

Qualidade da educação: avaliação, indicadores e metas

Reynaldo Fernandes
Amaury Gremaud

Este capítulo procura apresentar os diferentes objetivos e possibilidades dos sistemas de avaliação educacional, bem como identificar o atual estágio da avaliação da educação básica no Brasil. Ênfase é dada às mudanças recentes, em especial àquelas introduzidas pelo Plano de Desenvolvimento da Educação (PDE), que expandiram o papel da avaliação em larga escala para além de seu objetivo tradicional de diagnóstico dos sistemas educacionais e colocaram-na como um dos pilares da política educacional do Ministério da Educação.¹

Nesse sentido, as três principais inovações foram: i) a incorporação dos objetivos de *accountability*; ii) a criação de um indicador sintético da qualidade da educação básica, que considera tanto o desempenho dos estudantes em exames padronizados quanto a progressão desses alunos no sistema; e iii) a definição de metas tanto para o país quanto para cada sistema e escola em particular.²

Atualmente, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP/MEC) divulga resultados de avaliações por redes e escolas, sendo o Índice de Desenvolvimento da Educação Básica (IDEB) o principal indicador utilizado para monitorar a qualidade da educação básica. O IDEB, que combina as notas da Prova Brasil/Saeb com as taxas de aprovação, visa coibir tanto a reprovação indiscriminada como a prática de aprovar alunos que nada aprenderam. O IDEB foi, também, utilizado para estabelecer as metas para redes e escolas e, assim, propiciar uma movimentação nacional para que, até 2021, o Brasil atinja o estágio educacional atual dos países desenvolvidos.

O apoio à divulgação de resultados de avaliações educacionais, por redes e escolas, não é unânime. Embora as experiências com estes procedimentos proliferem em todo o mundo, e vários estudos apontem que suas vantagens superam seus possíveis defeitos, alguns ainda resistem à idéia. Os estudos mostram que, de modo geral, onde tais medidas foram adotadas o desempenho dos estudantes tendeu a crescer de forma mais acelerada. Não há evidências claras de exclusão de estudantes de baixo rendimento.³ Evidentemente, os resultados dos programas podem variar significativamente a depender do

¹ Assim, a ênfase do capítulo não é o PDE, mas as avaliações. Mesmo na questão da avaliação, outra mudança não será destacada: a introdução da Provinha Brasil: avaliação diagnóstica realizada a critério das redes na fase inicial do processo de aprendizagem a fim de se implementar ações o mais cedo possível. Para uma discussão do PDE, ver Haddad (2008).

² O termo *accountability* tem sido traduzido como transparência, responsabilização, rendição de contas e outros. Na falta de concordância sobre a melhor tradução, manteremos o termo em inglês.

³ Enquanto os estudos têm documentado que sistemas de *accountability* elevam o desempenho dos estudantes, alguns pesquisadores apontam que as escolas são incentivadas a adotar medidas para aumentar a pontuação dos estudantes que não significam mudanças na qualidade de ensino (treinamento para o exame, seleção de alunos para o teste etc). Nesse sentido, a principal preocupação é a exclusão de alunos de baixa proficiência. No entanto, as evidências não indicam que isso seja uma prática generalizada. Ver, por exemplo, Carnoy e Loeb (2002), Ladd e Zelli (2002), Jacob e Levitt (2003), Figlio e Winicki (2005), Hanushek e Raymond (2005), Jacob (2005), Chakrabarti (2006, 2008), Cullen e Reback (2006), Figlio e Getzler (2006), Figlio e Rouse (2006), Rouse et al. (2007), Neal e Schanzenbach (2007) e Krieg (2008).

desenho, da gestão, do ambiente etc. No Brasil, essas medidas são muito recentes, o que impossibilita uma avaliação dos seus resultados.

Assim, o presente capítulo se restringe a apresentar as medidas adotadas, clarificar sua lógica e destacar como os problemas usuais desse tipo de programa têm sido combatidos. O artigo está dividido em seis partes incluindo esta introdução. Na segunda discutem-se os potenciais e riscos dos sistemas de avaliação, com ênfase para sua função de *accountability*. A terceira analisa a incorporação dos objetivos de *accountability* com a introdução da Prova Brasil e a divulgação dos resultados do ENEM por escola. Na quarta parte discute-se a questão dos indicadores de qualidade da educação, sua adequação ao contexto brasileiro e a introdução do IDEB. Na quinta a discussão ocorre em torno dos sistemas de metas associados às avaliações educacionais. Por fim acrescentam-se algumas observações finais a título de conclusão do texto.

Avaliações educacionais em larga escala e *accountability*

As avaliações educacionais podem ser vistas como um caso particular das avaliações de programas e/ou políticas e, deste modo, estão relacionadas à idéia de resultados. Independente do contexto em que elas estão inseridas e de seus objetivos, as avaliações educacionais ou buscam aferir resultados passados ou inferir resultados futuros. Isso não significa que elas sejam sempre conduzidas com base em indicadores de resultados. Por exemplo, alguém que deseje avaliar uma escola e julgue que os indicadores de resultados educacionais, como aqueles obtidos em exames padronizados, são limitados e imprecisos pode preferir avaliá-la com base nos insumos de que ela dispõe (infra-estrutura, formação dos professores etc) e nos processos utilizados (método de ensino, estrutura curricular etc). Nesse caso, a idéia é que obtendo informações sobre insumos e processos-chaves é possível inferir sobre os resultados que serão alcançados pela escola.

O debate sobre avaliações orientadas por resultados e avaliações orientadas por insumos e processos é, em grande parte, um debate sobre medida. Caso duas medidas, uma baseada em resultados e outra baseada em insumos e processos, convergissem, no sentido de apresentarem a mesma ordenação de escolas, a medida baseada em insumos e processos deveria ser preferida. Isso porque ela dá mais retorno às escolas, pois, além de dar um diagnóstico da qualidade, sinaliza quais os pontos mais vulneráveis que deveriam ser atacados. Essa convergência, no entanto, não tende a ocorrer na prática.

Diversos estudos têm apontado para a fraca correlação (condicional nas características observáveis dos estudantes) entre as variáveis de insumos e processos e o desempenho dos estudantes em exames padronizados. Essa fraca correlação tem sido a base para as críticas de medidas de qualidade da escola baseadas em insumos e processos, pois com ressaltam Hanushek e Raymond “uma escola poderia ser recompensada por melhorar seus processos mesmo que isso não resulte em ganhos no desempenho dos estudantes”.⁴ Esse tipo de argumentação, no entanto, pressupõe que essa fraca correlação decorre da

⁴ Hanushek e Raymond (2003), pp 195.

deficiência das variáveis de insumos e processos em explicar a qualidade da escola e não da debilidade dos resultados dos testes em medir essa qualidade.

É preciso ter clareza de que o debate sobre avaliação de escolas e sistemas educacionais se dá em dois níveis de análise bastante distintos. O primeiro, de ordem mais conceitual, refere-se aos principais resultados que devemos esperar das escolas. Já o segundo, de ordem mais técnico-operacional, diz respeito à obtenção de indicadores para medir tais resultados.

