

Alternativa de política educacional para o Brasil: *School Accountability*

EDUARDO DE CARVALHO ANDRADE*

An alternative to the education policy in Brasil: School accountability. This paper examines the school accountability (SA) policies adopted in the US. Significant impacts on the quality of education occur when the SA incorporates a set of sanctions and rewards to schools based on their students' performance. In comparison with other policies, it is also more efficient. Potential problems of adopting the SA (bias toward cognitive ability, gaming and difficulty in measuring the school contribution) can be overcome. The analysis suggests that the SA should be considered as an alternative to improve the quality of education in Brazil.

Keywords: school accountability; education; education policy; incentive schemes; student performance.

JEL Classification: I21; I28.

INTRODUÇÃO

Desde o início dos anos 1990 no Brasil, as políticas educacionais obtiveram um êxito importante em elevar o nível de escolaridade da população. O Fundef (Fundo de Desenvolvimento do Ensino Fundamental e Valorização do Magistério), aprovado em 1996, contribuiu para esse processo. Dentre outros pontos, esse programa passou a distribuir os recursos para os estabelecimentos escolares no ensino fundamental em função do número de estudantes matriculados. Com isso, foi criado um forte estímulo para os gestores da educação, principalmente em nível municipal, atraírem as crianças para as escolas e evitarem a evasão escolar. O resultado foi significativo, levando à quase universalização do ensino fundamental. A parcela dos alunos entre 7 e 14 anos frequentando a escola passou de 88% em 1993 para 97%

* Insper Instituto de Ensino e Pesquisa, São Paulo. E-mail: Eduardo.Andrade@insper.org.br. Submetido: Agosto 2006; Aprovado: Março 2007.

em 2003. A taxa de escolarização líquida saltou de 20,8 em 1994 para 33,3 em 2000.¹ A taxa de analfabetismo caiu de 20% em 1991 para 12% em 2002.²

Se houve avanços importantes no quesito quantidade da educação, o mesmo não se verificou em relação à qualidade da educação. Em 1995, o governo federal instituiu o Saeb (Sistema Nacional de Avaliação da Educação Básica), que é uma prova de proficiência de matemática e língua portuguesa aplicada de dois em dois anos numa amostra de alunos da 4.º e 8.º séries do ensino fundamental e na 3.º série do ensino médio. Os resultados têm sido bastante insatisfatórios. Por exemplo, relatório do Ministério da Educação (MEC, 2003) indica que, em 2003, “59% dos alunos da 4.º série eram incapazes de qualquer leitura ou só conseguiam ler as frases mais simples, e que só 4,8% de crianças nessa série estaria com um nível de leitura considerado adequado para a sua idade”. Na série histórica do Saeb, pode-se observar uma ligeira queda nas médias de desempenho em língua portuguesa e matemática de 1995 a 2001, em praticamente todas as séries de ensino testadas. Esse resultado, em parte, deve estar associado com a maior incorporação de alunos à escola. Nos anos seguintes, ocorreu uma certa estabilização dos resultados.

Uma comparação internacional permite ter uma ideia da qualidade relativa da educação no Brasil. Desde 1995, o Brasil vem participando como convidado do PISA (Programa Internacional de Avaliação de Estudantes), coordenado pela OCDE. Esse é um teste de proficiência em matemática e leitura aplicado de três em três anos a uma amostra de alunos com 15 anos de idade dos países da OCDE e países convidados. O desempenho dos estudantes brasileiros foi bastante insatisfatório em todas as vezes que participou. Em 2000, por exemplo, o Brasil teve o pior desempenho dentre os trinta países que participaram.

Tendo avançado na quase universalização do ensino fundamental, a prioridade da política educacional deve ser a melhoria na qualidade do ensino. Principalmente, quando se tem em mente que a qualidade, e não a quantidade, da educação é uma variável mais importante para explicar crescimento econômico.³ No entanto, estudos sugerem que simplesmente aumentar os gastos não implica sua melhoria, quando mensurada pelo desempenho dos alunos nos testes de proficiência.⁴ Hanushek (2003) faz um extensiva revisão da literatura, para os Estados Unidos e internacional, sobre os possíveis impactos de diferentes políticas educacionais, que acarretam mais gastos em educação. Alguns exemplos dessas políticas são os seguintes: aumentar os salários dos professores, reduzir o número de alunos em sala de aula, exigir um maior nível de escolaridade dos professores e melhorar

¹ A taxa de escolarização líquida identifica o percentual da população em determinada faixa etária que se encontra matriculada no nível de ensino regular teoricamente adequado a essa faixa etária.

² Para maiores informações sobre o Fundef, ver Souza (2005) e Semeghini (2002).

³ Sobre a maior importância da qualidade, ver estudo de Hanushek & Kimko (2000).

⁴ Esses testes de proficiência capturam somente as habilidades cognitivas do indivíduo. Restringir os impactos da educação somente sobre essas habilidades é certamente uma visão simplista. Esse assunto será retomado com mais detalhes na terceira seção deste trabalho.

a infraestrutura nas escolas. Essas e outras variáveis afins não são significativas para explicar o desempenho escolar dos alunos e, por conseguinte, a qualidade da educação.⁵

O objetivo deste trabalho é discutir uma possível alternativa de política educacional que não seja simplesmente aumentar os gastos em educação e possa ser eficaz em melhorar a qualidade do ensino. Essa política é denominada *school accountability* (SA).⁶ Foi inicialmente implantada na década de 1980 na Inglaterra, tendo se transformado na principal política educacional dos Estados Unidos na década de 1990. Em linhas gerais, essa política tem os seguintes componentes básicos. Estabelece o conteúdo mínimo, para cada ano escolar, que o aluno tem de obter em cada disciplina. Obriga os estudantes a realizar testes de proficiência nessas disciplinas para checar o conhecimento adquirido. Adota como objetivo explícito de política educacional a melhoria do desempenho dos alunos nesses testes. Por fim, responsabiliza os professores/diretores da escola pelo resultado dos estudantes nos testes de proficiência. Inclusive, penalidades e bonificações podem ser estabelecidas para os gestores escolares em função desses resultados. Na prática, existem importantes diferenças na forma como o SA é de fato implementado, com impactos diferentes sobre a performance dos estudantes.

Inicialmente, este trabalho examina as diferentes políticas de SA adotadas nos Estados Unidos a partir da década de 1990, seus impactos sobre a qualidade do ensino e a sua eficácia, em termos de custo-benefício, em comparação com outras políticas. Como será visto, a bem-sucedida experiência americana com a SA indica que essa política deveria ser considerada para adoção no Brasil de forma a melhorar a qualidade do ensino.

