilização de professores e diretores, passaram a conformar a agenda para a reforma educacional nos Estados Unidos. O presente artigo, desenvolvido por meio de pesquisa bibliográfica e documental, questiona as premissas que fundamentam essas políticas, fortemente estruturadas em torno dos testes padronizados de larga escala, os quais constituem a “ferramenta por excelência” do programa governamental em questão. Ademais dos perversos efeitos impingidos aos educandos, mostram-se as limitações intrínsecas desses instrumentos para aferir a assimilação de conhecimentos por parte dos discentes, bem como a proliferação de todo tipo de expedientes de questionável validade ética com vistas a simular o alcance das metas de desempenho acadêmico estipuladas pelo mencionado programa.

Palavras-chave: Testes padronizados. Política educacional. Estados Unidos.

1 Introdução

Não obstante sua histórica onipresença na educação básica estadunidense, os testes padronizados têm-se transformado de mecanismo utilizado para aferir o nível de assimilação do conteúdo ministrado ao aluno na finalidade última do processo de ensino-aprendizagem naquela nação, com graves consequências para a formação de seu corpo discente. Dentre essas, destacam-se a acentuação das desigualdades educacionais oriundas das diferentes condições socioeconômicas

^a Universidade Federal do Ceará. Fortaleza, Ceará, Brasil.

Recebido em: 05 mai. 2016

Aprovado em: 01 ago. 2017

e da diversidade étnica presente naquele alunado, bem como o estreitamento do currículo que deveria ser ministrado, aspecto contemplado no presente artigo.

A proeminência desse tipo de exames no sistema educacional daquele país teve significativo impulso após aprovação e implementação nacional do programa *No Child Left Behind*, sobre o qual versará a seção a seguir.

2 *No Child Left Behind*: a estratégia tecnocrática para o salto educacional

Às voltas com um clima nacional de apreensão por conta de uma suposta estagnação no nível educacional de seus alunos de ensino básico, em comparação com outras nações desenvolvidas, as ideias que remetem a um controle mais rígido do desempenho acadêmico, por meio de testes padronizados de larga escala, tornam-se, no início do novo milênio, hegemônicas naquela formação social, culminando com a aprovação do programa federal *No Child Left Behind* (Nenhuma Criança Deixada para Trás; doravante NCLB) pelo Congresso em 2001 e com sua promulgação como lei federal¹ pelo então presidente George W. Bush em 8 de janeiro de 2002 (UNITED STATES OF AMERICA, 2002).

Em uma sucinta descrição do programa, podemos assinalar que os estados eram estimulados (mediante verbas federais) a implantar seus próprios testes e a adotar três níveis de desempenho (tais como **básico**, **proficiente** e **avançado**, por exemplo) de acordo com seus critérios. Todos os discentes da 3^a à 8^a série fariam provas de leitura e matemática, anualmente, naquelas séries do ensino fundamental e, uma vez, no ensino médio. Suas notas seriam desagregadas por etnia, *status* socioeconômico, condições de incapacidade física e de inglês como segundo idioma, com o intuito de monitorar o progresso de cada um desses grupos de estudantes. Os estados ficavam obrigados a apresentar, anualmente, uma “prestação de contas”, mostrando a variação (positiva ou negativa) das notas de seu alunado, referentes ao ano anterior², bem como um cronograma estabelecendo quando e como todos aqueles grupos de educandos alcançariam a meta de 100% de proficiência nessas duas matérias³ até 2014 (RAVITCH, 2010).

A lei previa que qualquer unidade escolar que não fizesse “progressos adequados” para cada subgrupo, com vistas à meta de 100% de proficiência, seria rotulada

¹ Trata-se da “lei de responsabilidade educacional” daquele país.

² Descrito na legislação do programa como *adequate yearly progress* (progresso anual adequado, doravante AYP). Trata-se de uma medida de acompanhamento anual das notas da escola ou do distrito escolar (USA, 2002).

³ Vide seção “O ardil da meta impossível”, no presente artigo.

como uma escola “em necessidade de melhoria” e enfrentaria uma série de sanções anualmente cada vez mais onerosas. Em um primeiro ano nessa condição, a instituição seria colocada em “aviso prévio”. No segundo ano, seria obrigada a oferecer a todos os seus alunos o direito de transferência para outra unidade escolar, com transporte pago pelo distrito a que pertencesse à sua congênera “em necessidade de melhoria” (descontado diretamente da dotação repassada pelo Governo Federal). No terceiro ano, a instituição seria obrigada a oferecer aulas gratuitas para estudantes de baixa renda, custeadas da mesma forma que o transporte retromencionado. No quarto ano, a unidade escolar teria de realizar “ações corretivas”, que poderiam significar mudanças curriculares, mudanças de pessoal, ou aumento na carga horária diária de aulas, ou do ano escolar. Se um educandário não conseguisse atingir suas metas para qualquer subgrupo durante cinco anos consecutivos, ele seria obrigado a se “reestruturar”, com cinco opções, tendo que optar por uma das seguintes alternativas: converter-se em uma escola *charter*⁴; substituir o diretor e toda a equipe; abrir mão do controle para a gestão privada; entregar o controle da escola para o estado; ou “qualquer outra grande reestruturação da governança da unidade escolar” (A maioria dos estados e distritos terminou por escolher a última, a mais ambígua, na esperança de evitar as outras sanções) (RAVITCH, 2010).

O NCLB exigia⁵ que todos os estados participassem dos exames do *National Assessment of Educational Progress* (doravante NAEP), os quais passariam a testar todos os estudantes em leitura e matemática, nas 4^a e 8^a séries, a cada dois anos (USA, 2002). Antes do NCLB, a participação dos estados nos exames do NAEP era voluntária e os testes não eram administrados bienalmente, bem como suas notas, que não tinham consequências para qualquer estudante, escola ou distrito, serviam apenas como auditoria externa para monitorar o progresso dos estados no cumprimento de suas metas, pois o exame era amostral, e não censitário (RAVITCH, 2010).

A ideia por trás do NCLB era que as desigualdades de oportunidades de educação seriam reduzidas ou desapareceriam, caso os alunos fossem forçados a fazer constantemente testes padronizados, as escolas desagregassem os dados para mostrar as notas obtidas pelos discentes de diferentes subgrupos, e os distritos interviessem com ações específicas focadas nos educandários de menor desempenho naqueles exames. No entanto, o amplo consenso observado entre a maioria dos

⁴ Escolas *charter* são financiadas com recursos públicos, mas administradas de forma privada por organizações com ou sem fins lucrativos. Para aprofundamento, conferir Ravitch (2013), especialmente seu capítulo 16.

⁵ O NCLP foi oficialmente extinto em 10 de dezembro de 2015, com a revogação da lei que o criou em 2002, tendo em vista a promulgação da *Every Student Succeeds Act* (ESSA), diploma legal que instituiu o novo programa educacional do Governo Federal dos Estados Unidos (USA, 2015).

pesquisadores, como se verá a seguir, deixa patente que testes padronizados são ineficazes e mesmo contraproducentes quando utilizados como instrumento de avaliação suficiente para conduzir qualquer reforma educacional.

A próxima seção abordará o massificado uso desses limitados mecanismos de aferição de aprendizagem e sua instrumentalização no processo de responsabilização punitiva imposto a professores e diretores escolares no contexto da “reforma” educacional, ora em vigor naquela nação.

3 Testes: a ferramenta “por excelência” do programa

O NCLB descortinou uma nova era no funcionamento das escolas públicas estadunidenses, marcada pela obsessão em aferir o desempenho acadêmico dos educandos e responsabilizar os professores pelos resultados, por meio da utilização de testes padronizados de larga escala (de aplicação censitária), a ponto de comprometer a quantidade de tempo destinada à própria instrução, principalmente dos discentes desfavorecidos. Nesse sentido, pontuam Guisbond, Neill e Schaeffer (2012) que

Os testes e a preparação para testes têm se proliferado – a quantidade de tempo gasto em testes em algumas escolas dobrou. Um estudo para o Congresso, feito pelo *Government Accountability Office* (Escritório Governamental de *Accountability*), estimou que os estados teriam que criar mais de 433 testes (a um custo entre US\$ 1,9 bilhões e 5,3 bilhões entre 2002 e 2008) para satisfazer as exigências da lei NCLB. Isso tornou-se apenas a ponta do *iceberg* de um aumento maciço no uso de testes. Não é incomum de 20 a 60 dias letivos por ano serem gastos em preparação, além dos dias gastos nos testes em si, que são consideráveis. [...] Um prejuízo considerável é o desvio do tempo de ensino destinado aos alunos que mais necessitam. A *Wisconsin Association for Supervision and Curriculum Development* (Associação Wisconsin para Supervisão e Desenvolvimento Curricular) tentou quantificar o tempo de aprendizagem perdido com os testes em geral e para os alunos com necessidades especiais. Eles descobriram que os professores de Wisconsin gastaram uma média por distrito de 976 horas [anuais] administrando testes. Isso foi particularmente prejudicial para alunos com necessidades especiais⁶ (p. 4–5).

