

Dados longitudinais em educação: um componente essencial da abordagem de valor agregado no que se refere à avaliação de desempenho escolar¹

VALERIE E. LEE*

POR QUE ESTOU AQUI?

Estou muito feliz por ter sido convidada a participar deste seminário que marca os cinco anos do projeto Geres. Este é um estudo longitudinal, ou seja, os dados foram coletados por cinco anos consecutivos com as mesmas crianças enquanto realizavam seus primeiros anos de escolaridade. Gostaria de elogiar o esforço realizado e a coragem de enfrentar todos os problemas que um estudo dessa natureza traz. Quero destacar que não é usual encontrarmos estudos longitudinais fora dos Estados Unidos. No entanto, é a forma mais segura de se pesquisar o “efeito escola”. É sobre essa questão que falarei aqui.

¹ Palestra proferida no Seminário Abave/Geres, realizado em Belo Horizonte, Minas Gerais, Brasil, em 25 de agosto de 2010.

* Docente da Faculdade de Educação da Universidade de Michigan (University of Michigan), no programa em fundamentos, administração, pesquisa e políticas (velee@umich.edu).

Há dois componentes essenciais da pesquisa avaliativa do efeito-escola: 1) a utilização de métodos estatísticos multinível que levem em conta a natureza composta do processo educacional escolar (alunos acomodados nas salas de aula e estas nas escolas); 2) a necessidade de dados longitudinais sobre os alunos – dados dos mesmos alunos, acompanhados por vários anos. Nas pesquisas avaliativas multinível, realizadas com pesquisadores brasileiros, percebemos que são necessários dados longitudinais para realmente avaliar a qualidade das escolas brasileiras, o que se passa dentro delas e como tais dados podem ser coletados. Por isso, é essencial desenvolver mais estudos dessa natureza.

UM PROJETO OUSADO

Há vários esforços de coleta de dados educacionais internacionais com alta qualidade, entre eles podemos citar: o *TIMSS – Third International Mathematics and Science Study* (Terceiro Estudo Internacional de Matemática e Ciências), promovido pelo *IES – Institute of Education Sciences* (Instituto das Ciências da Educação); o *PISA – Programme for International Student Assessment* (Programa Internacional de Avaliação de Alunos), patrocinado pela OCDE (Organização para a Cooperação Econômica e o Desenvolvimento); e o *SACMEQ – Southern and Eastern Africa Consortium for Monitoring Educational Quality* (Consórcio do Sudeste Africano para o Monitoramento da Qualidade da Educação), patrocinado pela Unesco. No entanto, cada um desses conjuntos de dados de avaliação educacional, realmente excelentes – que são projetados principalmente para comparar e avaliar sistemas nacionais de ensino – partilham uma fraqueza fundamental: são todos transversais. Ou seja, cada um desses programas coleta periodicamente, em essência, os mesmos dados em amostras nacionalmente representativas de escolas e alunos (em geral com dois ou três anos de diferença). As amostras de escolas e alunos são diferentes em cada aplicação. Não é possível acompanhar a trajetória de *cada aluno*. Tais estudos geralmente se concentram em grupos de alunos de determinada idade (15 anos no PISA) ou de determinada série (6º ano no SACMEQ). Estudos realizados com o TIMSS, PISA ou SACMEQ destinam-se a fazer comparações ao longo do tempo no conjunto (ou seja, nas coortes), com uma de duas finalidades: a) avaliar o progresso no desempenho escolar conjunto em países específicos, e/ou b) fazer comparações entre países. Mas, o modelo não permite saber quanto se agregou realmente em conhecimento para os alunos.

O Brasil foi incluído nos dados do PISA em 2000, que focalizou a leitura. Nas comparações entre os 39 países participantes ficou em último lugar! Depois que o

Peru foi incluído numa coorte posterior do PISA, o Brasil passou então ao penúltimo lugar. As avaliações do PISA eram exigentes, difíceis, e provavelmente não representaram o que era ensinado e o como era ensinado nas escolas brasileiras. Vocês podem perguntar: “O que os profissionais da educação no Brasil aprenderam sobre suas escolas e seus alunos ao participar do PISA?” Gostaria de sugerir que pouco aprenderam. Destaco que os estudos transversais realmente não permitem efetivamente que os pesquisadores aprendam coisas que lhes sejam realmente úteis para descobrir como fazer para que suas escolas e salas de aula funcionem melhor e, conseqüentemente, os estudantes aprendam mais.