No primeiro nível, o problema advém do fato das escolas possuírem múltiplos objetivos, enquanto as medidas de resultados utilizadas se detêm em apenas alguns deles. Por exemplo, grande parte das avaliações se limita a leitura e matemática e muitos críticos as consideram limitadas. Ainda que as avaliações ampliem seu escopo, incluindo outras disciplinas, elas tendem a se concentrar no desenvolvimento cognitivo dos estudantes e, segundo esses críticos, ignoram aspectos fundamentais como, por exemplo, ética, responsabilidade social, preservação ambiental e diversas habilidades não cognitivas que seriam importantes para formação dos estudantes. Por outro lado, esses críticos tendem a considerar os objetivos da escola tão gerais e abstratos (e.g. formar cidadãos comprometidos em construir uma sociedade democrática) que torna impossível a construção de instrumentos para medi-los. Na impossibilidade de elaborar instrumentos para aferir se as escolas estão cumprindo esses objetivos mais amplos, o melhor que se poderia fazer é imaginar que tipo de escola teria capacidade de concretizá-los e, portanto, avaliar as escolas existentes com base neste modelo. Isso envolve avaliar insumos e, principalmente, processos existentes, comparando-os com aqueles considerados ideais.⁵

Outros críticos até concordam que o aprendizado - por exemplo, em leitura e matemática - possa ser considerado o objetivo principal da escola, mas argumentam que os resultados dos exames padronizados não servem para medir o trabalho das escolas em proporcionar tal aprendizado. Uma escola que recebe alunos carentes, com escassa bagagem cultural, pode realizar um excelente trabalho e, ainda assim, não ter seus alunos entre aqueles com melhores resultados. Se um teste de matemática se propõe avaliar a capacidade das escolas em proporcionar aos seus estudantes um bom aprendizado na disciplina, ele é imperfeito porque seus resultados incorporam, além do esforço da escola, influências advindas da família, dos amigos e das habilidades natas dos estudantes, bem como do componente de erro aleatório de medida.⁶

Tentar identificar, nos resultados obtidos pelos estudantes nos exames, a parcela que advém do trabalho da escola tem sido a base para os Modelos de Valor Adicionado (*Value-Added Models* – VAMs). Infelizmente, isso não é uma tarefa fácil. A literatura de VAMs apresenta uma diversidade de modelos, envolvendo medidas relativamente simples de valor adicionado pela escola a complexos modelos estatísticos, que necessitam admitir questionáveis hipóteses

⁵ Para uma crítica aos objetivos parciais das avaliações baseadas em exames padronizados ver, por exemplo, Paro (2000a).

⁶ Kane e Staiger (2001 e 2002) notaram que os resultados de exames são sujeitos a erros de medida que variam muito, particularmente entre as pequenas escolas.

não testáveis.⁷ Além disso, medidas simples de valor adicionado tendem a produzir instabilidade de *rankings*, em virtude da grande variância do erro aleatório de medida.⁸ Por outro lado, os modelos mais sofisticados exigem uma quantidade de dados (e.g. testar todos os alunos, todos os anos e nas mesmas disciplinas) que são disponíveis apenas em poucos sistemas de avaliação no mundo.

É preciso reconhecer que todas as medidas utilizadas em avaliações educacionais são imperfeitas, por não levarem em conta todos os aspectos que se gostaria de considerar e por medir com erros (sistemáticos e/ou aleatórios) os aspectos que elas se propõem avaliar. Se isso vem a ser um problema grave ou não, depende da extensão dessas imperfeições e, principalmente, do uso que se pretende dar às avaliações. Embora boa parte dos problemas acima levantados já estivesse presente há muito tempo nas avaliações educacionais, eles não costumavam criar tanta polêmica como nos dias de hoje. A polêmica atual está ligada à nova função que os exames assumem nas políticas de *accountability*, cujo cerne é considerar não apenas os alunos, mas escolas, professores, diretores e gestores como responsáveis pelo desempenho dos estudantes.

Tradicionalmente as avaliações do aprendizado eram restritas às salas de aula, a partir de exames elaborados pelos próprios professores. O principal objetivo era verificar se os estudantes possuíam condições de avançar no sistema. Aqueles que não atingissem um padrão pré-determinado eram considerados não aptos e eram reprovados e isso não era visto como de responsabilidade do professor ou da escola. Avaliações de escolas tampouco eram incomuns. Elas costumavam ocorrer por meio dos processos de inspeção. Nesse caso, no entanto, elas se restringiam a averiguar as condições de funcionamento das escolas, se os professores cumpriam suas obrigações funcionais, se os currículos ministrados eram adequados etc. Ou seja, limitava-se a avaliar insumos e processos.

Com o crescimento dos exames em larga escala, a partir da década de 1950, as avaliações ganharam um papel adicional: a de diagnosticar e monitorar a qualidade do sistema educacional. Um marco nas avaliações diagnósticas é dado pelo relatório Coleman, que tinha como objetivo estudar a segregação racial no sistema educacional dos EUA. Coleman e seus colaboradores examinaram diferenças nos resultados de exames padronizados e procuraram explicá-las com base nas diferenças de recursos das escolas e das características sócio-econômica dos alunos. Esse trabalho é importante por uma série de razões, entre as quais a mudança de foco na forma de se identificar a qualidade das escolas: de insumos e processos para resultados.⁹

Hoje, praticamente, todos os países desenvolvidos e muitos em desenvolvimento contam com um sistema desse tipo. De modo geral, eles possuem escalas que permitem a comparação dos resultados entre diferentes anos e séries. Até o final dos anos 80, as avaliações tinham, usualmente, base amostral, com resultados divulgados de forma bastante agregada. É só a partir do

⁷ Para uma revisão e discussão desses modelos ver, por exemplo, McCaffrey et al. (2003) e o simpósio do *Journal of Education and Behavioral Statistics* (JEBS, 2004) sobre VAMs.

⁸ Ver Kane e Staiger (2001 e 2002) e Mizala et al. (2007).

⁹ Ver Coleman et al. (1966).

final dos anos 80 que as avaliações com fins de *accountability* ganham corpo, tendo como marco a reforma educacional inglesa de 1988.

Os chamados programas de *accountability* educacional diferem em uma série de aspectos.¹⁰ Entretanto, alguns elementos comuns podem ser destacados:

1. Ênfase no aprendizado dos alunos, o qual pode ser aferido por testes padronizados. Os indicadores para *accountability* podem incluir outras informações (e.g. frequência, evasão e alguma medida da qualidade e quantidade dos insumos escolares), mas o destaque é dado aos resultados dos testes padronizados.
2. Professores, diretores e gestores são co-responsáveis pelos resultados de seus estudantes e, assim, eles devem, ao menos em parte, responder pelo desempenho dos estudantes nos exames.
3. Presença de um sistema de incentivos para que os responsáveis pela educação dos estudantes moldem suas ações com o objetivo de que seus alunos obtenham bons resultados nos exames. Esse sistema de incentivos envolve, necessariamente, a publicidade dos resultados dos testes por unidade de *accountability* (sistemas educacionais, escolas e, em alguns casos, até por professores individuais). O sistema de incentivos pode conter, ou não, prêmios e punições atrelados aos resultados dos alunos.

Não é fácil realizar uma análise geral dos programas de *accountability*, dada a diversidade de desenhos e ambientes onde eles estão inseridos. Entretanto, acreditamos que na perspectiva de implementar um programa com essas características, se deveria procurar responder quatro questões básicas:

- Os objetivos estabelecidos pelos programas de *accountability* são adequados?
- Qual é o potencial de sucesso dos programas em atingir seus objetivos?
- Quais são os riscos?
- Quais os resultados das experiências pioneiras nesse tipo de programa?

i) Adequação dos objetivos

Vimos que tais programas buscam verificar se a aprendizagem proporcionou aos estudantes determinadas competências e conhecimentos. É importante ressaltar que os programas de *accountability* surgem nos países desenvolvidos, onde o problema de acesso à escola e permanência nela está, praticamente, resolvido. A defasagem idade-série não é, também, um problema sério nesses países. Em países onde esses problemas não estão resolvidos, enfatizar os resultados dos exames pode ser precipitado.

A idéia que o principal objetivo da escola é o aprendizado em algumas disciplinas básicas tem sido uma importante fonte de críticas aos programas de *accountability*. É difícil discordar da alegação de que os objetivos dos programas existentes são limitados, frente ao potencial das escolas em contribuir para a formação de seus estudantes. Entretanto, essa não é, a nosso ver, a forma correta

¹⁰ Ver Elmore et alli (1996), Kane e Staiger (2002) e Hanushek e Raymond (2003) para descrição dos programas americanos. Para o caso brasileiro, ver Brooke (2006) e Andrade (2008).

de colocar o problema. A questão fundamental é saber se os alunos teriam um futuro mais ou menos promissor, caso as escolas priorizassem suas ações para proporcionar-lhes os conhecimentos e habilidades exigidas nos exames.

Existem hoje evidências robustas de que a escolaridade, medida como anos de estudo completos, é uma variável chave na determinação do progresso econômico de indivíduos e nações. Além dos aspectos econômicos, a escolaridade tem sido associada, também, a uma variedade de benefícios não-econômicos: melhor saúde, redução da criminalidade, menor incidência de gravidez na adolescência, maior coesão social etc.¹¹ Enquanto os benefícios da educação são bem conhecidos, a maneira pela qual a educação gera esses benefícios é ainda pouco compreendida. Nesse aspecto, a hipótese mais difundida é que, na escola, os indivíduos adquirem determinados conhecimentos e desenvolvem habilidades que possuem um alto valor social.