Adicionalmente, este trabalho busca analisar os potenciais problemas associados à implantação da SA. Primeiro, essa política dá incentivos para os professores focarem no desenvolvimento das habilidades cognitivas dos alunos, pois em função dessas que seus trabalhos serão julgados e remunerados. Como decorrência disso, o desenvolvimento de habilidades não cognitivas fica em segundo plano com possíveis consequências negativas sobre a educação dos alunos. Segundo, os professores têm um incentivo em se envolver em atividades chamadas na literatura de *gaming*. No contexto da SA, isso significa ter atitudes indesejáveis para inflar artificialmente o desempenho dos alunos nos testes de proficiência. Por exemplo, aumentar a taxa de reprovação e/ou evasão escolar para evitar que os piores alunos façam os testes e, com isso, reduzira a nota da escola ou da turma. Por fim, existe uma dificuldade de mensurar a efetiva contribuição da escola para o desempenho do aluno. Sabe-se que variáveis não relacionadas a ela afetam a performance do estudante como, por exemplo, o nível educacional dos pais. Como a bonificação

⁵ Para evidência empírica com resultados semelhantes para o Brasil, ver Curi (2005).

⁶ Este trabalho utilizará o termo em inglês, pois não existe boa tradução para o português da palavra *accountability*. Alguns autores traduzem como responsabilização. Mas, na verdade, a política de SA incorpora características que vão além da simples responsabilização, como será visto em seguida.

ou a penalidade na política de SA decorre das notas dos alunos nos testes, é importante conseguir distinguir os diferentes fatores que afetam os resultados para que injustiças não sejam cometidas. Como será visto, esses problemas podem ser contornáveis e não devem ser considerados empecilhos à adoção da SA.

Além desta introdução, este trabalho tem três seções. Na segunda, é discutida a experiência dos Estados Unidos com o SA, enfatizando a existência de diferentes tipos e seus impactos diferenciados sobre o desempenho dos alunos nos testes de proficiência. Em seguida, são investigados alguns potenciais problemas ligados à implantação do SA. A última seção apresenta as conclusões.

SCHOOL ACCOUNTABILITY NOS ESTADOS UNIDOS

A partir da década de 1990, diversos estados americanos passaram a experimentar diferentes tipos de política de SA. Em 1996, 12 estados já haviam implantado alguma política desta natureza. Esse número pulou para 39 estados (de um total de 50) no ano 2000.

O Consortium for Policy Research in Education (CPRE) realizou uma extensa pesquisa sobre as diferentes políticas educacionais adotadas pelos 50 estados americanos nos anos 1999-2000.⁷ Baseado nesse banco de dados, Carnoy e Loeb (2002) criaram uma escala de zero a cinco, para classificar os diferentes estados americanos em função do grau de pressão externa exercida sobre as escolas para melhorar o desempenho escolar dos alunos nos testes de proficiência. A classificação que o estado recebe é a seguinte:

- 0, se ele não estipula testes de proficiência que todos os alunos têm de fazer ou não estipula padrões educacionais mínimos a serem seguidos pelas escolas ou distritos (exemplos: Iowa e Nebraska);
- 1, se ele estabelece testes de proficiência para todos os alunos nos ensinos fundamental e médio, com os resultados disponibilizados (exemplos: Montana e Nevada);
- 2, se além de estabelecer testes, estabelece penalidades/bonificações moderadas para as escolas em função dos desempenhos dos alunos ou estipula uma prova estadual que o aluno tem de passar para adquirir o diploma do ensino médio (exemplos: Virginia e Wisconsin);
- 3, se além de estabelecer testes, estabelece penalidades/bonificações moderadas para as escolas em função dos desempenhos dos alunos e estipula uma prova estadual que o aluno tem de passar para adquirir o diploma do ensino médio (exemplos: Ohio e Carolina do Sul);
- 4, se além de estabelecer testes, estabelece fortes penalidades/bonificações

⁷ Essa pesquisa está disponível na internet no seguinte link: http://www.cpre.org/Publications/Publications_Accountability.htm.

para as escolas em função dos desempenhos dos alunos (exemplo: Kentucky e Novo México);

- 5, se além de estabelecer testes e fortes penalidades/bonificações para as escolas em função dos desempenhos dos alunos, estipula uma prova estadual que o aluno tem de passar para adquirir o diploma do ensino médio (exemplos: Carolina do Norte e Texas).

Dos seis níveis acima, pode-se dizer que o único que não pode ser considerado uma política de SA é o primeiro, nível zero, já que não adota testes de proficiência. O nível 1 é uma forma bem branda de SA, pois não atrela ao desempenho dos estudantes nenhuma consequência para as escolas em termos de penalidades/bonificações. No entanto, para esse nível, existe a possibilidade de que a sociedade, principalmente através das associações de pais, exerça uma pressão sobre os gestores escolares para melhorar a performance dos alunos, caso essa seja vista como insatisfatória. Os níveis 2, 3 e 5 incorporam um aspecto além do que pode ser considerado de uma política de SA: exigir que o aluno passe numa prova para receber o diploma do ensino médio. Dessa forma, a responsabilidade em fazer com que o aluno alcance o desempenho considerado satisfatório recai não somente sobre as escolas como também sobre os estudantes.⁸

Na experiência americana, as penalidades e bonificações utilizadas são diversas. No estado do Texas, por exemplo, que tem uma política bastante agressiva, as penalidades são gradativas. Inicialmente, a escola tem de justificar por que os seus alunos tiveram um desempenho considerado insatisfatório e submeter um plano de melhora para a Secretaria Estadual de Educação. Caso os resultados não melhorem, é possível que seja formado um *board* por residentes locais para gerir a escola. Como última instância, a escola pode ser fechada, com o estudante adquirindo o direito de se transferir para outra escola ou distrito. Em relação a bonificações, o estado da Carolina do Norte, por exemplo, distribui placas de “distinção” e dinheiro para as melhores escolas. Essas são aquelas nas quais pelo menos 90% dos alunos alcançaram uma performance acima da estabelecida como mínima e que obtiveram um crescimento do desempenho dos alunos dentro do esperado. Nesse caso, cada membro do *staff* e cada assistente de professor da escola recebem, respectivamente, US\$ 1500 e US\$ 500.

Com a aprovação pelo Congresso dos Estados, no ano de 2001, do No Child Left Behind Act (NCLB), todo estado americano passou a ser obrigado a estipular uma política que o leve para o nível 4 da escala apresentada acima.⁹ Em particular, o NCLB obriga os estados a definir um plano que garanta que os alunos, num

⁸ O termo técnico em inglês para essa política que obriga o aluno a passar numa prova para obter o diploma do ensino médio é *student accountability*. Essa política alternativa, mais agressiva do que simplesmente adotar o SA, no entanto, não é o foco deste trabalho.