O Geres, que vem sendo realizado no Brasil por um conjunto de instituições universitárias e de pesquisa, é um estudo longitudinal de coleta de dados e representa um avanço em relação ao que vem sendo feito até aqui. Nesse projeto, os *dados foram coletados anualmente durante cinco anos, com as mesmas crianças, de modo a verificar o seu desempenho ao longo do ensino fundamental*, como já disse. As amostras do Geres, tecnicamente, não pretendiam ser nacionalmente representativas, mas, sim, incluir dados de amostras representativas de escolas, alunos e professores em cidades importantes do país, enfrentando o desafio de acompanhar as mesmas crianças por vários anos, o que pode acrescentar conhecimento significativo sobre suas aprendizagens, progressivamente. É um desafio.

Como reconhecem os profissionais da educação em todo o mundo, há vários e grandes desafios para a coleta de dados longitudinais. Menciono apenas alguns. Por exemplo, o que fazer quando as crianças mudam de uma escola para outra ou de uma localidade para outra? O que acontece quando a criança repete de ano (como é comum no Brasil)? A quem esses “repetentes” devem ser adequadamente comparados – aos seus homólogos de idade ou aos seus homólogos de série? E se uma criança muda de professor, na mesma escola, durante o ano letivo? Como é realmente possível localizar todas as crianças a cada ano? E se as crianças, professores ou escolas não preenchem os questionários? Como ter certeza de que os procedimentos de teste são comparáveis, tanto na configuração quanto no tempo, em todas as escolas e localidades? Como comparar os resultados de crianças nas diferentes séries? Como ter certeza de que o mesmo teste, numa classe em particular, não seja muito difícil para algumas crianças e muito fácil para outras? E se as crianças não conseguirem ler os testes ou os questionários? Ao pesquisar os pais, enviando questionários, o que fazer se não souberem ler? Como financiar esses esforços de coleta de dados caros e complicados?

Esses são apenas alguns dos inúmeros desafios que surgem no seio de qualquer esforço de coleta de dados longitudinais. Essas estão, naturalmente, entre as muitas

dificuldades que impediram estudos tão importantes como o TIMSS, o PISA, ou o SACMEQ de cogitar um projeto longitudinal (além da questão extremamente importante do custo). Embora os pesquisadores e estatísticos envolvidos com esses estudos reconheçam o valor dos dados longitudinais, também distinguem as muitas dificuldades (e têm sido desencorajados por elas) e os altos custos para a sua coleta. É por isso que gostaria de ressaltar o valor dos que conceberam, financiaram e realizaram o projeto Geres. O fato é que, apesar dos desafios, houve persistência.

Mas meu principal objetivo nesta fala, de fato, é oferecer-lhes algumas informações sobre uma proposição educacional em curso nos Estados Unidos – a avaliação do “valor agregado de professores” – para a qual os dados longitudinais sobre os alunos são absolutamente cruciais. Vou também discutir algumas das controvérsias que envolvem essa proposta educacional.

O QUE SIGNIFICA “VALOR AGREGADO”?

A estrutura e o planejamento da educação escolar nos Estados Unidos é sobretudo local – o governo federal fornece tradicionalmente apenas 10 a 15% do financiamento para a educação. Outros 40 a 50% vêm dos Estados. O restante é financiado com impostos municipais sobre a propriedade. No entanto, estamos caminhando lentamente em direção a um sistema mais centralizado.

Durante a última década, nos Estados Unidos, tem havido um enorme aumento no uso de testes educacionais, em especial de avaliações padronizadas para testar todos os alunos, a cada ano. Estados e cidades podem ter seus próprios testes. A cidade de Chicago tem avaliado anualmente seus alunos com testes padronizados – a partir do final do Grau 1 até o final do Grau 8. Os estudantes também são testados durante o ensino médio, mas com menos frequência.