Determinar quais são e como medir tais conhecimentos e habilidades não é algo que possua uma resposta simples. Entretanto, já existe um acúmulo de evidências de que os resultados dos testes estão positivamente correlacionados com a renda individual futura dos indivíduos, a produtividade e o crescimento econômico.¹² Assim, se podemos entender essas correlações como causais, as escolas estariam contribuindo para o futuro de seus estudantes ao priorizar o ensino das habilidades e conhecimentos exigidos nos exames. Por outro lado, se anos de estudos completos importam, os resultados nos exames de leitura e matemática nos anos iniciais de escolarização são bons preditores da extensão da carreira estudantil dos indivíduos. Quanto melhor os estudantes vão nesses exames, mais anos de estudos eles tendem a acumular.¹³ Em suma, dado o conhecimento disponível, somos levados a considerar que escolas que, nos anos iniciais de escolarização, orientem seus esforços para o aprendizado de leitura e matemática estão, provavelmente, fazendo a coisa certa.

ii) Potencial dos programas

A idéia básica destes programas é que os responsáveis (professores, diretores e gestores) podem alterar suas condutas e, assim, proporcionar aos estudantes um melhor ensino. Melhorar a qualidade de ensino pode exigir alterações em práticas e hábitos enraizados no processo escolar e, portanto, de difícil mudança. Alguém que considere que os professores já fazem o máximo para proporcionar aos estudantes o melhor aprendizado, tenderia a considerar qualquer política de incentivos para eles, no mínimo, inócua. Mas, alguém que considere que certos procedimentos adotados pelos professores são inadequados e que eles têm capacidade para alterá-los, mas que, por algum motivo, não o fazem, pode ver grande potencialidade nas políticas de *accountability*, na medida em que elas alteram a estrutura de incentivos dos professores.

¹¹ A respeito dos benefícios econômicos da educação ver, entre outros, Ashenfelter e Rouse (1999) e Doppelhofer et alli (2000). Com relação aos benefícios não-econômicos, eles se encontram bem descritos em Hanushek (2002).

¹² Para referências desses estudos, ver Hanushek (2003)

¹³ Ver Currie e Thomas (1999).

No debate sobre *accountability* existe uma confusão entre responsabilidade (ter de prestar contas) e culpa (ato ou omissão repreensível). Muitos críticos vêem os programas de *accountability* como uma forma de culpar professores e diretores de escola pela baixa qualidade do ensino. Acreditar que os responsáveis pela educação podem alterar procedimentos e, assim, melhorar o ensino, não significa dizer que sejam culpados por seus alunos não estarem aprendendo mais.

Suponha que um professor (diretor) adote um conjunto de procedimentos A e que o melhor conjunto de procedimentos seja B. Admita também que B não esteja além da capacidade desse professor (diretor). Então, pode-se perguntar: por que ele não adota B? Há, pelo menos, duas respostas para essa pergunta.

A primeira é que ele não conhece B ou, se conhece, não sabe que B produz melhores resultados. Nesse caso, uma política possível é ensinar a esse professor (diretor) o conjunto de procedimentos B ou informá-lo que ele produz melhores resultados. É essa a idéia por detrás das políticas de capacitação e assistência técnica. Uma política de *accountability* pode, também, ser útil nessa situação. Ao divulgar amplamente os resultados das escolas, os programas podem “ensinar” aos interessados que escolas que adotam B possuem melhores resultados.

A segunda resposta é que o professor (diretor), apesar de conhecer B e saber que ele produz melhores resultados, pode considerá-lo muito custoso. Nesse caso, ele pode necessitar de um incentivo adicional para adotar B. Tome-se como exemplo uma escola cujo diagnóstico para seu baixo desempenho reside no fato de uma parcela de seus professores não ser habilitada para a função. Uma saída seria a troca desses professores. Entretanto, processos de demissão podem ser desgastantes para quem os conduz e é possível que o diretor prefira evitá-los. A presença de um programa de *accountability*, ao expor os resultados das escolas e ao atrelar prêmios e sanções a eles, pode gerar o incentivo necessário para que esse diretor efetive as demissões, inclusive com respaldo dos demais professores. Para que os programas de *accountability* possam contribuir para mudar práticas e hábitos, é necessário que os agentes submetidos a eles possuam autonomia para fazê-lo. Nesse exemplo, não adiantaria ter um sistema de prêmios e/ou punição, caso o diretor não tivesse autonomia para contratar e demitir professores. E mais, prêmios e punições devem ser atrelados aos responsáveis pelas mudanças. Por exemplo, diminuir os recursos de escolas ou redes de ensino com baixo desempenho, sem proporcionar os meios para qualquer mudança, pode, simplesmente, punir uma segunda vez os alunos já mais desfavorecidos.

Em suma, o potencial dos programas de *accountability* depende muito da capacidade deles em induzir mudanças nos procedimentos adotados que contribuam para melhorar o ensino. Aqueles que acreditam que a qualidade da educação pode melhorar, implicitamente, admitem que alguns dos responsáveis pela educação (professores, diretores, gestores de rede ou governantes) podem fazer algo diferente do que vêm fazendo. Nessa perspectiva, em algum nível (professores, escolas ou sistemas de ensino) a *accountability* pode ser necessária.

Hanushek e Raymond dividem os programas de *accountability* em dois tipos: i) aqueles que se limitam a divulgar os resultados dos estudantes por escolas e sistema educacional, que vamos chamar de “*accountability* fraca” e ii)

aqueles que, além disso, atrelam prêmios, sanções e assistência a tais resultados, que vamos chamar de “*accountability forte*”.¹⁴

A divulgação dos resultados dos estudantes fornece parâmetros ao público interno (no caso das escolas, professores e diretor) e pode desfazer certas percepções que a escola ou a rede de ensino tem de si mesma. Na falta de resultados comparáveis, pode-se acreditar estar fazendo um bom trabalho, quando de fato não está. A divulgação funciona também como um elemento de pressão por melhoria da qualidade. Os pais, por exemplo, podem cobrar providências para que ela melhore, caso os resultados não sejam bons. Eles podem, também, utilizar tais informações na escolha da escola do filho. Ainda que na escola pública a possibilidade de escolha seja limitada, existe algum espaço isso e, do ponto de vista do sistema como um todo, seria desejável que melhores escolas tivessem mais alunos.

O fato do desempenho dos estudantes não refletir apenas a qualidade da escola, mas também as condições sócio-econômicas dos estudantes, não é grave nesse caso. O público interessado pode “extrair o sinal de qualidade” de uma escola, por exemplo, por compará-la com escolas próximas e/ou que possuem público similar. Pode ser mais fácil para o público relacionado à escola (pais de alunos, professores e diretores) estabelecer o parâmetro correto de comparação do que para qualquer analista distante dessa realidade.

Esse não é o caso, no entanto, para os programas de “*accountability forte*”. Para esses, a questão da medida de desempenho é um elemento sensível, já que as premiações e/ou punições são automaticamente atreladas a ela. Premiar ou penalizar escolas com base no perfil dos alunos que elas recebem, além de ser injusto, traria conseqüências indesejáveis à estrutura de incentivos implícita nos programas. Por exemplo, escolas mal avaliadas por receberem alunos com piores condições sócio-econômicas poderiam ser desmotivadas a melhorar a qualidade de ensino, já que o *ranking* de escolas reflete pouco o esforço realizado.