⁹ É interessante notar que o NCLB é uma lei federal que estipula regras a serem seguidas por todos os estados no que diz respeito a sua política educacional. Isso ocorre apesar do fato de que os Estados Unidos são uma federação na qual os estados, em geral, possuem grande autonomia para definir as suas políticas.

prazo de 12 anos, adquiram um nível considerado satisfatório de proficiência nas disciplinas básicas.¹⁰

Cada estado tem autonomia para estruturar o seu próprio plano. No entanto, tem de constar as seguintes linhas mestras básicas. Primeiro, o estabelecimento de padrões educacionais em cada disciplina e em cada ano escolar. Segundo, a obrigatoriedade de implantar testes de proficiência para averiguar se os padrões escolares estabelecidos são cumpridos. Os resultados agregados, das escolas ou dos distritos, são tornados públicos. Por fim, a implantação de um sistema de incentivos no qual a escola recebe bonificações ou penalidades em função do desempenho escolar dos seus alunos. Em particular, com a mensuração, é possível avaliar se o desempenho dos alunos está seguindo uma trajetória considerada satisfatória. Caso isso não ocorra, as escolas têm de ajudar os alunos para que alcancem desempenhos aceitáveis. As alternativas podem ir desde oferecer serviços suplementares até retirar o aluno da escola pública e ajudá-lo a encontrar uma outra escola mais eficiente. Em última instância, a escola pode, inclusive, ser fechada.

À luz da experiência recente dos Estados Unidos com SA, é possível fazer uma avaliação inicial dos seus impactos sobre a qualidade da educação. Três tipos de estratégias foram utilizados recentemente com este fim.

A primeira estratégia é uma análise de *cross-section* desenvolvida por Carnoy e Loeb (2002). Utilizando a classificação mencionada acima que distingue os estados americanos em função do grau de pressão externa exercida sobre as escolas para melhorar o desempenho dos alunos, eles examinam se a escala recebida pelo estado está relacionada com maiores taxas de crescimento das notas dos alunos no teste de matemática do NAEP (National Association of Educational Progress) no período de 1996-2000. Os resultados sugerem que estados com políticas mais agressivas de SA obtiveram maiores acréscimos nas notas. Em particular, um aumento de dois níveis na classificação do estado na escala de zero a cinco implica um ganho de 2,8% na proporção dos alunos da 8ª série da raça branca que alcançam pelo menos o nível básico de conhecimento. Para os alunos das raças negra e latina, este incremento é mais significativo sendo, respectivamente, aproximadamente igual a 5% e 9%.

A segunda estratégia é uma investigação utilizando dados de painel, feita por Hanushek e Raymond (2004), incorporando também todos os estados americanos na análise. Essa investigação com dados de painel se tornou possível porque os diferentes estados americanos introduziram em anos diferentes ao longo da década de 1990 o sistema de SA. Ao mesmo tempo, existe uma importante diferenciação no tipo do sistema de SA implantado nos diferentes estados. Hanushek e Raymond (2004) distinguem dois tipos. Um tipo é aquele que simplesmente divulga publicamente os resultados dos testes de proficiência dos alunos, chamado de *report card states*. O segundo tipo, além de reportar publicamente os resultados, atrela conse-

¹⁰ Para maiores detalhes sobre o NCLB, ver Hanushek (2004).

quências para as escolas em termos de bonificações ou punições em função dos mesmos, chamado *consequential states*.

A evidência empírica encontrada indica que a melhora do desempenho dos alunos ao longo da década de 1990 somente ocorre nos estados com o segundo tipo de SA (*consequential states*). Esse resultado sugere que uma política pouco agressiva, de simplesmente tornar público as notas dos alunos nos testes de proficiência, não é suficiente para acarretar uma melhora nas suas performances. Em outras palavras, simplesmente contar com uma possível pressão externa da sociedade, principalmente dos pais, para as escolas com resultados insatisfatórios se sentirem compelidas a melhorarem as notas dos seus alunos não parece ser uma política eficaz. É necessário a implantação de uma política mais agressiva, com incentivos diretos para os professores/diretores da escola, com consequências positivas ou negativas em função do desempenho dos alunos. Em suma, os resultados sugerem que “just reporting results has minimal impact on student performance and that the force of accountability comes from attaching consequences such as monetary awards or takeover threats to school performance” (Hanushek & Raymond, 2004).

É interessante tecer dois comentários finais sobre os resultados obtidos por Hanushek e Raymond (2004). Primeiro, diz respeito à magnitude da melhora do desempenho dos alunos nos *consequential states*. A melhora é significativa, mas não é espetacular, sendo algo em torno de um crescimento de 0,2 desvio padrão. Por fim, a evidência encontrada sugere que a adoção da política de SA nos estados americanos na década de 1990 foi capaz de reduzir somente a diferença de performance entre os alunos das raças branca e latina, não ocorrendo o mesmo entre os estudantes das raças branca e negra.

Ao contrário das duas primeiras estratégias, existe uma outra utilizada na literatura para avaliar os impactos da SA sobre a qualidade da educação. Ao invés de realizar uma análise que leva em consideração todos os estados americanos simultaneamente, vários estudos examinam, utilizando dados de painel, o que ocorre com o desempenho dos alunos nos testes de proficiência nas escolas de um determinado estado (ou cidade) que resolve adotar o regime de SA. Alguns exemplos são: Smith e Mickelson (2000) em Charlotte-Mecklenburg, Toenjes et al. (2000) e Carnoy, Loeb e Smith (2001) para o estado do Texas, e Jacob (2002) para a cidade de Chicago. Em geral, pode-se dizer que os resultados encontrados sustentam a visão de que a adoção da SA melhora a performance dos alunos.

A evidência apresentada acima, com os resultados preliminares da experiência dos estados americanos até o momento, sugere que a política de SA pode ser uma alternativa para melhorar a qualidade da educação. Aqueles estados (ou cidades) que adotaram uma política mais agressiva, chamadas com consequência ou com escalas mais elevadas de pressão externa sobre as escolas, registraram uma maior taxa de crescimento das notas dos seus alunos nos testes de proficiência. Ademais, os impactos mostraram-se mais significativos sobre alunos provenientes da minoria

latina. Em alguns casos, os alunos da raça negra se beneficiaram mais do que os da raça branca.¹¹

No entanto, é importante ressaltar que a melhora nos testes de proficiência devido ao SA é da ordem de 0,2-0,3 desvio padrão da nota média dos alunos.¹² Não é um resultado extraordinário, mas é significativo se comparado com o fracasso de várias outras reformas.¹³ É possível que, com o aprimoramento dos mecanismos de incentivos, os resultados das políticas de SA se tornem mais significativos ainda no futuro.

Talvez mais relevante ainda seja a eficácia relativa da política de SA numa análise custo-benefício *vis-à-vis* outras políticas. Sobre esse assunto, Jacob (2002) contrasta a política de SA com aquela que reduz o número de alunos em sala de aula. No experimento chamado Tennessee STAR, a redução do tamanho da sala de aula nos primeiros anos do ensino fundamental de 22 para 15 alunos teve um aumento do desempenho dos alunos de 0,2 desvio padrão. O custo estimado desta política é de US\$ 3.501 dólares por aluno por ano. Para um mesmo impacto sobre o desempenho dos alunos, o custo da SA é bem menor. Jacob (2002) faz uma estimativa de US\$ 447 dólares por aluno para o programa de SA na cidade de Chicago, que corresponde a algo em torno de 13% do custo da redução do tamanho das classes.