⁶ Todas as traduções do presente trabalho foram feitas por seu autor, que optou pela tradução livre.

Malgrado a objeção ultrajada de grande parcela de educadores e pais de alunos, os quais organizaram protestos e demandaram ações judiciais contra tal política, esses tipos de exames têm constituído a ferramenta “por excelência” do programa NCLB com vistas à operacionalização de seus fundamentos e ao alcance de seus objetivos.

Importa, não obstante, assinalar que o questionamento fundamental de professores, diretores, acadêmicos e pais de estudantes que se opõem aos testes padronizados de alto impacto, não se refere à avaliação em si, a qual reputam necessária e salutar para o aperfeiçoamento do próprio sistema, mas ao uso que se faz desses exames na tentativa de responsabilizar ou encontrar culpados por resultados acadêmicos insatisfatórios apresentados por educandos. Ravitch (2010) desnuda as limitações intrínsecas desses instrumentos para aferir a assimilação de conhecimentos por parte dos alunos, razão pela qual os especialistas no assunto frequentemente alertam as autoridades educacionais para o fato de que esse tipo de teste nunca deve ser utilizado de forma isolada para se tomar decisões importantes sobre discentes, docentes ou escolas (NATIONAL RESEARCH COUNCIL, 1999).

Nas seções posteriores, procurar-se-á expor as premissas, a fundamentação, a implementação e as consequências da adoção dessas avaliações standardizadas como elemento nuclear na estruturação da *accountability*⁷ educacional daquele país.

4 O ardil da meta impossível

SEÇÃO. 1111. PLANOS ESTADUAIS. (b) PADRÕES ACADÊMICOS, AVALIAÇÕES ACADÊMICAS E *ACCOUNTABILITY*. (2) *ACCOUNTABILITY*, (F): CRONOLOGIA – Cada estado deve estabelecer um cronograma para o progresso anual adequado. O cronograma deve assegurar que, **no mais tardar, 12 anos após o fim do ano escolar de 2001-2002, todos os alunos em cada grupo descrito no subparágrafo (C) (V) alcançarão ou excederão o nível de proficiência de desempenho acadêmico** estadual nas avaliações estaduais nos termos do parágrafo 3 (USA, 2002, p. 23–24, grifos nossos).

A letra da lei, expressa no trecho descrito na epígrafe, não deixa margem para qualquer dúvida: a totalidade dos alunos em todos os estados daquele país deveria ter 100% de proficiência em matemática e leitura até o final do ano escolar de

⁷ Conquanto esse termo seja predominantemente traduzido na mídia e na academia como “prestação de contas”, adota-se, no presente trabalho, bebendo em Afonso (2009), uma perspectiva de articulação entre suas três dimensões constitutivas, a saber, **avaliação, prestação de contas e responsabilização**, entendido que nenhuma delas, isoladamente, reflete a organicidade que caracteriza o termo em questão, razão pela qual optamos por sua não tradução.

2013/2014, o que, efetivamente, não ocorreu. A desmesuradamente elevada meta e as consequências para os professores de seu não atingimento suscitaram manifestações por parte daqueles, bem como de diretores e estudiosos de assuntos educacionais, no sentido de sua completa inexecutabilidade. De forma cabal, sustenta Ravitch (2013):

Vamos ser claros: 100% de proficiência é uma meta impossível; nenhuma nação no mundo já conseguiu isso, e nenhuma nação aprovou qualquer legislação para punir suas escolas por não atingir uma meta inatingível. Era [sic] como se o Congresso aprovasse uma lei dizendo que cada cidade na América deveria tornar-se totalmente livre de crime. Quem poderia reprovar uma meta tão louvável? Qual cidade não gostaria de ser livre de crime? Mas imagine se a lei fixasse um prazo de 12 anos e dissesse que qualquer cidade que não cumprisse a meta seria punida; suas delegacias seriam fechadas e privatizadas; seus policiais perderiam seus cargos. As primeiras a fechar seriam as delegacias de polícia nos bairros mais pobres, onde os índices de criminalidade são mais altos. Eventualmente, isso ocorreria mesmo nos bairros afluentes, porque nenhuma cidade é completamente livre de crime. **Desejar que possa ser assim, ou aprovar leis para exigir que assim seja, não faz com que o seja** (p. 16, grifos nossos).

Um dos desdobramentos lógicos da imposição autoritária de qualquer meta inatingível é a proliferação de todo tipo de expediente de questionável validade ética com vistas a simular o alcance daquela. Previsivelmente, foi o que ocorreu em todo o país, com a entrada em vigor do NCLB, utilizando-se as autoridades educacionais estaduais e municipais de um variado cardápio de subterfúgios para satisfazer, formalmente, os requisitos de proficiência insculpidos em sua lei, alguns dos quais expomos sucintamente a seguir.

A maioria dos estados, aproveitando-se da margem de discricionariedade concedida por aquele diploma legal, que delegava àquelas instâncias federativas o desenvolvimento de seus parâmetros curriculares, a escolha dos testes e a própria definição de proficiência, estabeleceu os níveis desta em escala consideravelmente inferior ao correspondente padrão federal, gerando, destarte, notas de corte mais baixas nos exames, com o intuito de “aprovar” maior número de estudantes e cumprir a meta prescrita na norma legal⁸. Isso fez com que a maioria dos

⁸ Uma rigorosa comparação entre os níveis de proficiência estaduais e os do NAEP conduzida pelo *National Center for Educational Statistics* (NCES) (Centro Nacional de Estatísticas Educacionais), uma agência interna do Departamento de Educação dos Estados Unidos, revelou que apenas dois dos 50 estados daquela nação, Massachusetts e Wyoming, definiram sua nota de corte para a proficiência no 4º série em matemática no mesmo nível da nota de corte correspondente utilizada pelo NAEP (NATIONAL CENTER FOR EDUCATION STATISTICS, 2007).

estados apresentasse estimulantes “progressos”, quase anualmente. O artifício ardiloso, contudo, era facilmente perceptível, bastando comparar o desempenho dos educandos naqueles testes com o obtido pelos mesmos discentes nos testes federais do NAEP para que se percebesse o embuste⁹.

Outra forma de gerar resultados mais elevados em testes, porém, sem que seja necessário baixar a nota de corte, consiste em elaborar provas mais fáceis, artifício certamente adotado na maioria dos estados, mas de difícil comprovação empírica.

Não obstante o recurso a esses mecanismos manipulatórios, Ravitch (2010) nos informa que, no ano letivo de 2006/2007, cerca de 25.000 escolas não alcançaram o AYP, tendo esse número aumentado em 2007/2008 para em torno de 30.000, ou 35,6% de todas as escolas públicas. Interessante observar que nessa preocupante estatística estão incluídas mais da metade das escolas públicas de Massachusetts, um dos dois únicos estados (o outro era Wyoming) que, conforme mencionado, estabeleceram notas de corte no nível do NAEP, o qual é nacionalmente aclamado pela excelência do nível acadêmico de seus discentes, os quais invariavelmente lideram os *rankings* daquele rigoroso exame nacional. Tal enigmática situação, com efeito, reforça sobremaneira os indícios, apontados por uma vasta gama de pesquisadores, da instrumentalização do estabelecimento da inatingível meta com o objetivo velado de engendrar condições políticas propícias à expansão das escolas *charter* naquele país¹⁰.