Existe certo consenso na literatura que programas de “*accountability forte*” deveriam utilizar alguma medida de valor adicionado. Entretanto, não há consenso sobre a melhor forma de fazer isso. Obter um modo totalmente convincente de decompor o resultado dos alunos em “efeito escola” e “efeito aluno” pode ser impossível. Isso não é necessariamente um problema, no sentido de induzir as mudanças necessárias, e existem desenhos que são menos sensíveis a esses erros de medida.¹⁵ De qualquer forma, quando prêmios e sanções são atrelados aos resultados, o cuidado deve ser redobrado. Bons desenhos podem reforçar as mudanças no sentido correto, maus desenhos podem ser contraproducentes.

iii) Riscos

Dois potenciais riscos têm sido destacados na literatura: distorção de incentivos e *gaming*. No primeiro, o destaque tem sido dado para o “estreitamento”

¹⁴ Hanushek e Raymond (2004 e 2005)

¹⁵ É possível, por exemplo, separar as escolas em grupos (A, B e C), onde as escolas do grupo A teriam baixíssima probabilidade de agregar menos valor que uma escola média, enquanto as escolas do grupo C teriam baixíssima probabilidade de agregar mais valor que a média.

do currículo e para a exclusão de alunos de baixa proficiência. Se os objetivos dos programas são adequados, o estreitamento do currículo não vem a ser um problema. Por outro lado, pode, de fato, haver um incentivo para a exclusão de alunos com baixa proficiência. Por exemplo, programas cuja medida é a parcela de estudantes acima de determinado nível estimulam as escolas a se concentrarem nos estudantes cujo desempenho está próximo a esse nível de proficiência e a abandonar os alunos de pior desempenho, com baixa probabilidade de alcançar o nível exigido. Assim, os programas devem procurar incluir formas de penalizar a exclusão de alunos com baixa proficiência.

O segundo risco (*gaming*) decorre do fato que as escolas podem adotar estratégias para alterar os resultados, mas que não mudam a qualidade do ensino ministrado como, por exemplo, treinar e motivar os estudantes para os testes ou excluir dos exames alunos de baixa proficiência. Parte desses problemas não se pode evitar como, por exemplo, o treinamento e motivação dos estudantes para os exames, mas não tendem a ser grave. O problema maior é quando o *gaming* se confunde com fraude. Nesses casos, só a fiscalização pode resolver.

iv) Os resultados

Estudos de casos para programas pioneiros mostram que, em alguns deles, o crescimento de desempenho dos estudantes foi excepcional. Esse é o caso, por exemplo, da Flórida e de Chicago¹⁶ Esses resultados ajudam a explicar a proliferação recente desses programas. No entanto, alguns autores têm colocado tais resultados sob suspeita ao argumentar que muito desse crescimento é devido ao *gaming*¹⁷. Há, por exemplo, evidências de seleção de alunos para os exames, de treinamento específico para os testes e de fraude de professores, que respondem questões para os alunos.¹⁸

Hanushek e Raymond providenciam importantes estudos para avaliar os impactos dos programas de *accountability* dos estados americanos. Por considerar um indicador não utilizado para *accountability*, seus resultados estão menos sujeitos ao *gaming*. Eles observam que estados que introduziram sistemas de *accountability* tenderam a mostrar um aumento mais rápido na proficiência durante os anos 90, quando comparados aos estados que não introduziram tais medidas. O impacto de um sistema de “*accountability* fraca” se mostrou menor, indicando que pode ser importante atrelar prêmios e sanções aos resultados dos exames. Eles não encontraram evidências de exclusão de alunos de baixa proficiência.¹⁹

Em suma, tanto a argumentação teórica como as evidências indicam que os programas de *accountability* possuem potencial para elevar o desempenho dos alunos nos aspectos medidos pelos exames. Vimos, também, que priorizar esses exames pode ser um procedimento bastante razoável. Para as séries iniciais, ter

¹⁶ Ver para Florida, Cullen e Reback (2006) e para Chicago, Jacob (2005).

¹⁷ Por exemplo, Jacob (2007) observa que o crescimento no desempenho dos estudantes tende ser maior nos testes usados pelos programas de *accountability* dos estados do que no *National Assessment of Educational Progress* (NAEP), que não é utilizado para efeitos de *accountability*.

¹⁸ Ver Cullen e Reback (2006), Jacob (2007) e Jacob e Levitt (2003).

¹⁹ Hanushek e Raymond (2004 e 2005).

como foco leitura e matemática parece adequado. Entretanto, para as séries finais do ensino básico isso pode ser insuficiente. Programas de *accountability* não estão isentos de riscos, sendo a exclusão de alunos de baixa proficiência o mais preocupante. Por fim, há indícios de que esses programas, de fato, incentivam a prática de *gaming*. Dessa forma, podemos esperar certa “inflação” na pontuação dos exames após a implementação dos programas. Por fim, os resultados podem variar, a depender do desenho, da gestão, do ambiente etc. Assim, os resultados apresentados devem ser vistos apenas como indicativos.

Accountability no Brasil: a Prova Brasil e a divulgação do ENEM por escola

No Brasil o principal marco da avaliação educacional é o SAEB (Sistema Nacional de Avaliação da Educação Básica). Este sistema teve seu início no final da década de 80 e se constituiu hoje no principal sistema de avaliação diagnóstica da educação básica brasileira. O SAEB é a conjugação de testes de desempenho aplicados aos estudantes com questionários, sobre fatores associados a tais resultados, endereçados a diferentes atores que compõem a escola.

O SAEB foi idealizado para conter amostras de escolas e de turmas representativas do país, das regiões e dos estados, para as escolas públicas e privadas brasileiras. Foram realizados, até 2008, nove ciclos de avaliação e incorporados avanços metodológicos importantes, como a teoria de resposta ao item (TRI) que permitiu, desde 1995, a comparabilidade entre anos e séries. As avaliações se concentraram em português (leitura) e matemática (resolução de problemas) no final dos ciclos²⁰

O SAEB garantiu qualidade e confiabilidade às avaliações realizadas no Brasil e por meio dele introduziu-se e consolidou-se uma cultura de avaliação no país. Com o SAEB detectaram-se as dificuldades com a qualidade da educação brasileira.²¹ Porém, ele não respondia a todas as necessidades de informação, pois, dado o seu caráter amostral, não apreendia toda a diversidade educacional do país, especialmente frente à ampla descentralização e municipalização que caracterizou a educação brasileira pós-redemocratização.

Esta é uma das explicações da entrada em cena, em 2005, da Prova Brasil: aumentar conteúdo informacional da avaliação diagnóstica e permitir que todos os sistemas educacionais públicos brasileiros se “enxergassem” na avaliação. Além disto, um novo passo tornou-se possível: agregar à perspectiva de avaliação para o diagnóstico a noção de *accountability*.

A Prova Brasil foi aplicada pela primeira vez em 2005 e sua principal diferença em relação ao SAEB é que, dentro de seu universo de referência, ela é censitária. O universo da Prova Brasil são as escolas públicas e urbanas nas 4ª e 8ª série (5º e 9º ano) do ensino fundamental.²² Nesse ano, as duas avaliações (SAEB e Prova Brasil) foram realizadas separadamente, com instrumentos

²⁰ Em alguns anos também foram aplicadas avaliações de ciências, história e geografia.

²¹ A participação do Brasil em avaliações internacionais como o PISA (OCDE) e o PERCE (Unesco) também conferiram importância aos problemas da qualidade da educação brasileira. O Brasil foi o único país não membro da OCDE a participar de todas as avaliações do PISA.

²² Em 2005, apenas as escolas com mais de 30 alunos na série/ano faziam parte da Prova Brasil. Este número foi reduzido para 20 em 2007.

diferentes (testes e questionários). Já em 2007 estas duas avaliações se fundiram e voltamos a ter uma única avaliação externa de caráter federal, em que os segmentos que compõem o universo da Prova Brasil foram avaliados censitariamente, enquanto que os outros segmentos que compõem o SAEB, mas não a Prova Brasil, foram avaliados amostralmente. Os instrumentos utilizados (testes e questionários) foram os mesmos.²³ Assim, voltamos a ter uma única avaliação (o SAEB) e o que entendemos por Prova Brasil constitui-se, na verdade, em estratos censitários do SAEB.

Um aspecto importante é que tanto a Prova Brasil quanto o SAEB são avaliações com adesão voluntária por parte das Secretarias de Educação. Nas duas edições realizadas a adesão das secretarias estaduais foi completa e também houve uma ampla adesão das redes municipais, atingindo mais de 99% das que faziam parte do universo.²⁴ Os resultados da Prova Brasil passaram a ser amplamente divulgados. Desde 2005, portanto, o Brasil passa a contar como um sistema de avaliação para diagnóstico e também com um programa de “*accountability* fraca”, por escolas e redes de ensino.