É importante ressaltar, no entanto, que não existe evidência de que, quando adotada em âmbito mais amplo, além de um simples experimento, a política de redução dos tamanhos das classes tenha um impacto significativo sobre o desempenho dos alunos. Na verdade, não existe uma política em âmbito nacional com impacto tão significativo quanto a SA. Essa evidência reforça a tese da eficácia relativa dessa política.

POTENCIAIS PROBLEMAS DO SA

Esta seção analisa algumas potenciais dificuldades que têm de ser levadas em consideração quando da implantação do sistema de SA.

Habilidades cognitivas vs. não cognitivas

A política de SA mais recomendada, por ter impactos mais significativos, estipula que professores/diretores da escola devem ser premiados ou bonificados em

¹¹ Adiante serão mostradas evidências que sugerem que esses resultados não são explicados por comportamentos oportunistas dos professores para inflar artificialmente o desempenho dos alunos nos testes.

¹² Dois exemplos podem dar uma ideia do que significa o aumento de 0,2-0,3 desvio padrão na média dos alunos. A diferença da média das notas nos testes de proficiência dos brancos e negros nos Estados Unidos é de aproximadamente 1 desvio padrão da média nacional. Também é de aproximadamente 1 desvio padrão da média das notas dos alunos que participam do PISA a diferença na média dos alunos brasileiros e dos alunos da OECD. Para maiores informações sobre os resultados do PISA, ver o site <http://www.pisa.oecd.org>.

¹³ Sobre o fracasso de várias dessas experiências, ver Hanushek (2003).

função do desempenho dos seus alunos nos testes de proficiência. Como resultado, os mesmos têm um grande incentivo em focar o seu trabalho no aprimoramento e desenvolvimento das habilidades cognitivas dos estudantes, cobradas nas provas, possivelmente em detrimento das habilidades não cognitivas. Portanto, um possível efeito colateral não desejável decorrente da aplicação do SA é o desbalanceamento na formação educacional do estudante. Em outras palavras, é possível que exista uma discrepância entre o aprendizado alcançado pelo aluno que é auferido pelos testes de proficiência e aquele desejável.¹⁴

Para analisar esse potencial problema, é importante primeiro averiguar quais são os possíveis impactos, para os estudantes e para a sociedade, de um melhor desempenho nos testes de proficiência. Se esses impactos forem positivos, pode ser interessante adotar a política de SA, mesmo que ocorra o “desbalanceamento” mencionado acima.

No que diz respeito à sociedade, a evidência empírica existente sugere que os países com melhor qualidade do ensino, mensurado pelas notas nos testes de proficiência, obtêm maiores taxas de crescimento econômico. Hanushek e Kimko (2000) realizaram uma regressão para explicar crescimento econômico no período 1960-1990 utilizando uma *cross-section* com 31 países, entre desenvolvidos e em desenvolvimento. Os resultados apontam que o aumento de um desvio padrão na nota dos testes de proficiência¹⁵ acarretam um acréscimo na taxa de crescimento do PIB *per capita* de 1,4% ao ano. Em contraste, o mesmo acréscimo de um desvio padrão nos anos de escolaridade está associado com o aumento de somente 0,26% na taxa de crescimento do PIB *per capita* por ano. Esse resultado indica a importância de encontrar políticas públicas que causem impactos significativos sobre a qualidade da educação, como a SA.

O melhor desempenho dos estudantes nos testes de proficiência acarreta benefícios para os mesmos também, não somente para a sociedade através de um maior crescimento do PIB *per capita*. Estudos mostram que alunos com maiores notas nesses testes obtêm maiores salários depois de formados nas escolas.¹⁶ Por exemplo, os resultados encontrados por Murphy e Peltzman (2004) sugerem que um decréscimo de um desvio padrão na nota corresponde a uma redução de 3% no salário médio dos jovens que ingressam na força de trabalho. Como era de esperar, dada a relação da produtividade de um indivíduo com o seu salário e da produtividade dos trabalhadores e o crescimento econômico de um país, estudos também mostram que as notas dos testes de proficiência dos estudantes afetam suas produtividades quando trabalhadores. Por exemplo, Bishop (1989) conclui que a

¹⁴ Crítica semelhante é feita sobre algumas escolas de ensino médio que priorizam a preparação do aluno para o vestibular, em vez de se preocupar com uma formação mais ampla e completa.

¹⁵ Os testes de proficiência utilizados na análise são os International Math and Science Tests, realizados em diferentes anos.

¹⁶ Para exemplos de evidência empírica sobre esse assunto, ver Boissiere et al. (1985), Heckman (1995), Murnane et al. (1995) e Murphy & Peltzman (2004).

piora no desempenho dos estudantes americanos na prova conhecida como SAT, iniciada em 1967 e se estendendo até 1980, é coincidente com o declínio do crescimento da produtividade da economia dos Estados Unidos. Por fim, os estudantes que apresentam melhores desempenhos nos testes de proficiência têm maior probabilidade de concluir o ensino médio e tendem a ingressar e concluir o ensino universitário com mais frequência¹⁷.

A evidência empírica apresentada acima, sugerindo impactos positivos dos resultados nos testes de proficiência sobre os indivíduos e a sociedade, corrobora a visão de que é desejável a adoção de políticas públicas na área da educação voltadas para elevar o desempenho dos estudantes. Em outras palavras, do ponto de vista do *policymaker*, é interessante ter como objetivo de política a melhoria das habilidades cognitivas dos alunos. Essa é exatamente uma importante característica da política de SA.

No entanto, focar nas habilidades cognitivas é somente um lado da moeda. Heckman et al. (2006) mostram a importância do desenvolvimento de habilidades cognitivas e não cognitivas¹⁸ para determinar comportamentos sociais (como probabilidade de se envolver com drogas no futuro, de ser preso até os 30 anos de idade, engravidar até os 18 anos de idade e participar de atividades ilegais) e variáveis econômicas (como salários, empregabilidade e nível de escolaridade). A conclusão do trabalho é a seguinte: “For many dimensions of behavior [...] noncognitive ability is as important, if not more important, than cognitive ability. Our findings challenge a pervasive view in the literatures in economics and psychology that cognitive ability, as measured by test scores, plays a dominant role in explaining personal achievement. Although cognitive skills explain much more of the variance of (log) wages, their effects on (log) wages (as measured by skill gradients) are similar to the effects of the noncognitive traits. In fact, noncognitive skills are about equally strong in many outcomes and are stronger for some outcomes”.

Em vista desses resultados e a importância das habilidades não cognitivas, pode-se questionar a adequabilidade de se implantar uma política como a de SA, que foca no desenvolvimento das habilidades cognitivas. Sobre esse assunto, Cunha et al. (2005) escrevem: “Current systems of evaluating educational reforms are based predominantly on changes in scores on cognitive tests. These tests capture only one of the many skills required for a successful life. A more comprehensive evaluation of educational systems would account for their effects on producing the noncognitive traits that are also valued in the market”.¹⁹

¹⁷ Sobre esse tópico, ver Bishop (1991), Rivkin (1995) e Hanushek (1996).