Mais um ardil sagazmente empregado por diretores com vistas a melhorar a nota dos alunos de determinada escola consiste em evitar a presença de estudantes “problemáticos”, considerados mais difíceis de ensinar, geralmente oriundos de famílias de baixa renda ou filhos de imigrantes, os quais têm o inglês como segundo idioma, além daqueles com necessidades especiais. Nesse estratagema, a exclusão desses estudantes se dá no próprio processo de admissão à escola¹¹ ou *a posteriori*¹², por meio do “endurecimento dos códigos disciplinares”, elaborados em consonância com tais objetivos, ou mediante outros artifícios, tais

⁹ O caso paradigmático, já quase “um clássico” na literatura, é o do estado de Mississippi, cujas autoridades educacionais jactavam-se de possuir 89% dos seus alunos da 4ª série no nível ou acima da proficiência em leitura, quando nos correspondentes testes do NAEP tal nível de proficiência era obtido por apenas 18% do mesmo corpo discente (HULL, 2008).

¹⁰ Burriss (2013a, 2013b) desvela a manipulação na atribuição da nota de corte que estabeleceria o nível de proficiência dos alunos da 3ª à 8ª série nos testes do *Common Core* (correspondente, no Brasil, à Base Nacional Comum em leitura e matemática), no estado norte-americano de Nova Iorque.

¹¹ No afã de se livrar de alunos com tal perfil, os quais, no estreito entendimento dessas administrações escolares, diminuem as notas da escola ou do distrito, utiliza-se todo tipo de subterfúgio para eliminá-los já na admissão. Para aprofundamento, conferir Ravitch (2010).

¹² Orel (2003) relata como 522 estudantes na cidade de Birmingham, no estado do Alabama, foram afastados do ensino médio poucos dias antes das provas finais.

como suspender aqueles alunos, às vésperas do dia dos testes, ou encaminhá-los indevidamente para o subsistema de educação especial, com o intuito de removê-los de um subgrupo no qual sua nota baixa possa impedir esse grupo de atingir o AYP (RAVITCH, 2010).

Outro mecanismo, mais refinado e mais difícil de ser identificado, consiste em alterar a forma de cômputo das notas desses testes, operando uma *sui generis* “lógica aritmética”, ilustrativamente descrita no caso empírico a seguir: entre 2006 e 2009, o estado de Nova Iorque sutilmente promoveu uma alteração no modo de calcular a nota mínima de proficiência em matemática e em inglês, de forma que um estudante que em 2006 precisaria acertar 59,6% das questões para atingir aquela nota, três anos depois precisaria acertar apenas 44,0% para ser considerado proficiente (NEW YORK STATE EDUCATION DEPARTMENT, 2015). O bom senso entende que dificilmente, em qualquer sistema escolar sério, tal pontuação possa ensejar a proficiência de um educando.

As tentativas de burla do sistema chegam a ultrapassar os limites da própria legalidade jurídica, com a disseminação de todo tipo de fraude envolvendo docentes, discentes e diretores na administração e na correção de testes, passando por vazamento de questões e “vista grossa” na aplicação dos testes, consoante informam Guisbond, Neil e Schaeffer (2012):

Uma epidemia de trapanças tomou conta da nação. Em Atlanta, onde foram confirmadas fraudes em 44 escolas públicas, envolvendo 178 professores e diretores, o relatório do Bureau de Investigações da Geórgia (GBI) descreveu uma cultura de “medo, intimidação e retaliação espalhados por todo o distrito”. No final de 2011, investigadores da Geórgia divulgaram outro relatório documentando a generalizada “cola” em provas no município de Dougherty, 200 milhas ao sul de Atlanta. Eles encontraram evidências de fraude em cada uma das 11 escolas do município e evidências de professores coagidos a corrigir como corretas as respostas erradas dos alunos. [...]. **Motivo número um: “Pressão para atender ao progresso anual adequado” exigido pela lei** No Child Left Behind (p. 7, grifo nosso).

Longe de constituírem casos isolados ou idiossincráticos¹³, esses lamentáveis expedientes consubstanciam o resultado previsível ante a pressão para atender aos irracionais requisitos da legislação em questão. Na ótica de Ravitch (2013):

¹³ “[...] incidentes de fraude nos últimos três anos têm sido confirmados em 30 estados e no Distrito de Columbia [...]” (GUISBOND; NEILL; SCHAEFFER, 2012, p. 5).

Dada a importância dos resultados dos testes, não é de estranhar que os professores e diretores criem várias formas de burlar o sistema de testes, isto é, truques e atalhos para alcançar os resultados desejados, sem a melhoria da educação. Quando o objetivo do teste é informativo e de diagnóstico, não há nenhuma razão para professores e administradores alterarem os resultados, exceto por meio do aprimoramento do processo de ensino e aprendizagem. Mas, quando o objetivo do teste é somente a prestação de contas, então os professores e os administradores entendem que há consequências reais se as pontuações em sua sala de aula ou em sua escola mudarem. Se as notas subirem, eles podem receber um bônus considerável; se elas descerem, sua escola será estigmatizada, e eles podem perder seus empregos. A intensa pressão gerada por exigências de prestação de contas leva muitos educadores e diretores a aumentarem as notas de maneira que nada têm a ver com a aprendizagem (p. 109).

A seção seguinte problematizará precisamente essa concepção de *accountability* adotada pelo NCLB, analisando as premissas e os efeitos de sua implantação no sistema público de educação básica daquela nação.

5 Recompensa e punição: o *modus operandi* do sistema de *accountability* do *No Child Left Behind*

O uso de punições, mesmo se referido eufemisticamente como incentivos negativos, sanções ou consequências, cria um clima de medo, e o medo gera raiva e ressentimento. Ele também leva as pessoas a mudarem para o modo de “controle de danos” e a agir com mais cautela. Seres humanos simplesmente não pensam de forma criativa e não alcançam a excelência quando se sentem ameaçados. A filosofia por trás dos testes de alto impacto tem sido habilmente satirizada por um cartaz em alguns escritórios e salas de aula: “OS ESPANCAMENTOS CONTINUARÃO ATÉ QUE O MORAL MELHORE”. Quando os professores são privados de segurança no emprego e de aumentos salariais, em um esforço para melhorar seus desempenhos, eles geralmente se tornam desmoralizados em vez de motivados (KOHN, 2000, p. 97).

Emergem progressivamente mais nítidas as intrínsecas limitações de tais ferramentas, com vistas à operacionalização de um sistema de *accountability* destinado a impor graves consequências àqueles de quem pretende medir o

desempenho, as habilidades e até mesmo os conhecimentos. As fragilidades técnicas desses instrumentos para desenvolver e aferir dimensões cognitivas em seres humanos, cuja complexidade mostra-se infensa a tais pretensões de mensuração¹⁴, manifestam-se não apenas nas amplas margens de erro estatístico ou de variação aleatória de seus resultados, mas em todas as fases do desenvolvimento, da aplicação e até mesmo da avaliação dos testes, instâncias essas operadas por seres humanos, portanto suscetíveis a todos os tipos de enganos ou incorreções, ocasionados de forma deliberada ou não¹⁵ (BRAUN, 2013).

Ravitch (2010) aprofunda a análise sobre as limitações desses testes:

O problema com a responsabilização com base em testes é que ela impõe sérias consequências sobre as crianças, os educadores e as escolas com base em notas que podem refletir erros de medição, erros estatísticos, variação aleatória, ou uma série de fatores ambientais ou atributos dos estudantes. Nenhum de nós gostaria de ser avaliado, com a nossa reputação e meios de subsistência em risco, somente por meio de um instrumento que é propenso a erros e ambiguidades. Os testes atualmente em uso não são suficientes por si só para a tarefa de medir a qualidade das escolas ou dos professores. Eles foram projetados para fins específicos: para medir se os alunos podem ler e podem executar operações matemáticas; mesmo nessas tarefas, eles devem ser usados com consciência de suas limitações e de sua variabilidade. Eles não foram projetados para capturar as dimensões mais importantes da educação, para as quais não existe medida (p. 117).

Do binômio “recompensa e punição”, *modus operandi* vigente no NCLB, no que concerne ao primeiro termo, a perspectiva empresarial, hegemônica na burocracia educacional estadunidense, entende, não obstante farta evidência teórica e empírica em contrário disponível na literatura científica concernente à área¹⁶, que os professores se esforçarão mais e obterão melhor desempenho, caso lhes sejam

¹⁴ Campbell (2010) instituiu uma máxima nas ciências sociais que ficou conhecida como “Lei de Campbell”, pela qual: “[q]uanto mais qualquer indicador social quantitativo for usado para tomada de decisões sociais, mais ele será sujeito a pressões corruptoras e mais propenso ele estará a distorcer e a corromper os processos sociais que ele deve monitorar”. Nesse sentido, entendemos que essa descoberta científica de Campbell explica, de forma lapidar, as fraudes e as tentativas de burla do sistema de accountability instituído pelo NCLB. Congruentemente, manifesta-se Koretz (2008), bem como Rothstein, Jacobsen e Wilder (2008).