Isso, no entanto, se restringe ao ensino fundamental. No caso do ensino médio, como a Prova Brasil não atinge censitariamente as escolas públicas, não há resultados por escola. O sistema de *accountability* para ensino médio com base na Prova Brasil/SAEB só alcança as redes estaduais de ensino. Para o ensino médio, no entanto, o Brasil dispõe de outra avaliação: o ENEM (Exame Nacional de Ensino Médio).

Quando criado, em 1998, o ENEM tinha o objetivo de fornecer uma auto-avaliação aos estudantes que terminavam o ensino médio. O exame, de caráter voluntário, permitia apenas a comparação dos resultados obtidos por um participante com a média dos demais participantes na edição daquele ano. Contudo, várias instituições de ensino superior passaram a se valer dos resultados do ENEM como critério, parcial ou exclusivo, de seleção para ingresso nos seus cursos. Em 2005, o governo federal passou a utilizar o ENEM como critério para a concessão de bolsas no âmbito do Programa Universidade Para Todos (ProUni).²⁵ Tais elementos, junto com a gratuidade do exame para aqueles oriundos de escolas públicas, fizeram com que o ENEM, apesar de continuar sendo voluntário, tenha atingido 4.018.070 inscritos em 2008 (conforme figura abaixo) e passou a ter uma boa cobertura dos concluintes do ensino médio.²⁶

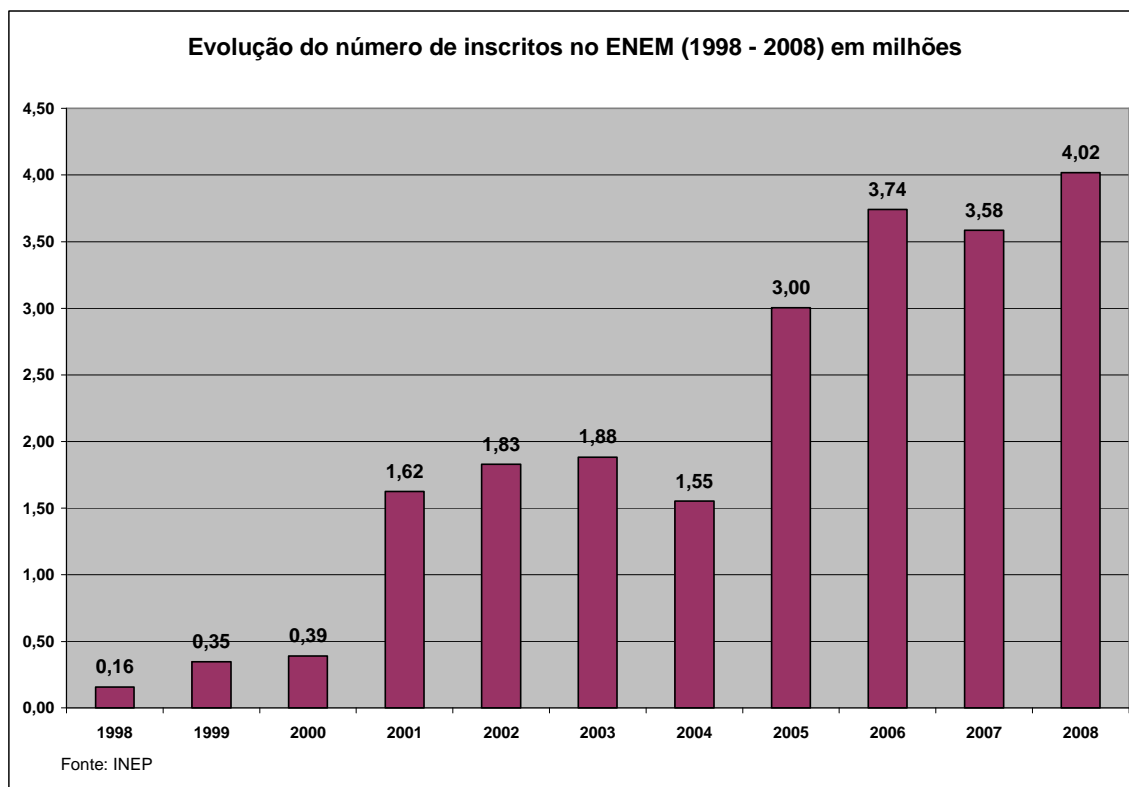
Assim, além de auto-avaliação e de credenciamento, o ENEM passa a ser utilizado, também, como instrumento de diagnóstico e de *accountability*. Deste modo, a partir de 2005, os resultados do ENEM passaram a ser divulgados por estado, sistema de ensino e por escola, tomando-se apenas os resultados dos concluintes do ensino médio que realizam cada edição do ENEM.

²³ No SAEB um aluno respondia ou ao teste de Português ou ao de Matemática. Na Prova Brasil o mesmo aluno respondia aos testes das duas disciplinas. Em 2007 este último procedimento foi estendido ao SAEB.

²⁴ Em 2005, nem todos os alunos da rede estadual de Estado de São Paulo participaram, ainda que todas as escolas tenham participado. Assim, não foi possível ter resultados por escolas.

²⁵ Atualmente, 2/3 dos candidatos declaram fazer o ENEM com vistas ao ingresso na universidade.

²⁶ Das mais de 24.250 escolas de ensino médio que constam do Censo Escolar, cerca de 23.000 tinham alunos inscritos no ENEM em 2007.



Um cuidado teve que ser tomado para que esta divulgação fosse feita. O fato do exame ser voluntário poderia enviesar os resultados por escola. Para minimizar este possível viés, foram estabelecidos critérios mínimos de participação para a divulgação por escola e criado um corretor de participação. Mesmo para as escolas com alta participação, a amostra dos alunos de cada instituição pode não representar o desempenho médio da escola que seria obtido caso todos os alunos participassem. O objetivo do corretor foi representar a nota média da escola, caso todos os alunos matriculados no final do ensino médio tivessem realizado o exame.²⁷

Assim, desde 2005 o ENEM passou a contribuir para o estabelecimento de um diagnóstico do sistema educacional brasileiro e a fazer parte de um sistema de “*accountability* fraca”, ao lado da Prova Brasil/SAEB. O ENEM não permite que a comparação seja feita entre duas edições do exame ou em relação aos resultados da Prova Brasil/SAEB. Por outro lado, o ENEM envolve também as escolas particulares e não enfrenta problemas de falta de motivação, por parte dos participantes, para se obter um bom resultado no exame. Portanto, valendo-se do ENEM e da Prova Brasil /SAEB, o Brasil passa a ter, a partir de 2005, um sistema nacional de *accountability* da educação básica.

Deve-se ressaltar que existem experiências de sistemas de *accountability* efetuados por iniciativa de Secretarias Estaduais de Educação. Esse foi o caso,

²⁷ Ver INEP (2008)

por exemplo, de experiências, já abandonadas, nos estados do Rio de Janeiro (projeto Nova Escola) e Paraná (Boletim da Escola). Há divulgação de resultados por escolas em Minas Gerais e, mais recentemente, no Ceará, São Paulo e Rio Grande do Sul.²⁸

Avaliação externa e fluxo escolar: o IDEB

Como já ressaltado, os programas de *accountability* possuem potencial para elevar o desempenho dos estudantes, mas apresentam riscos. No Brasil, um dos principais riscos de um programa de *accountability*, baseado apenas em exames padronizados, é o de contribuir para agravar o já dramático quadro de repetência e evasão escolar.

A questão do acesso à escola mereceu, nos anos 80 e 90, um enorme esforço, sendo que o país conseguiu colocar a quase a totalidade das crianças na escola. Porém, em relação aos indicadores de fluxo escolar, o Brasil ainda permanece com fortes dificuldades: altas taxas de repetência e elevada proporção de adolescentes que abandonam a escola sem concluir a educação básica²⁹. Isso, aliado à baixa proficiência obtida em exames padronizados. Assim, a idéia foi criar um indicador da educação básica que, tanto para efeitos de monitoramento quanto para *accountability*, combinasse informações de desempenho em exames padronizados com informações sobre fluxo escolar.