Com a introdução da proposta nacional *No Child Left Behind* – NCLB (Nenhuma Criança Deixada Para Trás), no início do governo Bush (por volta de 2003 ou 2004), a aplicação de testes padronizados foi bastante ampliada, embora o desenvolvimento e a adoção de determinados testes tenham sido deixados para os Estados (e às vezes para os municípios). No início, os estudantes eram testados apenas em séries-alvo (geralmente nos Graus 3, 6 e 8) e os Estados podiam escolher quais exames e quais padrões curriculares usar (e muitos desenvolveram seus próprios testes e seus próprios padrões). As escolas que não atingem o *Adequate Yearly Progress* – AYP (Progresso Anual Adequado), conforme definido pela proposta NCLB, e calculado a partir de diferentes testes, são penalizadas com a redução dos recursos federais. Como era de esperar, as escolas que não alcançaram o AYP

têm sido, quase sempre, aquelas cujos alunos provêm de camadas sociais de baixa renda ou de minorias. A proposta NCLB ainda está em vigor, embora haja planos para alterá-la (e os aspectos políticos relacionados a ela). A NCLB gerou muita polêmica em torno de várias questões, incluindo as seguintes: nem todos os Estados usam os mesmos testes – não temos nenhuma prova nacional nos Estados Unidos;

- Estados foram autorizados a definir suas normas de conteúdo, de sucesso e “progresso”, o que resultou em padrões progressivamente mais baixos (às vezes chamados “corrida para o fundo” – “*race to the bottom*”);
- a forma como é definido o “progresso” é falha, pois nem todos os alunos (até nos mesmos Estados) têm sido testados todos os anos;
- muitos acreditavam que estudantes de minorias, de baixa renda ou com deficiências não deveriam ser avaliados pelos mesmos critérios, embora a legislação insistisse que deviam;
- muitos argumentaram que os testes padronizados (geralmente de leitura e matemática) não eram as únicas medidas de “sucesso” para as crianças na escola; e
- as escolas foram “culpadas” pelo baixo desempenho de seus alunos, mas não havia uma ideia clara de como a legislação poderia ajudá-las a melhorar o rendimento de seus alunos.

Embora a aplicação de testes nas escolas americanas tenha continuado a se expandir exponencialmente na última década, e os debates tenham sido muitos, a maioria dessas questões continua sem solução. O governo Obama não propôs eliminar a proposta NCLB; o Congresso teria que ser acionado para fazer isso. No entanto, o governo Obama lançou novas medidas por meio das quais a legislação federal pode ajudar as escolas a melhorar o desempenho de seus alunos.

POR QUE OS DADOS LONGITUDINAIS SÃO ESSENCIAIS?

Faço algumas considerações para simplificar a discussão. Primeiro, vamos supor que os testes padronizados **podem** avaliar adequadamente o desempenho dos alunos em determinado momento. Mas todos sabemos que algumas crianças chegam à escola com mais condições de ter sucesso nesses testes do que outras. Alguns anos atrás, meu colega David Burkam e eu publicamos um livro, *Inequality at the Starting Gate* (Desigualdade na Linha de Partida), no qual documentamos diferenças enormes no desempenho em leitura e matemática numa grande amostra

nacionalmente representativa de crianças que estavam na pré-escola nos Estados Unidos – diferenças por raça/etnia, *status* socioeconômico, experiência antes da escola, estrutura familiar, e (intrinsecamente) habilidade cognitiva. É claro que as crianças em todos os países entram na escola com vasta gama de habilidades cognitivas. Segundo, vamos assumir que essa diferença na entrada não seja algo pelo qual as escolas e professores devam ser responsabilizados. No entanto, todas as escolas e professores **devem ter a responsabilidade** de desenvolver as crianças a partir de onde estão ao entrar na escola, em termos de *status* cognitivo. Com certeza, por meio de nosso processo educacional, podemos nos esforçar muito para eliminar as diferenças no desempenho inicial dessas crianças, em razão de seu *status* social; mas temos também de atentar para aquelas que já trazem um cabedal maior. Em todo caso, as escolas e os professores devem ser responsabilizados pelo que as crianças “ganham” ou “se desenvolvem” em termos de capacidade cognitiva durante o tempo que frequentam determinada escola ou sala de aula. O ganho não é um termo absoluto, mas relativo a um ponto inicial.