¹⁵ As limitações dos testes padronizados com vistas à aferição da efetiva aprendizagem dos alunos são contempladas em seção subsequente do presente trabalho.

¹⁶ Recomendamos reportar-se a Fleming (2011), Glazerman e Seifullah (2012), Sparks (2011), Springer et al. (2011) e Viadero (2009), dentre outras inúmeras fontes, copiosidade essa que, tendo em vista as limitações de escopo do presente trabalho, pode-se apenas, como ora se procede, indicar ilustrativamente.

oferecidos benefícios e estímulos materiais suficientemente atraentes. Um dos modelos de remuneração por desempenho que seguem essa filosofia é o *merit pay* (pagamento por mérito), o qual, pelo menos na educação dos Estados Unidos, já foi tentado inúmeras vezes desde a segunda década do século passado, sempre com resultados decepcionantes¹⁷. Ravitch (2013, p. 99) refere-se a ele como “a ideia que nunca funciona e nunca morre”, aludindo à insistência em sua aplicação visando ao estímulo de educadores, desconsiderando (ou desconhecendo) a básica premissa de que a própria natureza da profissão docente remete a uma motivação calcada eminentemente em aspectos **intrínsecos à atividade**, tendo os fatores extrínsecos (prêmios financeiros, principalmente) menores peso e consistência na motivação dos professores, em comparação com o que ocorre com profissionais de outras áreas. Segundo Harris (2007):

[...] professores podem ser, de alguma maneira, menos sensíveis a prêmios em dinheiro do que os trabalhadores de outras profissões. Os próprios professores e grande parte do público em geral veem a docência como uma profissão mal paga. Em estudos recentes, 75% de novos professores concordam que eles são seriamente mal pagos, e 78% de recentes graduados na faculdade concordam que os professores são seriamente mal pagos. Dada a impressão de uma perspectiva de baixa remuneração, seria surpreendente se aqueles que escolhem a docência fossem fortemente motivados pelo pagamento (p. 15).

O autor salienta que os docentes, ao optarem por sua profissão, não o fazem seduzidos por remuneração elevada, razão pela qual os prêmios em dinheiro não exercem no magistério a atração que exercem em outras profissões, o que é corroborado por Ravitch (2013, p. 98, grifo nosso), quando esta observa que “[...] quem se torna um professor **com a expectativa de fazer dinheiro** não está fazendo uma sábia escolha de carreira. Ele ou ela estaria melhor em outra área”.

As argumentações de Harris (2007) e Ravitch (2013) não contradizem, todavia, a pertinência e a necessidade de uma remuneração justa, digna e mesmo atraente aos professores, em um nível tal que propicie o suprimento adequado de docentes qualificados de que o país precisa. O que se questiona não é o interesse dos professores por bons salários, mas a premissa empresarial de que, para se obter maior produtividade desses profissionais, basta acrescentar alguns benefícios salariais variáveis provisórios (bônus) ao seu salário. Com efeito, o docente

¹⁷ Vide Murnane e Cohen (1986) em consistente análise desse sistema de remuneração na educação estadunidense abrangendo todo o século XX.

deve auferir remuneração suficientemente justa e satisfatória, de forma que possa desempenhar suas atividades de docência sem se sentir pressionado pelos constrangimentos que as carências materiais impingem àqueles que auferem rendimentos injustos e mesmo insuficientes para seu sustento e o de suas famílias.

Sobre as deficiências desse modelo de remuneração, acrescenta ainda aquela historiadora da educação:

O paradoxo do pagamento por mérito em educação é que, mesmo que ele funcionasse, ele ainda seria um fracasso. Quanto mais os professores e as escolas são compelidos a se concentrarem somente em elevar as notas dos testes, mais eles serão suscetíveis a limitar o currículo, mais os distritos e escolas irão burlar o sistema para inflar as notas, mais trapaça ocorrerá e mais os professores procurarão evitar estudantes que tiram notas baixas. Então, na medida em que se promete recompensar as escolas pela elevação das notas nos testes (e puni-las pela não elevação dessas notas), a qualidade da educação é que vai sofrer. As escolas reduzirão o tempo para as artes, para a história, para as ciências, para as línguas estrangeiras, para a educação física e para qualquer outra disciplina que não seja testada. Ou então testarão tudo, o que reduziria o tempo de instrução ainda mais do que apenas testando habilidades básicas (RAVITCH, 2013, p. 99, grifo nosso).

Cassettari (2008), por sua vez, critica a aplicação indiscriminada desses mecanismos de gestão privada em instituições públicas de ensino¹⁸:

[...] uma vez que seus objetivos são muito diferentes, senão opostos. A lógica do setor privado é a de aumentar a produtividade dos trabalhadores para aumentar o lucro das empresas. Essa lógica não pode ser simplesmente transportada para as escolas, já que o produto do trabalho dos professores não é facilmente identificado ou mensurado. Além disso, por trabalharem com seres humanos, os professores não podem descartar os ‘produtos’ que apresentarem resultados diferentes dos esperados [como ocorre, por exemplo, em uma linha de montagem de uma indústria] (p. 4–5).

¹⁸ Com vistas a um aprofundamento sobre a questão do pagamento de adicional remuneratório por desempenho aos professores, reportar-se a Santos (2004).

Isso não significa, entretanto, que estímulos da espécie não possam ser empregados em situações atípicas, tais como para atrair professores para atuar em regiões inóspitas, ou para ensinar matérias nas quais haja escassez de profissionais habilitados no mercado de trabalho, além de outras circunstâncias cuja excepcionalidade justifique a adoção de tal mecanismo. Contudo, mesmo nesses casos em que os docentes auferem uma remuneração adicional, fatores de outra natureza mostram-se sobremaneira relevantes no que concerne à motivação desses profissionais, como explica Ravitch (2010):

[...] Os professores, como outros profissionais, precisam sentir-se competentes, eficazes e admirados. Saber que eles estão mudando as vidas de seus alunos, um por um, é uma fonte de satisfação, certamente. Se os professores são tratados com desdém pelos administradores, trabalham em prédios mal conservados, são-lhes atribuídas grandes turmas de alunos mal preparados, são confrontados por educandos indisciplinados e obrigados a cumprir metas irrealistas, eles certamente não desfrutarão de um senso de satisfação pessoal e profissional (p. 135).

O problema é que, mesmo que se encontrasse um mecanismo de motivação e estímulo do corpo docente adequado e eficaz, não se pode olvidar que os objetivos programáticos do NCLB, **necessária e legalmente**, vinculam a materialização do êxito de qualquer escola ou professor ao sucesso dos seus alunos **nos testes**¹⁹. Subentende-se que, **se o professor está “devidamente estimulado”, automaticamente fará o discente aprender**, como se o único fator a atuar na aprendizagem escolar fosse a instrução fornecida pelo docente²⁰, **o qual converterá tal aprendizado em notas elevadas nas avaliações estandardizadas**.

Essa é precisamente a premissa fundamental do *Value-added Model* (Modelo de valor agregado), ferramenta estatística que se propõe a aferir quantitativamente, com base nos resultados dos testes de seus alunos, a contribuição do professor no desempenho escolar de seus discentes. A seguir, tratar-se-á das limitações dos testes padronizados de larga escala.

¹⁹ O desempenho dos discentes no NCLB é aferido exclusivamente por meio dos testes padronizados de alto impacto, os quais não possuem nem a abrangência nem a idoneidade necessárias para contemplar em toda sua complexidade as dimensões cognitivas do desenvolvimento intelectual do educando, sem falar nas igualmente relevantes dimensões não cognitivas.

²⁰ A maioria das pesquisas relevantes e consistentes na área mostram que o professor somente é responsável por cerca de 10% da variação no desempenho dos estudantes. Nesse sentido, reportar-se à American Statistical Association (MORGANSTEIN; WASSERSTEIN, 2014), Goldhader, Brewer e Anderson (1999), Haertel (2013) e Nye, Konstantopoulos e Hedges (2004).