Medidas de proficiência em exames padronizados e os indicadores de fluxo escolar não são independentes. Por exemplo, a adoção de uma medida que melhore o aprendizado, fixado o padrão de aprovação, tende tanto a elevar a proficiência nos testes padronizados, como a reduzir a taxa de repetência. Se a relação entre indicadores de proficiência em exames padronizados e indicadores de fluxo fosse sempre desse tipo, não haveria grandes problemas em escolher entre eles, uma vez que a melhora (piora) de um indicador implica melhora (piora) do outro. A dificuldade de escolha é maior ante a possibilidade de eles terem sentidos opostos. Suponhamos que se adote uma redução no padrão de aprovação. Nesse caso, as taxas de repetência seriam reduzidas, mas, muito provavelmente, teríamos uma redução na proficiência média dos estudantes.

Não se pode descartar, assim, a possibilidade das escolas e/ou redes de ensino adotarem medidas que melhorem o fluxo e piorem o desempenho nos exames padronizados e vice-versa. Nesse caso, se a cobrança for restringida aos indicadores de fluxo, ela pode incentivar os professores, diretores e gestores a adotarem medidas que impliquem redução no desempenho médio dos estudantes nos testes padronizados, como, por exemplo, reduzir o padrão de aprovação. Por outro lado, se a cobrança for apenas sobre o resultado dos alunos nos exames, o incentivo é o oposto, como, por exemplo, elevar o padrão de aprovação. Um sistema educacional que reprova sistematicamente seus estudantes, fazendo com que grande parte deles abandone a escola antes de completar a educação básica, não é desejável, mesmo que aqueles que concluem essa etapa atinjam

²⁸ Sobre as experiências estaduais ver Andrade (2008) e Brooke (2006).

²⁹ Os motivos para as elevadas taxas de retenção escolar têm sido bastante discutidos na literatura. Ver, por exemplo, Ribeiro (1991), Paro (2000b) e Schiefelbein e Wolff (1993).

elevadas pontuações nos exames padronizados. Mas um sistema em que os alunos concluem o ensino médio no período correto não é de interesse caso eles aprendam muito pouco.

Uma visão mais integrada de desenvolvimento educacional, que considera tanto o desempenho dos estudantes em testes como informações de fluxo, é a de que as avaliações deveriam ser aplicadas aos alunos de determinada geração, em vez aos de determinada série. Nesse caso, o desempenho dos alunos atrasados, que se pressupõe menor que o dos que estão na série correta, estaria contemplado. Porém as avaliações aplicadas no Brasil (SAEB, Prova Brasil e ENEM) têm como base a série, sendo o PISA (Programme for International Student Assessment), que é aplicado aos alunos de 15 anos de idade, a única exceção. Nesse caso, no entanto, a avaliação é amostral, com resultados disponíveis só para estados e para o país.

Assim, o IDEB surge para contornar os riscos de agravar os problemas de fluxo que um sistema de *accountability* tradicional poderia gerar em um país como o Brasil. O IDEB é o resultado da combinação de dois outros indicadores: a) a pontuação média dos estudantes em exames padronizados ao final de determinada etapa da educação básica e b) a taxa média de aprovação dos estudantes da correspondente a etapa de ensino. Ele é de fácil compreensão, simples de calcular e explícito em relação à “taxa de troca” entre probabilidade de aprovação e proficiência dos estudantes. Ou seja, o indicador torna claro o quanto se está disposto a perder na pontuação média dos estudantes nos exames para se obter determinado aumento na taxa média de aprovação.

A dificuldade em se fixar a “taxa de troca” decorre da falta de clareza de como a relação entre repetência e desempenho é determinada nas escolas e, em que medida um aumento da proficiência esperada é melhor para o futuro dos estudantes quando esse aumento vem acompanhado de um aumento na probabilidade de repetência. As elevadas taxas de repetência do país apontam para a existência de algo errado com nosso sistema educacional. Entretanto, isso não significa que a reprovação de alunos em si não possua alguma funcionalidade e que o ideal seria ter um sistema com taxa de reprovação igual a zero.

Três justificativas usuais para se promover a retenção de alunos são: i) permitir aos que não obtiveram desempenho adequado a oportunidade de se recuperarem e, assim, prosseguirem no sistema de forma mais apropriada; ii) incentivar os alunos a aumentar o empenho na obtenção de habilidades e conhecimentos; e iii) impedir que alunos sem os conhecimentos mínimos avancem no sistema, o que viria a atrapalhar o desenvolvimento dos que estão aptos para a série seguinte. A validade de cada uma é, evidentemente, questionável, mas não podem ser descartadas *a priori*. Essas justificativas têm em comum o fato de que, enquanto a reprovação depende do desempenho obtido pelos alunos, também ela afeta esse desempenho. Elas pressupõem que reprovações são produtivas, no sentido que contribuem para o aprendizado. Ainda que reprovações não fossem produtivas, as escolas poderiam desejar promovê-las para impedir que alunos de baixa proficiência cheguem ao final do ciclo e realizem os exames externos.

Fernandes mostra que uma taxa de troca igual a “um” (um aumento na taxa de reprovação de 10% pode ser compensado pelo aumento de 10% nas notas dos estudantes) seria, provavelmente, suficiente para eliminar esse tipo de

reprovação, caso as escolas tivessem como objetivo obter o maior IDEB possível e o ganho de aprendizado do aluno reprovado não seja muito grande.³⁰ Entretanto, se as reprovações contribuíssem para elevar o desempenho dos alunos pelas justificativas (ii) e (iii), a taxa ótima de reprovação seria diferente de zero, mas, provavelmente, muito baixas. Isso se confirmada a crença, de grande parte dos pesquisadores em educação, de que reprovações são pouco produtivas. Com base nesses argumentos, o IDEB foi fixado para ter uma taxa de troca igual a “um”. Isso é obtido pela simples multiplicação da proficiência média dos alunos da escola (N) pela taxa média de aprovação (P), como pode ser visto na seguinte formulação:

$$\text{IDEB} = N \cdot P$$

A média de proficiência dos estudantes combina as notas de matemática e português e a proficiência média é padronizada para estar entre zero e dez, enquanto que a taxa de aprovação se situa entre zero e um, de modo que o IDEB também se situa entre zero e 10.³¹

O IDEB é, assim, um indicador que sintetiza as informações de desempenho em exames padronizados com as informações sobre rendimento escolar (taxa média de aprovação dos estudantes na etapa de ensino). Este último (rendimento escolar) pode ser também visto como o inverso do tempo médio de conclusão de uma série. Assim, o IDEB também pode ser interpretado da seguinte maneira: para uma escola A cuja média padronizada da Prova Brasil, 4ª série, é 5,0 e o tempo médio de conclusão de cada série é de dois anos, a rede/escola terá o IDEB igual a 5,0 multiplicados por 0,5, ou seja, $\text{IDEB} = 2,5$. Já uma escola B com média padronizada da Prova Brasil, 4ª série, igual a 5,0 e tempo médio para conclusão igual a um ano, terá $\text{IDEB} = 5,0$.

Vários aprimoramentos são, evidentemente, possíveis, mas o índice proposto é uma inovação em relação aos indicadores até então utilizados para monitorar o sistema de ensino e passou a ser o principal indicador de *accountability* do país. O IDEB é calculado e disponibilizado para a sociedade para todos os sistemas de ensino que fazem parte do SAEB e todas as redes e escolas da Prova Brasil.

Um sistema de metas para o avanço da qualidade da educação

Com a implementação do IDEB, o Ministério da Educação sinaliza às redes e escolas públicas um objetivo a ser perseguido: obter o maior valor possível no indicador. Pode-se, no entanto, questionar a capacidade da simples divulgação do IDEB fornecer o incentivo necessário para promover mudanças que resultem em aumentos significativos no indicador. É importante ressaltar que as escolas públicas de educação básica estão, quase em sua totalidade, nas mãos de estados e municípios e que esses possuem total autonomia para geri-las. Isso impõe restrições às ações que o governo federal pode adotar. Para reforçar o objetivo de melhorar o IDEB, o MEC cria o sistema de metas educacionais.

³⁰ Fernandes (2007).

³¹ Sobre a padronização ver INEP (2007a).

A idéia central do sistema de metas foi obter um maior comprometimento das redes e escolas com o objetivo da melhoria do IDEB. Um sistema de metas – pactuadas entre o Ministério da Educação e Secretarias de Educação de estados e municípios – serviria para aumentar a mobilização da sociedade em favor da qualidade da educação. Para isso, os desafios eram: i) estabelecer metas de aumento do IDEB – para o país, redes de ensino e escolas públicas - que fossem desafiadoras e factíveis; ii) comprometimento de estados e municípios com essas metas e iii) estabelecimento de metas intermediárias para monitoramento.