¹⁸ Exemplos de habilidades não cognitivas importantes são: motivação, persistência, autocontrole e preferência temporal.

¹⁹ Em uma discussão mais completa sobre qual é a política educacional mais eficiente deve-se levar em consideração não somente os benefícios como os custos. É provável que os custos de elevar as habilidades não cognitivas sejam maiores que aumentar as cognitivas. Se esse é o caso, focar nas últimas pode ser mais adequado mesmo sabendo dos impactos positivos das primeiras.

A dificuldade em adotar um sistema de avaliação educacional que leve em consideração também as habilidades não cognitivas é que, ao contrário das habilidades cognitivas, não existe nenhum indicador capaz de sumariá-las e servir como parâmetro para mensurar o desenvolvimento do aluno nessas aptidões. É pouco provável também que surja algo nesse sentido. A questão que se coloca, portanto, é a seguinte. Levando em consideração os pontos apresentados até agora, do ponto de vista da política educacional, é mais eficiente manter o atual regime ou implantar um regime de SA, que prioriza o desenvolvimento de habilidades cognitivas?

Para melhor responder a esta pergunta, pode-se pensar, em princípio, numa formulação na qual a produtividade (ou o salário) de um indivíduo depende dos insumos, habilidades cognitivas e não cognitivas, adquiridos na escola.²⁰ O objetivo da política educacional é alcançar a maior produtividade possível, dados os recursos disponíveis.

Com os incentivos atuais na escola, os estudantes adquirem certa quantidade de cada um dos insumos e uma correspondente produtividade. Com a alteração na política e a introdução da SA, o sistema de incentivos passa a priorizar as habilidades cognitivas. É muito provável que, como consequência, os alunos aumentem as suas habilidades cognitivas e diminuam as suas habilidades não cognitivas na escola. Como os incentivos do atual regime são baixos e as atuais habilidades, tanto cognitiva como não cognitiva, adquiridas pelos alunos são possivelmente muito baixas, não é improvável que o aumento nas habilidades cognitivas com o SA mais que compensem as possíveis reduções nas habilidades não cognitivas. O efeito líquido pode ser um aumento da produtividade média dos indivíduos. Esse resultado, no entanto, é influenciado pelo grau de complementaridade e substituíbilidade dos dois tipos de insumos. Quanto mais complementares forem os dois tipos de habilidades, é mais provável que a introdução da política de SA reduza a produtividade média do indivíduo. O inverso ocorre quanto mais substitutos forem os insumos.

A discussão acima, no entanto, desconsidera um aspecto importante no processo de desenvolvimento das habilidades dos indivíduos, qual seja, o fato de existir uma interdependência das habilidades cognitivas e não cognitivas. Sobre isso Cunha et al. (2005) escrevem: “Higher levels of noncognitive skills promote success on achievement tests even when they do not affect IQ. This effect operates because noncognitive skills affect schooling and schooling raises measured achievement”.²¹

Levando em consideração esse efeito, pode-se pensar que, para aumentar as

²⁰ Formalmente, pode-se pensar num modelo simples no qual a produtividade “P” (ou o salário) de um indivíduo depende dos insumos, habilidades cognitivas (HC) e não cognitivas (HNC), de acordo com a seguinte função de produção: $P = f(HC; HNC)$, sendo “f” uma função com as propriedades usuais. Uma estrutura semelhante é utilizada em Cunha et al. (2005).

²¹ Para mais detalhes, ver Hansen, Heckman & Mullen (2004) e Heckman, Larenas & Urzua (2004).

habilidades cognitivas dos indivíduos, é necessário ou imprescindível desenvolver as suas habilidades não cognitivas.²² Desta forma, ao contrário do sugerido acima, a introdução da política de SA e o seu mecanismo de incentivos não necessariamente acarretarão uma redução das habilidades não cognitivas dos alunos. Isto porque os professores/diretores na escola terão necessariamente que se preocupar que os alunos desenvolvam essas competências para conseguirem aumentar suas habilidades cognitivas e, por conseguinte, alcançarem um melhor desempenho nos testes de proficiência, baseado no qual eles serão julgados. Por conseguinte, o mais provável é que as habilidades cognitivas sejam mais desenvolvidas com a SA, como a experiência americana sugere, sem ser necessariamente em detrimento das habilidades não cognitivas.

Pode-se concluir da análise nesta seção os seguintes pontos. Primeiro, ter uma política educacional que foque na melhoria das notas nos testes de proficiência, como a de SA, pode ser interessante em vista dos seus impactos positivos sobre a sociedade e o futuro profissional dos alunos. É verdade que, dado os incentivos intrínsecos na política de SA, professores e diretores das escolas tendem a priorizar o desenvolvimento das habilidades cognitivas dos estudantes vis-à-vis às não cognitivas. Mesmo levando em consideração esse fator, o efeito líquido sobre a produtividade (e salários) dos futuros profissionais tende a ser positivo. Isto porque o desenvolvimento das habilidades não cognitivas afeta o desempenho nos testes de proficiência. Portanto, de fato, essas habilidades não podem ser desconsideradas pelos gestores educacionais quando submetidas a uma política como a SA. Em suma, o potencial problema da adoção da SA discutido nessa seção não parece ser significativo.

Gaming

Existe um problema conhecido como *gaming* na literatura de economia organizacional. Esse é entendido como uma forma de comportamento oportunístico por parte dos empregados de forma a afetar a avaliação de sua performance. Ele ocorre quando o empregado é avaliado e, por conseguinte, remunerado, em função de um indicador objetivo. Para melhorar esse indicador, ele se envolve em atividades disfuncionais. Por exemplo, um funcionário remunerado em função da sua produção total pode reduzir a qualidade do produto para conseguir aumentar a produção.²³

²² Levando em consideração esse efeito, a modelagem da nota de rodapé 20 se alteraria para: $P = f[HC(HNC); HNC]$.

²³ Sobre o assunto *gaming*, ver Brickley, Smith & Zimmerman (2004). Um exemplo curioso relatado nesta referência ocorreu na empresa Lincoln Electric. Ela instalou medidores nas máquinas de escrever das secretárias que tiveram seus salários atrelados ao número de caracteres por elas datilografados. Essa política foi abandonada quando se descobriu que uma secretária, no horário de almoço e intervalos, pressionava repetidamente uma tecla na máquina sem executar nenhum trabalho produtivo, tendo recebido um salário bem superior às demais como decorrência disto.