6 As limitações intrínsecas dos testes padronizados de aprendizagem

Prevalece no senso comum o entendimento de que os testes padronizados de aprendizagem são instrumentos científicos confiáveis tão objetivos e precisos como um termômetro ou qualquer outro dispositivo utilizado para medições no âmbito da ciência. Nada mais ilusório. Examinemos algumas das limitações desse tipo de avaliação discente. A primeira limitação refere-se à abrangência do conteúdo testado. Tais tipos de testes conseguem atingir apenas uma pequena amostra de um amplo campo cognitivo ou de uma determinada disciplina ministrada em uma série escolar específica. Por essa razão, a nota do educando em um teste dessa espécie não consegue espelhar o seu nível de aprendizagem de todo o conteúdo ministrado. Explica-se: segundo Popham (1999), o intuito dos profissionais que elaboram a grande maioria dos testes padronizados de larga escala é criar um instrumento de avaliação **referenciado à norma**, que, com apenas um punhado de itens, produza interpretações válidas do domínio do aluno a respeito de uma parte substancial do conteúdo **em comparação com o grupo normativo**²¹.

Referenciado à norma, termo cunhado por Glaser (1963), contrapõe-se a **referenciado ao critério** (*norm-referenced e criterion-referenced*, respectivamente). Segundo Brito (1997):

[...] [b]asicamente, há duas formas de analisar dados sobre o rendimento dos alunos. A primeira seria uma interpretação dos resultados tomando-se por base um referencial de desempenho, representado pelos objetivos de ensino e respectivos critérios de alcance. Esse tipo de interpretação descreve quais os objetivos que o aluno atingiu dentre aqueles que lhe foram propostos, sem relação a qualquer outro desempenho, seja individual ou de turma, série ou grupo. [...] Esse tipo de medida "referenciada ao critério" é tido como uma medida de padrão absoluto, cuja base é o desempenho individual. Em contraposição, a medida "referenciada à norma" baseia-se no desempenho do grupo, seguindo um padrão relativo. Por essa forma de análise, pode-se determinar os desempenhos, discriminando-os entre si. Na construção dos instrumentos, deve-se atentar para que os itens maximizem as diferenças e promovam uma grande variabilidade de resultados (p. 136–137).

Os itens que fazem melhor esse trabalho de **diferenciar os alunos** são aqueles respondidos corretamente por cerca de metade deles. Evitam-se itens a que possam

²¹ "[...] uma amostra nacional de estudantes da mesma idade ou nível escolar" (POPHAM, 1999, p. 9).

ter respondido corretamente muitos ou poucos discentes, o que não contribuiria para **espalhar os resultados dos testes dos estudantes**, que é precisamente o propósito desse tipo de avaliação²².

A necessidade de diferenciação das notas dos discentes, para atender aos requisitos desse tipo de avaliação, elimina itens em que os estudantes teriam um bom desempenho. No entanto, esses itens geralmente cobrem precisamente o conteúdo que, devido à sua importância, os professores priorizam. Assim, por mais que os docentes privilegiem esses conhecimentos e habilidades importantes, o menos provável é que questões que cubram esse conteúdo sejam contempladas em um teste de desempenho referenciado à norma. Dessa forma, **avaliar a eficácia do ensino dos professores, utilizando instrumentos de avaliação de seus alunos²³ que evitam deliberadamente conteúdos importantes, é fundamentalmente insensato**.

Observe-se, ademais, outro efeito desfavorável da adoção desse tipo de avaliação: com vistas a espalhar as notas dos alunos e premiados pela limitação do número de itens, os elaboradores desses testes têm uma predileção toda especial por questões que exijam do educando uma habilidade intelectual inata, não facilmente adquirida na escola²⁴. Todavia, itens que essencialmente medem as diferenças de capacidades intelectuais inatas dos educandos, ou do que eles aprenderam fora da escola, obviamente, não contribuem para inferências válidas sobre como as crianças foram ensinadas **na escola**, assim não poderiam ser utilizados para avaliar nem a escola, nem o professor. Perceba-se, igualmente, o viés social e de classe embutido na questão: indubitavelmente, os discentes oriundos de famílias de maior capital intelectual e social têm probabilidade bem maior do que os das classes inferiores de acertar esse tipo de item.

Kohn (2000) esclarece ainda que:

²² A maioria das empresas de testes padronizados de larga escala utilizam as avaliações **referenciadas à norma**, em contraposição às **referenciadas ao critério**. Nos últimos anos, vem crescendo o número de empresas que vêm adotando a metodologia da **teoria de resposta ao item**, sendo a Pearson, a maior do mundo no segmento, uma delas. No Brasil, o Exame Nacional do Ensino Médio (Enem) vem utilizando essa metodologia em suas últimas edições, a qual, segundo o pesquisador estadunidense Walter Stroup, igualmente padece das mesmas limitações da **referenciada à norma** (SMITH, 2012).

²³ Essa é a premissa fundamental do Value-added Model (Modelo de valor agregado). Para aprofundamento, vide Adler (2014), American Statistical Association (MORGANSTEIN, D.; WASSERSTEIN, 2014), Amrein-Beardsley (2008), Darling-Hammond et al. (2011), Hewitt (2015), Kupermintz (2003), Ravitch (2013) e Rothstein (2011).

²⁴ A título de ilustração, segue o exemplo do item citado por Popham (1999, p. 13), uma questão de um teste de estudos sociais aplicado a uma turma da 6ª série: "Se alguém quer realmente conservar os recursos, uma boa maneira de fazer isso é: a) deixar luzes acesas mesmo se elas não forem necessárias; b) lavar pequenas quantidades de roupa em vez de grandes quantidades em uma máquina de lavar; c) escrever em ambos os lados de uma folha de papel; d) colocar jornais utilizados no lixo" (a resposta certa é a letra C).

Não importa quantos alunos façam o teste, não importa quão bem ou mal eles sejam ensinados, não importa quão difícil sejam as perguntas, o padrão de resultados é garantido para ser o mesmo: exatamente 10% daqueles que fazem o teste pontuarão entre os 10% com maior escore, e a outra metade vai sempre cair abaixo da mediana. Isso não é porque nossas escolas estão falhando; isso é por causa do que significa “mediana”. Uma boa pontuação em um teste referenciado à norma significa “melhor do que outras pessoas”, mas não sabemos mesmo **quanto** melhor. Pode ser que as pontuações reais de cada um sejam todas muito semelhantes, caso em que as distinções entre elas são sem sentido – um pouco como dizer que eu sou a pessoa mais alta no meu quarteirão, embora eu seja apenas meio centímetro mais alto do que a pessoa mais baixa no meu quarteirão (p. 77, grifos nossos).

Tal característica das avaliações referenciadas à norma leva a outra séria deficiência, que consiste na questão de que, como bem salienta o autor supra, pelo fato de que toda distribuição de notas sempre conterá um patamar mais baixo, vai aparecer sempre que algumas crianças estão indo muito mal. Isso, por sua vez, reforça a sensação de que as escolas estão falhando. Pior, isso contribui para a insidiosa suposição de que algumas crianças simplesmente não podem aprender – especialmente se as mesmas crianças sempre parecem ficar abaixo da mediana (KOHN, 2000).

Por fim, observamos que os testes padronizados não conseguem mensurar as habilidades não cognitivas interpessoais²⁵, tampouco as individuais, tais como as capacidades de escutar e de refletir, o controle dos impulsos, a paciência, a persistência, a empatia, a curiosidade, a criatividade, o pensamento crítico, a resiliência, a autodisciplina, a desenvoltura, as responsabilidades social e ambiental, a dimensão cívica e outras de mesma índole e importância social. Ambos os tipos de capacidades não cognitivas, as interpessoais e as individuais, interagem com as habilidades cognitivas para o desenvolvimento integral do ser humano.

A seção a seguir tratará do estreitamento curricular resultante da adoção de políticas focadas na instrumentalização dessas avaliações estandardizadas como esteio para legitimar a segregação educacional, ocupacional e socioeconômica dos estudantes, do início ao final de sua vida escolar.

²⁵ As capacidades interpessoais referem-se à forma como nos relacionamos com os outros em um ambiente social e inclui aspectos como colaboração, respeito, amizade, liderança, comunicação, solidariedade e partilha, entre muitas outras características (LEVIN, 2013).