O sistema de metas foi introduzido pelo PDE, no denominado “Compromisso Todos pela Educação”. As metas do IDEB foram estipuladas para 2021 (divulgação em 2022), com metas intermediárias para cada dois anos, a partir de 2007. A meta para a média do Brasil foi estabelecida a partir de um indicador externo. Ou seja, procurou-se simular o cálculo do IDEB para os países da OCDE e verificar qual seria o desempenho médio desses no indicador. Para isso foi utilizada a ordenação dos alunos brasileiros no PISA e no SAEB e a ordenação dos alunos da OCDE no PISA. Admitiu-se que os exames ordenam os alunos de forma similar e, assim, fez-se a conversão das escalas.³² Para o cálculo do IDEB dos países da OCDE, supôs-se uma taxa de aprovação de 96%. O IDEB nacional em 2005 foi, para a 4ª série, de 3,5 e a meta estipulada foi de 6,0.

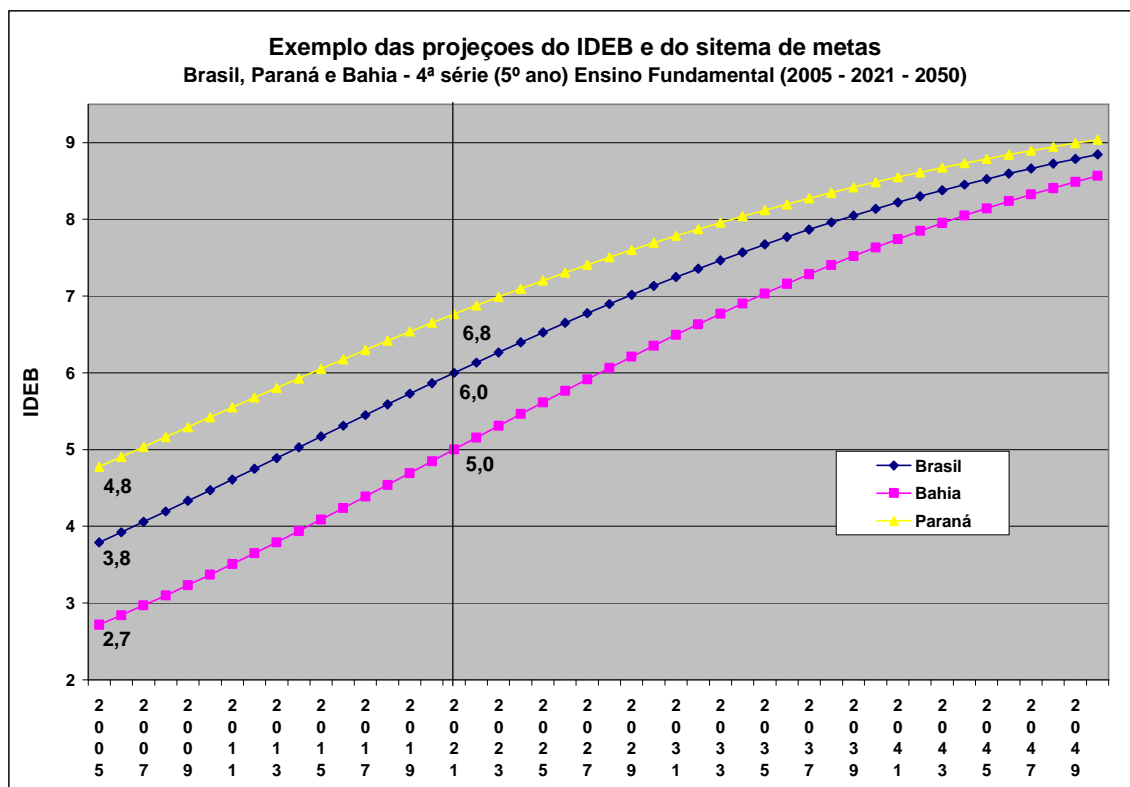
Para projetar as trajetórias esperadas para o IDEB, tanto para o Brasil como para redes e escolas, foram feitas algumas hipóteses

- I. As trajetórias do IDEB - para o Brasil, redes e escolas - têm o comportamento de uma função logística;
- II. As trajetórias do IDEB por rede de ensino devem contribuir para a redução das desigualdades regionais de qualidade educacional. Isso significa a adoção de um processo de convergência dos IDEBs no longo prazo
- III. As metas de cada rede e escola devem ser baseadas no “esforço” que o Brasil necessita fazer para atingir suas metas. Assim, cada rede de ensino e escola possui “esforço” similar, mas metas diferenciadas, a depender do ponto que elas se encontravam em 2005.³³ (ver trajetórias no gráfico):

Com estes pressupostos, desenhou-se e divulgou-se para cada escola e/ou sistema de ensino participante da Prova Brasil, índices progressivos para o IDEB a serem alcançados de dois em dois anos, de modo que a média brasileira em 2021 fosse seis (para a primeira fase do ensino fundamental), cada rede e cada escola tem suas metas próprias (no gráfico visualizam-se as metas do Paraná e da Bahia). Os primeiros dados do IDEB foram divulgados em 2007, referentes a 2005; em 2008 foram divulgados os dados de 2007 e, de uma forma geral, atingiram-se as metas estabelecidas, mas o percurso até 2021 é longo e as metas não são simples de serem alcançadas.

³² Ver INEP (2007b).

³³ O esforço é medido pelo parâmetro que dá a curvatura da função logística. Assim, calcula-se o valor do parâmetro que garante o comprimento da meta do país. Dado esse parâmetro, pode-se calcular a data em que o IDEB do país atingirá o valor 9,9. Então, forçou-se que todas as redes e escolas tenham IDEB 9,9 nessa data. Ver INEP (2007c).



O outro lado do Compromisso Todos pela Educação é a possibilidade de intervenção dos governos com base nos indicadores. O MEC estabeleceu acordos com as redes de ensino para que elas acolhessem as metas e se comprometessem com elas. As escolas que atingirem as metas são beneficiadas com aumento de seus recursos no programa de dinheiro direto da escola (PDDE) do MEC, mas a principal atenção do governo foi às redes com piores IDEBs. O MEC estabelece convênios com estados e municípios, por meio da elaboração local de um Plano de Ações Articuladas (PAR). Pelo PAR, os gestores municipais e estaduais se comprometem a promover um conjunto de ações, responsabilizando-se pelo alcance das metas estabelecidas. Em contrapartida, passam a contar com transferências voluntárias e assessoria técnica da União.

Conclusão

O presente capítulo procurou analisar o papel que as avaliações externas ocupam hoje na política educacional brasileira. A ênfase se deu nas mudanças recentes, em especial as introduzidas no contexto do Plano de Desenvolvimento da Educação (PDE). Procuramos apresentar os motivos de se incorporar os objetivos de *accountability* e clarificar as especificidades do modelo adotado.

As avaliações constituem-se em um dos pilares de toda política educacional do Ministério da Educação, por meio do plano de metas do Compromisso Todos pela Educação, cujo objetivo é atingir, até 2021, o padrão de qualidade dos países desenvolvidos. Por ele, todas as esferas de governo (federal, estadual e municipal) se comprometem com metas de melhoria do IDEB.

Ao analisarmos o potencial e os riscos de um programa de *accountability* educacional, concluímos que tais programas podem contribuir para elevar o desempenho dos alunos, mas não estão isentos de riscos. No Brasil, a principal preocupação com um sistema de *accountability* focado apenas no desempenho dos estudantes em exames externos é o de agravar os problemas de repetência e evasão. Essa foi a principal justificativa para a criação do IDEB. Ao combinar as notas da Prova Brasil/Saeb com as taxas de aprovação, o IDEB visa coibir tanto a reprovação indiscriminada como a prática de aprovar alunos que nada aprenderam. O IDEB foi, também, utilizado para estabelecer metas para redes e escolas e, assim, obter um maior comprometimento das secretarias de educação de estados e municípios para que, até 2021, o Brasil atinja o estágio educacional atual dos países desenvolvidos.