No contexto de SA, esse é um fenômeno que pode ocorrer. Como visto acima, os professores e/ou diretores de escola sofrem penalidades ou são bonificados em função do desempenho dos seus alunos nos testes de proficiência. Em decorrência disso, eles podem tomar atitudes indesejáveis para “inflar” artificialmente as notas dos alunos e serem recompensados por isso. Por exemplo, Levitt e Dubner (2005) mostram evidências que, nos Estados Unidos, professores passaram, em provas por eles aplicadas, “cola” para os seus próprios alunos. O objetivo era inflar as médias dos alunos para receber maiores bonificações.

Mesmo sem serem os responsáveis por aplicar as provas, existem pelo menos duas outras maneiras que os professores podem utilizar para tentar aumentar o desempenho dos seus alunos com o intuito de se beneficiar com isso. A primeira delas é incentivar os piores alunos a abandonar a escola. Ao aumentar a evasão escolar, somente os melhores alunos realizam os testes de proficiência. Como resultado, a escola passa a registrar um resultado melhor. A segunda maneira é aumentar a exigência sobre os alunos, sem necessariamente uma contrapartida na melhoria da qualidade da educação, simplesmente para elevar as taxas de reprovação. Pelo menos nos Estados Unidos, os alunos com maiores dificuldades acadêmicas que forem transferidos para classes especiais ficam isentos de fazer as provas de proficiência. A consequência também é um melhor desempenho dos alunos. Note que essas duas artimanhas levam ao registro de uma melhor performance dos estudantes, mas, se elas de fato ocorrerem, são decorrentes de posturas inadequadas e indesejáveis.

À luz da experiência nos Estados Unidos, a literatura econômica examinou se a implantação do SA efetivamente levou os professores a adotarem o comportamento oportunístico discutido acima. Em outras palavras, a pergunta central é se os resultados apresentados na segunda seção deste trabalho, que mostram uma melhora no desempenho dos alunos nos testes de proficiência como decorrência da implantação do SA, são reflexos realmente da melhoria da qualidade do ensino ou são explicados pelo *gaming* por parte dos professores aumentando as taxas de evasão e/ou reprovação escolares.

Os trabalhos apresentados na segunda seção, que utilizam uma base de dados com todos os estados americanos, realizando ou uma análise de *cross-section* ou painel, investigam a possível existência de *gaming*. Carnoy e Loeb (2002) não encontram nenhuma evidência de que a melhora dos desempenhos dos alunos nos testes seja resultado de modificações ou na taxa de retenção ou reprovação dos alunos. Hanushek e Raymond (2004) mostram que o efeito da SA, nos *report card states* e *consequential states*, sobre as taxas de retenção dos alunos é insignificante (em termos de magnitude e significância estatística), uma vez considerado a tendência de longo do tempo. Sobre esse assunto, Hanushek e Raymond indicam que as estimativas “do not indicate that gaming is irrelevant to accountability but they do suggest caution in interpreting analyses of the magnitude of gaming under accountability systems. If such gaming were generally important in the case of special education, it should show up in the national data – but it does not”.

Existem vários outros estudos que investigaram o efeito, num determinado

estado ou cidade, da SA sobre as taxas de reprovação e evasão escolar: Jacob (2002) em Chicago; Koretz e Barron (1998) em Kentucky; Deere e Strayer (2001), Cullen e Reback (2002), Haney (2000), Carnoy, Loeb e Smith (2001) e Toenjies e Dworkin (2002) no Texas; Figlio e Getzler (2002) na Flórida. Os resultados são mistos, com alguns estudos apontando de fato um aumento da taxas mencionadas acima como consequência de *gaming* e comportamentos oportunistas por parte dos professores. Por exemplo, Jacob (2002) mostra que, em Chicago: “Prior to the accountability policy, the retention rate was roughly 4 to 5 percent in first grade, 2,5 percent in second grade and a little over 1 percent in grades four, five and seven. Retention rates began to increase in 1996, possibly in anticipation of the new standards the students would face in 1997. In most grades, the rates peaked in 1997 and then declined somewhat. However, the first grade retention rate continued to increase over time. This is consistent with the fact that first grade is likely the first opportunity for retention for most students, while teachers in other grades take prior retentions into consideration when in deciding whether or not to hold a student back”.

A evidência existente até o momento, a partir da experiência nos Estados Unidos, não indica que *gaming* seja um problema muito significativo como decorrência da implantação do SA. Os ganhos decorrentes da política em termos de melhoria no desempenho dos alunos parecem mais do que compensar os possíveis custos associados ao *gaming*. Adicionalmente, existem dois outros motivos que fortalecem a visão de que esse não é mesmo um problema muito significativo.

O primeiro é apontado em Hanushek e Raymond (2002). Mesmo que ocorra *gaming*, os seus impactos tendem a ocorrer somente uma vez e serem de curto prazo. As taxas de reprovação e evasão escolares têm de ser continuamente crescentes para que as escolas apresentem contínuas melhoras nos seus resultados, o que é pouco razoável que aconteça.²⁴ Sobre esse assunto, Hanushek (2004) sugere: “The movement toward stronger accountability in schools has also suggested to many that there would be adverse consequences – more exclusions, higher dropout rates, a narrowing of the curriculum, and the like. While some existing research supports these presumptions, it appears that the negative impacts are likely to be considerably overstated. Importantly, many of the adverse effects that involve ‘gaming’ the system come from short-run incentives that are unlikely to be present over time”.

Mesmo que a estratégia de *gaming* fosse possível nos médio e longo prazos, existe um outro fator que dificulta a sua utilização por parte dos professores. Os responsáveis por estipular o arcabouço institucional que rege a política de SA estão atentos para esse possível comportamento oportunista por parte dos professores. E não parece ser muito difícil incorporar mecanismos que inibam o *gaming*. Por

²⁴ Sobre casos como o da fraude captada no trabalho de Levitt & Dubner (2005) apresentada acima, por exemplo, Hanushek & Raymond (2002) escrevem: “Take, for example, efforts to teach all students how to fill in mechanical scoring sheets for standardized exams. Once students know how to do this [...] it would not be expected to have any continuing effects on their scores as they progress through the grades. Similarly, any cheating on a given test must be repeated in subsequent years just to stay at the same level, but scores will only improve if the level of cheating is increased over time”.

exemplo, a legislação que rege a política de SA no estado do Texas estabelece critérios claros para uma escola ser considerada exemplar e ser passível de bonificação. Dentre esses critérios consta um no qual a taxa de evasão escolar tem de ser inferior a 1% do total de alunos. Portanto, não é possível obter vantagens com o sistema incentivando os piores alunos a sair da escola. Já na Carolina do Norte, um dos itens que compõem o indicador que avalia a escola é a porcentagem dos alunos que obtêm o padrão de competência do estado nos exames de proficiência. Deste modo, retira-se o estímulo para elevar a taxa de reprovação.

A partir da análise acima, pode-se concluir que, tanto do ponto de vista teórico quanto empírico, *gaming* atrelado à política de SA não é um problema significativo. Certamente, esse não deveria ser visto como um empecilho para a adoção da SA.