7 Adestramento para testes padronizados e estreitamento curricular: a diferença entre treinar e educar

Um dos efeitos mais daninhos da instituição dos testes padronizados de larga escala como ferramenta fulcral das políticas educacionais de viés empresarial tem sido a subtração do tempo anteriormente destinado à instrução, com o fito de intensificar a preparação para aqueles testes, os quais se transmutaram de meio de aferição (com todas as suas limitações) do conteúdo efetivamente ministrado em objetivo último de todo o processo de ensino-aprendizagem, seu zênite supremo. Para atingi-lo, não se medem as consequências da marginalização das disciplinas do campo das humanidades e mesmo das ciências naturais. Todavia, tal procedimento tem causando um estreitamento curricular que tem engendrado um corpo discente hábil na aplicação de técnicas e “bizus” com vistas a se sair bem naqueles testes, todavia ignorante de conhecimentos básicos daquelas disciplinas menosprezadas, não obstante essenciais para a conformação de um adulto minimamente instruído para o convívio nas sociedades contemporâneas. Ravitch (2010), a partir de Montefinise (2007), ilustra a situação:

Na cidade de Nova Iorque, professores disseram a uma jornalista que eles **eliminaram estudos sociais, arte e ciências por um mês antes dos testes estaduais de leitura e matemática para se concentrarem nas atividades de preparação para tais exames**. Uma professora disse que **seus alunos não sabem quem era o presidente durante a Guerra Civil, mas eles podem te dizer como eliminar respostas em um teste de múltipla escolha**. “E, enquanto nossos resultados dos testes estiverem altos, todos estarão felizes”. Seu diretor ordenou-lhe que “esquecesse de tudo, exceto da preparação para o teste”. Outra professora disse que os diretores são parcialmente avaliados de acordo com os resultados dos testes, por isso “naturalmente, eles querem as notas altas, [e] essa é a nossa prioridade. Educação verdadeira é secundário” (RAVITCH, 2010, p. 79, grifos nossos).

O episódio é paradigmático da forma como as escolas vêm comprometendo o ensino de disciplinas indispensáveis a uma sólida formação na desesperada tentativa de apresentar notas elevadas nos testes estaduais²⁶. Trata-se, com efeito, de uma distorção do próprio propósito original dos testes, qual seja, avaliar a efetiva

²⁶ Para mais exemplos desse modus operandi, que constitui a regra, e não a exceção, reportar-se ilustrativamente a Fox e Hacker (2008), McMurrer (2008), McMurrer e Kober (2007) e Perlstein (2007). Para acesso a evidências consistentes e abundantes dos efeitos do estreitamento curricular no sistema educacional estadunidense, remetemos a Au (2007), Meyer (2005) e Morton e Dalton (2007), dentre outros autores.

aprendizagem, e não tentar “a ferro e fogo” produzir notas altas, independentemente da real e consistente assimilação do conteúdo por parte do aluno.

O aspecto possivelmente irônico ou paradoxal da questão consistiria na constatação de que o foco estreito, apenas nas habilidades básicas no idioma nacional e em matemática, é incompatível com o desenvolvimento de um “capital humano” criativo, preparado para o desenho de novas tecnologias, dotado de habilidades de interação interpessoal para trabalhar em equipe e outras da mesma natureza²⁷, características essenciais na conformação do trabalhador (“colaborador”, no sádico e cruel jargão empresarial) de perfil adequado ao que Sennet (2006) cunha de **novo capitalismo**. Tal incongruência minaria ou dificultaria o suprimento da força de trabalho apta a laborar na decantada “economia do conhecimento”. Nossa hipótese, a qual exploramos em pesquisa em andamento, é a de que inexistente tal paradoxo, porquanto são outras as necessidades do capital na presente etapa histórica de seu ciclo de reprodução, no que se refere à exploração do **trabalho vivo**, na acepção marxiana.

Ravitch (2010, 2013), lembrando que o propósito essencial da escola pública, a razão pela qual ela recebe fundos públicos, é a formação dos jovens para o exercício da cidadania, advoga que toda escola deve ter um currículo rico e balanceado, incluindo artes, ciências, história, literatura, civismo, geografia, línguas estrangeiras e educação física²⁸. Observa a pesquisadora:

Hoje, os formuladores de políticas pensam em educação apenas em termos dos seus propósitos secundários. Eles falam das crianças como futuros concorrentes mundiais. Eles, às vezes, referem-se às crianças, em termos bastante infelizes, como “ativos humanos”, esquecendo-se que elas são pessoas únicas e que não são fungíveis. Eles querem que todos os alunos estejam “prontos para a faculdade e para a carreira”. Eles tendem a falar apenas de preparação para o mercado de trabalho, e não da educação para a cidadania. Mas isso é um engodo. A formação da força de trabalho pode ocorrer nas escolas ou no local de trabalho. Isso não é sem importância. Também não é a preparação para a

²⁷ “Pessoas inovadoras não podem vir de escolas que forcem os alunos a memorizar respostas corretas em testes padronizados ou recompensam aqueles que se destacam em regurgitar conhecimento mastigado” (ZHAO, 2009, p. vii). Na mesma perspectiva, assevera o autor, em outra obra, que “[...] o que gera ótimas notas em testes pode arruinar as qualidades empreendedoras. Os testes padronizados e o foco em memorização mecânica, por exemplo, são talvez os maiores inimigos da capacidade empreendedora” (ZHAO, 2012, não paginado).

²⁸ Argumenta a historiadora, na primeira dessas duas obras, que, não obstante o currículo não seja a “varinha de condão” que resolverá, por um passe de mágica, todos os problemas educacionais, não tê-lo indica uma inabilidade ou mesmo uma falta de vontade de definir o que se está tentando realizar, tendo como efeito o movimento de andar em círculos e não chegar a lugar nenhum. Mas o pior de todos os efeitos, salienta a autora, é deixar que as editoras de livros didáticos decidam o que as crianças devem aprender.

faculdade sem importância. Mas **preparar para a faculdade não é o objetivo central da educação. Nem o é o treinamento da força de trabalho. O objetivo central da educação é preparar todos para assumirem os direitos e responsabilidades da cidadania em uma democracia.** (RAVITCH, 2013, p. 182, grifo nosso).

Os efeitos do estreitamento curricular, todavia, não atingem igualmente os alunos de todas as classes sociais, sendo os oriundos de famílias de menor *status* socioeconômico e de etnias minoritárias os mais prejudicados, como explicam Finn e Ravitch (2007):

Precisamos de escolas que preparam as crianças para se destacar e competir não apenas no mercado de trabalho global, mas também como participantes de pleno direito em nossa sociedade, nossa cultura, nossa política e nossa economia. [...] Abandonar as disciplinas de humanidades em nome das ciências duras [STEM²⁹] sozinhas também tende a alargar as divisões sociais e aprofundar as desigualdades internas. Os ricos que entendem o valor da aprendizagem das disciplinas de humanidades podem ser os únicos capazes de pagá-las para seus filhos. Escolas privadas de elite e alguns sistemas escolares dos subúrbios ficarão com a educação em sentido lato, assim como as faculdades de elite. Crianças ricas vão estudar filosofia e arte, música e história, enquanto seus colegas pobres preencherão bolinhas em folhas de teste. Os poucos sortudos darão origem à próxima geração de magnatas, líderes políticos, inventores, autores, artistas e empresários. As massas menos sortudas terão oportunidades mais estreitas. Alguns não encontrarão oportunidade nenhuma, e a frustração vai tentá-los a saquear os mais afortunados, que, por sua vez, irão trancar-se em condomínios fechados, clubes exclusivos e “privado isso – privado aquilo”, alargando assim as clivagens internas e piorando as nossas perspectivas para a coesão social e a civilidade³⁰ (não paginado).

²⁹ Acrônimo para Science, Technology, Engineering and Math (Ciência, Tecnologia, Engenharia e Matemática).