Enquanto há divulgação de resultados e fixação de metas por escolas, a ênfase do sistema de *accountability* é dada para estados e municípios. O Brasil possui um sistema educacional descentralizado, com mais de cinco mil redes de ensino com autonomia para gerir suas escolas. A vantagem da descentralização é que ela tem maior potencial para gerar inovações. Mas para que todos se beneficiem é preciso que os casos de sucesso sejam identificados e difundidos. Por outro lado, a falta de resultados comparáveis entre redes pode levá-las ao isolamento. Um sistema de nacional de avaliação possibilita que: i) inovações de sucesso sejam compartilhadas por todos; ii) o MEC identifique as redes em pior situação, orientando suas ações de assistência técnica e financeira; e iii) a opinião pública local tenha elementos para cobrar seus representantes por melhorias na educação. Em fim, o plano de metas Compromisso Todos pela Educação é um sistema de *accountability* compatível com o federalismo existente no país.

Bibliografia:

- Andrade, Eduardo (2008). “*School accountability’ no Brasil: experiências e dificuldades*” **Revista de Economia Política**. v. 28, n. 3: 443-453.
- Ashenfelter, Orley e Rouse, Cecilia (1999) “*Schooling, Intelligence, and Income in América: Cracks in the Bell Curve*”. **NBER Working Paper** 6902. Cambridge: NBER.
- Brooke, Nigel (2006) “*O futuro das políticas de responsabilização educacional no Brasil*.” **Cadernos de Pesquisa**, v. 36, n. 128: 377-401.
- Carnoy, Martin e Loeb, Susana (2002) “*Does external accountability affect student outcomes? A cross-state analysis*.” **Educational Evaluation and Policy Analysis**, v.24, n.4: 305-331.
- Chakrabarti, Rajashri (2008) “*Impact of voucher design on public school performance: Evidence from Florida and Milwaukee voucher programs*.” **Federal Reserve Bank of New York Staff Reports**, n. 315.
- . (2006) “*Vouchers, public school response and the role of incentives: Evidence from Florida*.” **Working Paper - Federal Reserve Bank of New York**.
- Coleman, James; Campbell, Ernest; Hobson, Carol; McPartland, James; Mood, Alexander; Weinfeld, Frederic; e York, Robert. (1966) **Equality of Educational Opportunity**. Washington DC.

- Cullen, Julie Berry e Reback, Randall (2006). "*Tinkering toward accolades: School gaming under a performance accountability system.*" In: Gronberg, T. J. e Jansen, D. W. (Eds), **Advances in Applied Microeconomics**, 14. Elsevier.
- Currie, Janet e Thomas, Duncam (1999) "*Early test Scores, Socioeconomic Status and Future Outcomes.*" **NBER Working Paper** 6943. Cambridge : NBER.
- Doppelhofer, Geot; Miller, Ronald e Sala-I-Martin, Xavier. (2000) "*Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach.*" **NBER Working Paper** 7750. Cambridge: NBER.
- Elmore, Richard, Abelman, Charles. e Fuhrman. Susan (1996). "*The New Accountability in State Education Reform: From Process to Performance.*" In Ladd, H. F. (Ed) **Holding Schools Accountable: Performance-Based Reform in Education**, Washington, DC: Brookings Institution: 65-98.
- Fernandes, Reynaldo (2007) "*Índice de desenvolvimento da educação básica (IDEB)*" **Texto para discussão INEP**, n.26. Inep.
- Figlio, David e Getzler, Lawrence (2006) "*Accountability, ability, and disability: Gaming the system?*" In: Gronberg, T. J. e Jansen, D. W. (Eds), **Advances in Applied Microeconomics**, 14. Elsevier.
- Figlio, David e Rouse, Cecilia (2006) "*Do accountability and voucher threats improve low -performing schools?*" **Journal of Public Economics**, v.90, n.1-2: 239-255.
- Figlio, David e Winicki, Joshua (2005) "*Food for thought? The effects of school accountability plans on school nutrition.*" **Journal of Public Economics**, v.89, n. 2-3: 381-394.
- Haddad, Fernando (2008). "*O Plano de Desenvolvimento da Educação: razões, princípios e programas.*" **Texto para discussão INEP** n.30. Inep.
- Hanushek, Eric (2003). "*The failure of input-based schooling policies.*" **Economic Journal**, v. 113: F64-F98.
- . (2002) "*Publicly Provided Education.*" **NBER Working Paper** 8799. Cambridge: NBER.
- Hanushek, Eric. e Raymond, Margareth (2005) "*Does school accountability lead to improved student performance?*" **Journal of Policy Analysis & Management**, v. 24, n, 2: 297-327
- . (2004). "*The effect of school accountability systems on the level and distribution of student achievement.*" **Journal of the European Economic Association**, v.2, n.2-3: 406-415.
- . (2003). "*Improving educational quality: How best to evaluate our schools?*" In Kodrzycki, Y. (Ed) **Education in the 21st Century: Meeting the Challenges of a Changing World**. Boston, MA: Federal Reserve Bank of Boston: 193-224.
- INEP (2007a). **Nota técnica: Índice de Desenvolvimento da Educação Básica.**
- . (2007b). **Nota metodológica sobre a compatibilização de desempenhos do PISA com a escala do SAEB.**
- . (2007c). **Nota técnica: Metodologia utilizada para o estabelecimento das metas intermediárias para a trajetória do Ideb no Brasil, Estados, Municípios e Escolas**
- . (2008). **Nota Técnica: Notas Médias do Enem 2007 por Município e por Escola dos Alunos Concluintes do Ensino Médio em 2007.**

- Jacob, Brian (2007). "Test-Based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments." **NBER Working Paper** 12817. Cambridge: NBER.
- . (2005). "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools." **Journal of Public Economics**, v.89, n. 5-6: 761-796.
- Jacob, Brian e Levitt, Steven (2003). "Rotten apples: An investigation of the prevalence and predictors of teacher cheating." **Quarterly Journal of Economics**, v.118, n. 3: 843-877.
- JEBS. (2004). "Special Issue of the Journal of Educational and Behavioral Statistics on Value-Added Assessment." **Journal of Educational and Behavioral Statistics**. v. 29, n. 3.
- Kane, Thomas. e Staiger, Douglas (2002). "The promise and pitfalls of using imprecise school accountability measures." **Journal of Economic Perspectives**. v.16, n.4: 91–114.
- . (2001). "Improving school accountability measures." **NBER Working Paper** 8156. Cambridge: NBER
- Krieg, John (2008) "Are students left behind? The distributional effects of No Child Left Behind." **Education Finance and Policy**, v.3, n. 2: 250-281.
- Ladd, Hellen e Zelli, Arnaldo (2002). "School-based accountability in North Carolina: The responses of school principals," **Educational Administration Quarterly**, v.38, n.4: 494-529.
- McCaffrey, Daniel; Lockwood, J.R; Koretz, Daniel e Hamilton, Laura. (2003). **Evaluating Value-Added Models for Teacher Accountability**. Santa Monica, CA: RAND.
- Mizala, Alejandra; Romaguera, Pilar e Urquiola, Miguel (2007). "Socioeconomic Status or Noise? Tradeoffs in the Generation of School Quality Information." **Journal of Development Economics**. v. 84, n. 1: 61-75.
- Neal, Derek e Schanzenbach, Deane Whitmore (2007) "Left behind by design: Proficiency counts and test-based accountability." **NBER Working Paper** 13293. Cambridge: NBER.
- Paro, Vitor Henrique (2000a) "Educação para a democracia: o elemento que falta na discussão da qualidade do ensino." **Revista Portuguesa de Educação**, Braga, Portugal., v. 13, n. 1: 23-38,
- . (2000b) "Por que os professores reprovam: resultados preliminares de uma pesquisa" **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v.8, n.28: 273-282.
- Ribeiro, Sergio da Costa. (1991) "A pedagogia da repetência." **Estudos Avançados**, São Paulo, v.5, n.12: 7-21.
- Rouse, Cecilia; Hannaway, Jane; Goldhaber, Dan e Figlio, David (2007) "Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure." **NBER Working Paper** 13681. Cambridge: NBER.
- Schiefelbein, Ernesto e Wolff, Laurence (1993) "Repetition and inadequate achievement in Latin America's primary schools: a review of magnitudes, causes, relationships and strategies". **Estudos em Avaliação Educacional**, São Paulo, n.7: 45-87.