Mensuração da contribuição da escola

Como visto acima, uma das bases da política de SA é bonificar ou penalizar as escolas em função do desempenho dos seus alunos. Para esse esquema de incentivos ser eficaz, é fundamental que diretores/professores sejam recompensados com base na sua real contribuição para os resultados obtidos pelos estudantes. No entanto, pode não ser uma tarefa trivial obter essa verdadeira contribuição e, com isso, comprometer o bom funcionamento do regime de SA.

A dificuldade está relacionada com o fato de que vários fatores, fora do controle das escolas, afetam o desempenho dos alunos. Um dos fatores de maior importância é o background familiar. Por exemplo, o nível de educação da mãe é uma variável significativa para explicar a performance escolar de uma criança. Outros fatores são importantes tais como: a habilidade do indivíduo, fatores históricos da formação educacional do aluno e a qualidade/formação dos colegas de classe.

Portanto, simplesmente utilizar uma *cross section* e comparar a média das notas dos alunos nas diferentes escolas não é um procedimento adequado. É perfeitamente possível que a contribuição ou valor adicionado de uma escola (ou seja, quanto do aumento do aprendizado do aluno é decorrente do trabalho desenvolvido pela escola) para o aprendizado seja maior do que a de outra escola, mesmo sendo a média dos seus alunos inferior. E, portanto, num sistema de incentivo adequado, maior deveria ser a sua bonificação em comparação com um regime no qual a recompensa financeira da escola esteja relacionada somente com a média das notas dos alunos.

É fácil entender como isso pode ocorrer com um simples exemplo. Suponha duas escolas, “A” e “B”, que sejam idênticas em todos os aspectos com exceção de dois. Primeiro, os professores são mais eficientes e contribuem de maneira mais eficaz para a melhora no desempenho dos seus alunos em “A”. Segundo, o nível educacional dos pais dos seus alunos é mais elevado em “B”. Suponha também que, como resultado da influência destes dois fatores, a média das notas dos indivíduos seja maior na escola “B”. Ou seja, os maiores esforços/competência dos professores da escola “A” não são capazes de compensar o pior background familiar dos seus alunos. Um sistema de incentivos adequado deveria dar maior compensação finan-

ceira para os professores da escola “A”. No entanto, se a métrica utilizada for a média das notas dos alunos, isso não ocorreria e os professores da escola “B” receberiam uma remuneração variável maior.

Para contornar este potencial problema, uma alternativa é aquela utilizada pelo estado do Texas. As escolas são divididas em grupos com características semelhantes. Por exemplo, no mesmo grupo são colocadas escolas cujo perfil de renda dos alunos é parecido. Procura-se, dessa maneira, eliminar as principais diferenças dos fatores não escolares que afetam o desempenho dos estudantes. Com isso, as possíveis diferenças das notas dentro de cada grupo estão fortemente relacionadas com o trabalho realizado pela escola. Por conseguinte, aquelas com maiores e menores notas dentro de cada grupo recebem, respectivamente, as bonificações e as penalidades.

Na verdade, busca-se com isso atrelar o sistema de incentivos ao valor adicionado da escola. Existem outras alternativas para alcançar este objetivo.²⁵ A primeira é fixar uma determinada série escolar (digamos, a 4.º série do ensino fundamental) e comparar a média das notas dos alunos de um ano letivo para outro. Obviamente, o grupo de alunos é diferente e o efeito *cohort* pode ser relevante e influenciar o resultado. A segunda alternativa é comparar a média dos alunos numa determinada série com a obtida por este mesmo grupo de alunos na série escolar seguinte. A dificuldade dessa comparação é que esse grupo de alunos pode se alterar devido, por exemplo, a migração ou a evasão escolar. A última alternativa é mais demandante em termos de coleta de dados. Ela acompanha a evolução das notas de todos os alunos individualmente (ao invés da nota da escola). Com isso, ela pode levar em consideração as questões migratórias ou de evasão escolar quando computar a contribuição da escola. É importante salientar, no entanto, que se fatores como background familiar e habilidade do indivíduo influenciar também a taxa de crescimento do desempenho dos alunos, mesmo com essa alternativa mais elaborada, não será possível separar totalmente a contribuição da escola.

As dificuldades de mensuração da real contribuição da escola para o aprendizado do aluno não devem ser vistas como um empecilho para a adoção da política de SA. O foco deve ser implementar um sistema de incentivos no qual os professores sejam julgados por um indicador que seja o mais próximo possível do valor adicionado da escola. A experiência americana sugere que o impacto da SA sobre o desempenho dos alunos pode ser significativo mesmo que a métrica escolhida não seja perfeita.

CONCLUSÃO

Este trabalho analisa diversos aspectos relacionados à política de SA. Em linhas gerais, pode-se dizer que a SA deveria ser considerada pelos gestores da educação como uma alternativa eficaz para melhorar a qualidade da educação no Brasil.

²⁵ Para mais detalhes, ver Hanushek & Raymond (2002).

A experiência com a SA nos Estados Unidos, a partir da década de 1990, sugere que essa política propicia uma melhora significativa no desempenho dos alunos nos testes de proficiência, desde que acompanhada por um sistema de incentivos no qual as escolas recebem penalidades/bonificações em função do desempenho dos seus alunos. Os estudos mostram que o impacto pode chegar a 0,3 desvio padrão. Impactos semelhantes no Brasil corresponderiam a reduzir em aproximadamente 30% a diferença entre o desempenho da média dos países da OECD e do Brasil no teste internacional do PISA.

Uma grande vantagem da SA é a sua eficácia relativa numa análise custo-benefício *vis-à-vis* outras políticas. Por exemplo, Jacob (2002) faz uma estimativa de US\$ 447 dólares por aluno para o programa de SA na cidade de São Paulo. Este valor corresponde a 13% do custo da redução do tamanho das classes do experimento chamado Tennessee STAR, que obteve o mesmo impacto sobre o desempenho dos alunos.

Críticas à política de SA não devem ser vistas como empecilhos sérios para a sua implantação. A primeira refere-se ao fato de o foco do aprendizado ser o desenvolvimento de habilidades cognitivas, que são mensuradas pelos testes de proficiência, *vis-à-vis* as não cognitivas. No entanto, as evidências mostram que priorizar as habilidades cognitivas é benéfico para a sociedade, pois gera maiores taxas de crescimento na economia, e para o indivíduo, pois aumenta a sua produtividade e salário. Adicionalmente, dado que o desenvolvimento das habilidades não cognitivas afeta o desempenho nos testes de proficiência, os gestores da educação não podem desconsiderar as primeiras, mesmo que sejam recompensadas pelas últimas.

No Brasil, já está sendo feita uma tentativa de se avaliar, em teste de proficiência, algumas habilidades não cognitivas, através do Exame Nacional do Ensino Médio (ENEM). Neste, uma das cinco competências que o teste avalia é a de “recorrer aos conhecimentos na escola para elaboração de propostas de intervenção solidária na realidade, respeitando valores humanos e considerando a diversidade sociocultural”. Num regime em que os professores são avaliados em função do desempenho dos alunos nos testes, estes acabam influenciando o currículo. Por conseguinte, ao ser incorporado nos testes questões que avaliam habilidades não cognitivas, é possível corrigir o possível foco em habilidades cognitivas mencionado acima.