³⁰ Trazendo a questão para a realidade brasileira, Freitas (2012) contextualiza o dano que tal processo inflige à nossa juventude pobre: “O argumento para justificar a limitação ao básico é que os outros aspectos mais complexos dependem de se saber o básico primeiro. Um argumento muito conhecido no âmbito do sistema capitalista e que significa postergar para algum futuro não próximo a real formação da juventude, retirando dela elementos de análise crítica da realidade e substituindo-se por um ‘conhecimento básico’, um corpo de habilidades básicas de vida, suficiente para atender aos interesses das corporações e limitado a algumas áreas de aprendizagem restritas (usualmente leitura, matemática e ciências). A consequência é o **estreitamento curricular** focado nas disciplinas testadas e o esquecimento das demais áreas de formação do jovem, em nome de uma **promessa futura: domine o básico e, no futuro, você poderá avançar para outros patamares de formação. Todos sabem que a juventude mais pobre depende fundamentalmente da escola para aprender; e, se à sua passagem pela escola for limitada às habilidades básicas, nisso se resumirá sua formação**” (p. 389–390, grifos nossos).

Na sociedade estadunidense, a forte resistência dos pais, dos professores e dos alunos³¹ levou o Governo Federal a encerrar o NCLB (o qual, como se esperava, não atingiu a insana meta de 100% de proficiência de todos os discentes do ensino básico nas disciplinas testadas), com a promulgação de uma nova lei educacional, *Every Student Succeeds Act* (ESSA – Lei Todo Estudante Vencerá), a qual destina aos estados maior protagonismo na condução de sua política de educação, ao permitir que essas unidades federativas estabeleçam múltiplas medidas de aferição do desempenho acadêmico de seus alunos, possibilitando assim que os testes deixem de constituir a única forma de avaliação do aprendizado dos educandos (BROWN, 2015; FOSTER, 2016). Cabe, agora, àquele corpo social a vigilância com vistas a impedir que a obsessão por esses tipos de testes continue a prejudicar o alunado daquele país, notadamente os vitimados pela desigualdade e pela segregação socioeconômica e étnica.

³¹ Notadamente por meio de massivos boicotes (Opt out) aos testes estaduais. Cf. Foster (2016).

The role of standardized testing in the policy for basic education in the United States

Abstract

With the advent of the federal education program known as No Child Left Behind, a disproportionate increase in hours devoted to testing, narrowing of curricula and vertical assessments focusing on accountability of teachers and principals began to shape the agenda for educational reform in the United States. This article, developed through bibliographical and documentary research, contests the assumptions underlying these policies, which are strongly structured around standardized large-scale testing, known to be the "tool par excellence" of the government program in question. Besides perverse effects forced upon the students, this paper exposes the intrinsic limitations of these instruments in order to measure the assimilation of knowledge by the students, as well as the proliferation of all kinds of expedients of questionable ethical validity with the view to simulate the reach of academic performance targets set by the mentioned program.

Keywords: Standardized tests. Educational policy. United States.

El papel de las pruebas estandarizadas en la política educativa para la enseñanza básica en los Estados Unidos

Resumen

Con el advenimiento del programa federal No Child Left Behind, un aumento desproporcional de las horas destinadas a pruebas, el estrechamiento de contenidos curriculares y las evaluaciones con verticalidad, enfocando la responsabilidad de profesores y directores, pasaron a conformar la agenda para la reforma educativa en los Estados Unidos. El presente artículo, desarrollado por medio de investigación bibliográfica y documental, cuestiona las premisas que fundamentan esas políticas, fuertemente estructuradas en torno a las pruebas estandarizadas a gran escala, que constituyen la "herramienta por excelencia" del programa gubernamental en cuestión. Además de los perversos efectos impuestos a los educandos, se muestran las limitaciones intrínsecas de esos instrumentos para medir la asimilación de conocimientos por parte de los discentes, así como la proliferación de todo tipo de expedientes de cuestionable validez ética con miras a simular el alcance de las metas de desempeño académicos estipulados por el mencionado programa.

Palabras clave: Pruebas estandarizadas. Política educativa. Estados Unidos.

Referências

- ADLER, M. Review of measuring the impacts of teachers. *Think Twice Review National Education Policy Center Report*, 2014. Boulder, CO: National Education Policy Center. Disponível em: <http://greatlakescenter.org/docs/Think_Twice/TT_Adler_TeachImpact.pdf>. Acesso em: 22 mar. 2016.
- AFONSO, A. J. Nem tudo o que conta em educação é mensurável ou comparável: crítica à *accountability* baseada em testes estandardizados e *rankings* escolares. *Revista Lusófona de Educação*, Lisboa, v. 13, n. 13, p. 13–29, 2009. Disponível em: <<http://revistas.ulusofona.pt/index.php/rleducacao/article/view/545>>. Acesso em: 29 abr. 2017.
- AMREIN-BEARDSLEY, A. Methodological concerns about the education value-added assessment system. *Educational Researcher*, v. 37, n. 2, p. 65–75, 2008. doi: <https://doi.org/10.3102/0013189X08316420>
- AU, W. High-stakes testing and curricular control: a qualitative metasynthesis. *Educational Researcher*, v. 36, n. 5, p. 258–267, 2007. doi: <https://doi.org/10.3102/0013189x07306523>
- BRAUN, H. Value-added modeling and the power of magical thinking. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 21, n. 78, p. 115–130, 2013. doi: <http://dx.doi.org/10.1590/S0104-40362013000100007>
- BRITO, M. S. T. Norma e critério de desempenho como parâmetros da avaliação da aprendizagem. *Estudos em Avaliação Educacional*, São Paulo, n. 15, p. 135–198, 1997. doi: <http://dx.doi.org/10.18222/ae01519972276>
- BROWN, E. How schools would be judged under ‘Every Student Succeeds’, the new No Child Left Behind. *The Washington Post*, [S.l.], 30 Nov. 2015. Disponível em: <<https://www.washingtonpost.com/news/education/wp/2015/11/30/how-schools-would-be-judged-under-every-student-succeeds-the-new-no-child-left-behind/>>. Acesso em: 27 abr. 2016.
- BURRIS, C. How come officials could predict new test score results? *The Washington Post*, [S.l.], 12 Aug. 2013a. Disponível em: <<https://www.washingtonpost.com/blogs/answer-sheet/wp/2013/08/12/how-come-officials-could-predict-results-on-new-test-scores>>. Acesso em: 27 fev. 2016.
- _____. What big drop in new standardized test scores really means. *The Washington Post*, [S.l.], 7 Aug. 2013b. Disponível em: <<https://www.washingtonpost.com/news/answer-sheet/wp/2013/08/07/what-big-drop-in-new-standardized-test-scores-really-means/>>. Acesso em: 27 fev. 2016.

CAMPBELL, D. T. Assessing the impact of planned social change. *Journal of MultiDisciplinary Evaluation*, v. 7, n. 15, p. 3–43, 2010.

CASSETTARI, N. Pagamento por *performance* na educação básica. In: REUNIÃO ANUAL DA ANPED. 31., 2008, Caxambu. *Anais...* Caxambu: Anped, 2008. p. 1–15.

DARLING-HAMMOND, L. et al. Getting teacher evaluation right: A background paper for policy makers. [S. l]: American Educational Research Association; National Academy of Education. 2011. Disponível em: <<http://files.eric.ed.gov/fulltext/ED533702.pdf>>. Acesso em: 24 mar. 2016.

FINN, C. E.; RAVITCH, D. Not by geeks alone. *The Washington Post*, [S.l.], 8 Aug. 2007. Disponível em: <<http://www.wsj.com/articles/SB118653759532491305>>. Acesso em: 10 abr. 2016.

FLEMING, N. Some efforts on merit pay scaled back. *Education Week*, Bethesda, v. 31, n. 4, p. 1–18, 2011.

FOSTER, J. B. The opt out revolt: democracy and education. *Monthly Review*, New York, v. 67, n. 10, p. 1, 2016.

FOX, L.; HACKER, H. K. Dallas-Fort worth students struggle with TAKS' Shortresponse Written Test. *The Dallas Morning News*, Dallas, 20 Jul. 2008. Disponível em: <<http://www.dallasnews.com/sharedcontent/dws/news/localnews/stories/072008dnmettaksscores.443a381.html>>. Acesso em: 10 abr. 2016.

FREITAS, L. C. Os reformadores empresariais da educação: da desmoralização do magistério à destruição do sistema público de educação. *Educação e Sociedade*, Campinas, v. 33, n. 119, p. 379–404, 2012. doi: <http://dx.doi.org/10.1590/S0101-73302012000200004>

GLASER, R. Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist*, v. 18, n. 8, p. 519–521, 1963. doi: <http://dx.doi.org/10.1037/h0049294>

GLAZERMAN, S.; SEIFULLAH, A. *An evaluation of the Chicago teacher Advancement Program (Chicago TAP) after four years*. Washington, DC: The Joyce Foundation, 2012.