A segunda crítica à SA é a possibilidade de os professores se envolverem em comportamentos oportunistas (*gaming*) para “inflar” artificialmente as notas dos seus alunos e serem recompensados por isso. As duas maneiras mais óbvias de fazer isso é incentivar os piores alunos a abandonar a escola e aumentar a taxa de reprovação sem uma contrapartida na melhoria da qualidade da educação. É possível penalizar as escolas que tenham relativamente elevadas taxas de evasão e reprovação escolares e inibir tais ações. Alternativamente, pode-se utilizar uma combinação das duas metas para estabelecer a recompensa: resultado nos testes mais taxa de retenção/aprovação, apesar da evidência empírica nos Estados Unidos não indicar que os impactos da SA sobre essa última sejam significativos.

Como visto, o impacto da SA sobre a qualidade da educação é significativo,

mas não revolucionário, girando em torno de 0,2-0,3 desvio padrão. Tendo em vista esses resultados, a sua implantação não deveria inibir a discussão e a análise de outras políticas alternativas que talvez possam gerar uma melhoria na qualidade da educação. Um sistema de *vouchers* e as *charter schools*, que possibilitam uma maior competição entre as escolas, são duas dessas alternativas.

REFERÊNCIAS BIBLIOGRÁFICAS

- BISHOP, J. (1991) "Achievement, test score, and relative wage"; em Marvin Koster, ed., *Workers and Their Wages*, Washington, DC: The AEI Press.
- BOISSIERE, M.; KNIGHT, J. & SABOT, R. (1985) "Earning, schooling, ability, and cognitive skill"; *The American Economic Review*, v. 75(5): 1016-30.
- BRICKLEY, J.; SMITH, C. & ZIMMERMAN, J. (2004) *Managerial Economics and Organizational Architecture*, Boston: McGraw Hill.
- CARNOY, M. & LOEB, S. (2002) "Does external accountability affect student outcomes? A cross-state analysis"; *Educational Evaluation and Policy Analysis*, v. 24, n. 4: 305-31, Winter.
- CARNOY, M., LOEB, S. & SMITH, T. (2001) "Do higher state test scores in Texas make for better high school outcomes?"; *CPRE Research Report Series RR-047*, November.
- CULLEN, J. & REBACK, R. (2002) "Tinkering toward accolades: School gaming under a performance based accountability system"; Department of Economics, University of Michigan, mimeo.
- CUNHA, F., HECKMAN, J., LOCHNER, L. & MASTEROV, D. (2005) "Interpreting the evidence on life cycle skill formation"; *NBER Working Paper #11331*.
- CURI, A. (2005) "A relação entre o desempenho escolar e os salários no Brasil"; mimeo.
- DEERE, D. & STRAYER, W. (2001) "Putting schools to the test: School accountability, incentives, and behavior"; *Working Paper #113*, Private Enterprise Research Center; Texas A&M University, March.
- FIGLIO, D. & GETZLER, L. (2002) "Accountability, ability and disability: gaming the system"; *NBER Working Paper #9307*.
- HANEY, W. (2000) "The myth of the Texas miracle in education"; *Education Policy Analysis Archives*, v. 8, n. 41, August.
- HANUSHEK, E. (1996) "A more complete picture of school resource policies"; *Review of Educational Research*, v. 66, n. 3: 397-409, Fall.
- HANUSHEK, E. (2003) "The failure of input-based schooling policies"; *The Economic Journal*, v.113, n. 485: 64-98, February.
- HANUSHEK, E. (2004) "United lessons about school accountability"; *CESifo DICE Report*, 2(4): 27-32, Winter.
- HANUSHEK, E. & RAYMOND, M. (2002) "Lessons about the design of state accountability systems"; em Paul E. Peterson and Martin R. West (ed.), *No Child Left Behind? The Politics and Practice of Accountability*, Washington, DC: Brookings.
- HANUSHEK, E. & RAYMOND, M. (2005) "Does school accountability lead to improved student performance?"; *Journal of Policy Analysis and Management*, v. 24(2, Spring): 297-327.
- HANUSHEK, E. & KIMKO, D. (2000) "Schooling, labor-force quality, and the growth of nations"; *The American Economic Review*, v. 90(5): 1184-1208, December.
- HANSEN, K., HECKMAN, J. & MULLEN, K. (2004) "Educational attainment, ability, and test scores"; *Journal of Econometrics*, 121(1-2): 39-98.
- HECKMAN, J. (1995) "Lessons from the bell curve"; *Journal of Political Economy*, v. 103, n. 5: 1091-1120.
- HECKMAN, J., STIXRUD, J. & URZUA, S. (2006) "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior"; *NBER Working Paper #12006*.

- HECKMAN, J., LARENAS, M. & URZUA, S. (2004) "Accounting for the effect of schooling and abilities in the analysis of racial and ethnic disparities in achievement test scores"; mimeo.
- JACOB, B. (2002) "Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago public schools"; *NBER Working Paper* #8968.
- KORETZ, D. & BARRON, S. (1998) *The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)*, Santa Monica, CA: RAND Corporation.
- LEVITT, S. & DUBNER, S. (2005) *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, New York, HarperCollins Publishers.
- MEC (2003) "Qualidade da educação: uma nova leitura do desempenho dos estudantes da 4.º série do ensino fundamental"; Ministério da Educação/INEP, Brasília, abril.
- MURNANE, R.; WILLETT, J.; LEVY, R. (1995) "The growing importance of cognitive skill in wage determination"; *Review of Economics and Statistics*, v. 77, n. 2, 1995: 251- 266, May.
- MURPHY, K. & PELTZMAN, S. (2004) "School performance and the youth labor market"; *Journal of Labor Economics*, v. 22(2): 299-327.
- RIVKIN, S. (1995) "Black/white differences in schooling and employment"; *Journal of Human Resources*, v. 30, n. 4: 826-52, Fall.
- SEMEGHINI, U. (2002) "Fundef: uma revolução silenciosa"; mimeo.
- SMITH, S. & MICKELSON, R. (2000) "All that Glitters is Not Gold: School Reform in Charlotte-Mecklenburg"; *Educational Evaluation and Policy Analysis*, 22(2): 101-27.
- SOUZA, P. (2005) *A Revolução Gerenciada*, Editora Financial Times BR.
- TOENJES, L., LORENCE, A. & HILL, A. (2000) "The lone star gamble: high stakes testing, accountability and student achievement in Texas and Houston"; mimeo.
- TOENJES, L. & DWORKIN., G. (2002) "Are increasing test scores in Texas really a myth, or is Haney's myth a myth?"; *Educational Policy Analysis Archives*, v. 10, n. 17, March.