GOLDHADER, D. D.; BREWER, D. J.; ANDERSON, D. J. A three-way error components analysis of educational productivity. *Education Economics*, v. 7, n. 3, p. 199–208, 1999. doi: <https://doi.org/10.1080/09645299900000018>

GUISBOND, L.; NEILL, M.; SCHAEFFER, B. NCLB's lost decade for educational progress: what can we learn from this policy failure?. *Educação e Sociedade*, Campinas, v. 33, n. 119, p. 405–430, 2012. doi: <http://dx.doi.org/10.1590/S0101-73302012000200005>

HAERTEL, E. H. Reliability and validity of inferences about teachers based on student test scores. Washington, DC: ETS, 2013. Disponível em: <<https://www.ets.org/media/research/pdf/picang14.pdf>>. Acesso em: 24 fev. 2016.

HARRIS, D. C. The promises and pitfalls of alternative teacher compensation approaches. Arizona: Arizona State University, 2007.

HEWITT, K. K. Educator Evaluation Policy that Incorporates EVAAS Value-Added Measures: undermined intentions and exacerbated inequities. *Education Policy Analysis Archives*, v. 23, n. 76, p. 1–49, 2015. doi: <http://dx.doi.org/10.14507/epaa.v23.1968>

HULL, J. *The proficiency debate: a guide to NAEP achievement levels*. Alexandria, VA : The Center for Public Education, 2008. Disponível em: <<http://www.centerforpubliceducation.org/main-menu/evaluating-performance/the-proficiency-debate-at-a-glance/the-proficiency-debate-a-guide-to-naep-achievement-levels.html>>. Acesso em: 31 dez. 2015.

KOHN, A. *The schools our children deserve: Moving beyond traditional classrooms and “tougher standards”*. Boston: Houghton Mifflin Harcourt, 2000.

KORETZ, D. M. *Measuring up*. Harvard: Harvard University, 2008.

KUPERMINTZ, H. Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis*, v. 25, n. 3, p. 287–298, 2003. doi: <https://doi.org/10.3102/01623737025003287>

LEVIN, H. M. The limits of test score comparisons. Unesco Office in Santiago: *Newsletter LLEC*. [S.l.], 13 Jun 2013. Disponível em: <http://www.unesco.org/new/en/santiago/press-room/single-new/news/the_limits_of_test_score_comparisons_henry_m_levin_teachers_college_columbia_university/#vutgendw7iu>. Acesso em: 12 mar. 2016.

MORGANSTEIN, D.; WASSERSTEIN, R. ASA statement on using value-added models for educational assessment. *Statistics and Public Policy*, v. 1, n. 1, 2014. doi: <http://dx.doi.org/10.1080/2330443X.2014.956906>

McMURRER, J. *Instructional time in elementary schools: a closer look at changes for specific subjects*. Washington, DC: Center on Education Policy, 2008.

_____. KOBER, N. *Choices, changes, and challenges: curriculum and instruction in the NCLB era*. Washington, DC: Centre on Education Policy, 2007.

MEYER, L. The complete curriculum: ensuring a place for the Arts in America's Schools. *Arts Education Policy Review*, v. 106, n. 3, p. 35–39, 2005.

MONTEFINISE, A. Lost lessons in test-prep craze. *The New York Post*, [S.l.], 28 Jan. 2007. Disponível em: <<http://nypost.com/2007/01/28/lost-lessons-in-test-prep-craze/>>. Acesso em: 6 abr. 2016.

MORTON, B. A.; DALTON, B. National Center for Educational Statistics, and Institute of Education Sciences (US). *Changes in instructional hours in four subjects by public school teachers of grades 1 through 4*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, 2007.

MURNANE, R.; COHEN, D. Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Educational Review*, Harvard, v. 56, n. 1, p. 1–18, 1986. doi: <https://doi.org/10.17763/haer.56.1.18q2334243271116>

NATIONAL CENTER FOR EDUCATION STATISTICS. *Mapping 2005 state proficiency standards onto the NAEP scales*. National Assessment of Educational Progress. Washington, DC: Department of Education, 2007.

NATIONAL RESEARCH COUNCIL. *High stakes: testing for tracking, promotion, and graduation*. Washington, DC: National Academies, 1999. doi: <https://doi.org/10.17226/6336>

NEW YORK STATE EDUCATION DEPARTMENT (NYSED). *English Language Arts (ELA) and Mathematic Assessment Results*. New York: NYSED, 2015. Disponível em: <www.emsc.nysed.gov/irts/ela-math/>. Acesso em: 31 dez. 2015.

NYE, B.; KONSTANTOPOULOS, S.; HEDGES, L. V. How large are teacher effects? *Educational Evaluation and Policy Analysis*, v. 26, n. 3, p. 237–257, 2004. doi: <https://doi.org/10.3102/01623737026003237>

OREL, S. Left behind in Birmingham: 522 pushed-out students. In: LENT, R. C.; PIPKIN, G. (Ed.). *Silent no more: voices of courage in American schools*. Portsmouth: Heinemann, 2003. p. 1–14.

PERLSTEIN, L. *Tested: one American school struggles to make the grade*. New York: Henry Holt, 2007.

POPHAM, W. J. Why standardized test scores don't measure educational quality. *Educational Leadership*, v. 56, n. 6, p. 8–15, 1999.

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

_____. *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York: Basic, 2013.

ROTHSTEIN, J. *Review of learning about teaching*. Boulder, CO: National Education Policy Center, 2011.

ROTHSTEIN, R.; JACOBSEN, R.; WILDER, T. *Grading education: Getting accountability right*. Washington, DC: Economic Policy Institute, 2008.

SANTOS, L. L. C. P. Formação de professores na cultura do desempenho. *Educação e Sociedade*, Campinas, v. 25, n. 89, p. 1145-1157, 2004.

SENNET, R. *A cultura do novo capitalismo*. Rio de Janeiro: Record, 2006.

SMITH, M. A serious design flaw is suspected in state tests. *The New York Times*, [S.l.] 28 Jul. 2012. Disponível em: <http://www.nytimes.com/2012/07/29/education/texas-studies-suggest-test-design-flaw-in-taks.html?_r=0>. Acesso em: 14 mar. 2016.

SPARKS, S. Study leads to end of New York City merit-pay program. *Education Week*, July 20, 2011. Disponível em: <http://blogs.edweek.org/edweek/inside-school-research/2011/07/a_new_study_by_the.html>. Acesso em: 9 jan. 2016.

SPRINGER, M. G. et al. Teacher pay for performance: experimental evidence from the Project on Incentives in Teaching (Point). Society for Research on Educational Effectiveness, 2011. Disponível em: <<http://files.eric.ed.gov/fulltext/ED518378.pdf>>. Acesso em: 9 jan. 2016.

UNITED STATES OF AMERICA (USA). *No Children Left Behind Act*. Washington, DC: US Department of Education, Jan. 2002. Disponível em: <<http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>>. Acesso em: 15 dez. 2015.

UNITED STATES OF AMERICA (USA). *President Obama signs the every student succeeds act*. Washington, DC: The White House, 2015. Disponível em: <<https://www.whitehouse.gov/photos-and-video/video/2015/12/10/president-obama-signs-every-student-succeeds-act>>. Acesso em: 22 dez. 2015.

VIADERO, D. Texas merit-pay pilot failed to boost student scores, study says. *Education Week*, Nov. 4, 2009. Disponível em: <http://blogs.edweek.org/edweek/inside-school-research/2009/11/texas_merit-pay_pilot_failed_t.html>. Acesso em: 9 jan. 2016.

ZHAO, Y. *Catching up or leading the way: American education in the age of globalization*. Alexandria, VA: ASCD, 2009.

_____. Double-Think: The Creativity-Testing Conflict. *Education Week*, v. 31, n. 36, p. 26–32. 2012. Disponível em: <http://www.edweek.org/ew/articles/2012/07/18/36zhao_ep.h31.html>. Acesso em: 07 jun. 2017.



Informações do autor

Remo Moreira Brito Bastos: Doutor e Mestre em Educação pela Universidade Federal do Ceará (UFC). Servidor Público Federal (IBGE, Analista). Contato: remomoreira@gmail.com