De acordo com os estudiosos, as medidas de “valor agregado” de desempenho são definidas por dois elementos-chave. Em primeiro lugar, focam a mudança do desempenho dos alunos de um ano para outro (o que exige uma avaliação anual de todas as crianças). Em segundo lugar, o crescimento do desempenho do aluno é calculado para cada estudante dentro de determinada escola. Isso poderia ser problemático para as escolas onde a mobilidade de estudantes é elevada, o que ocorre em muitos dos nossos grandes distritos escolares urbanos. Por exemplo, em Chicago 40% dos alunos não terminam o ano na mesma escola em que começaram.

A essência do “valor agregado” é que os dados longitudinais são cruciais. Cada ano que a criança é educada em determinada escola ou sala de aula, irá aprender alguma coisa. Mas só podemos medir quanto a criança vai ganhar, progredir e aprender se levarmos em conta em que ponto a criança começou – em termos de desempenho cognitivo – quando entrou na escola ou na sala de aula. O que precisamos saber é quanto o desempenho aumentou durante esse período. Queremos medir **o valor do que foi agregado**, em termos de desempenho cognitivo da criança, durante esse período. Medir esse “valor agregado” exige dados longitudinais sobre as mesmas crianças. Se estamos interessados em saber quanto uma escola contribuiu para o “valor” cognitivo da criança, precisamos ter resultados do mesmo teste aplicado a cada criança ao entrar e sair da escola. Se nos interessa saber quanto determinada criança aprendeu ao longo de determinado ano, em determinada sala, com determinado professor, precisamos ter uma medida do desempenho dessa

criança, tanto quando entrou como quando deixou aquela sala de aula. Para fazer isso, temos de recolher dados sobre o desempenho de cada criança ao final de cada ano escolar. Precisamos também certificar-nos de que os testes que usamos são escalonados e equalizados de modo que os resultados das crianças nesses testes reflitam esses ganhos (e que o desempenho seja medido de forma adequada para as diferentes idades e classes). Esse é o desafio – e a promessa – dos dados longitudinais a serviço de programas escolares mais efetivos.

O PROGRAMA RACE TO THE TOP NOS ESTADOS UNIDOS

O *American Recovery and Reinvestment Act* (Ato de Recuperação e Reinvestimento Americano), chamado de “o pacote de estímulo”, foi transformado em lei pelo presidente Barack Obama em fevereiro de 2009. Essa lei incluiu um nível de investimento na educação, historicamente memorável – US\$ 4.35 bilhões – dentro do programa *Race to the Top* (Corrida para o Topo). Esse programa de financiamento competitivo tem como alvo os Estados que decidiram reformar substancialmente seus sistemas de ensino numa determinada direção. A “teoria da ação” que embasa essa legislação apoia-se nos meios propostos pelos sistemas estaduais pelos quais a aprendizagem dos alunos poderia ser melhorada, as deficiências no desempenho corrigidas e os índices de conclusão no ensino médio melhorados. Observem o uso do termo “aprendizagem dos alunos” e não “desempenho dos alunos”. Projetos estaduais bem sucedidos deveriam incluir planos inovadores em quatro áreas de reforma:

1. adoção de normas e avaliações que preparem os estudantes para serem bem sucedidos na faculdade e no trabalho (isso agora mudou para um conjunto de padrões normatizados comuns);
2. construção de sistemas de dados que meçam o progresso e o sucesso do aluno e que informem os professores e diretores sobre como eles podem melhorar a instrução;
3. recrutar, desenvolver, recompensar e manter professores e diretores eficazes, especialmente onde são mais necessários; e
4. reverter a situação das escolas de menor desempenho no país.

A primeira rodada de concessões do programa *Race to the Top* foi concedida a apenas dois dos quarenta Estados que se candidataram: Delaware (US\$ 100 milhões) e Tennessee (US\$ 500 milhões). Dos 38 Estados que se candidataram inicialmente,

mas que não foram beneficiados, 35 revisaram e reapresentaram seus projetos em junho de 2010, e, recentemente, essa lista foi reduzida pelo Departamento de Educação americano para dezoito Estados elegíveis. O resultado final estava previsto para ser anunciado em setembro de 2010.

Para poder se candidatar ao financiamento do programa *Race to the Top*, as agências estaduais de educação deveriam criar ou ampliar seus sistemas de dados longitudinais. A lei é bastante específica sobre o que esses sistemas devem conter, inclusive dados sobre alunos de educação especial, aprendizes de língua inglesa, programas de primeira infância, bem como informações sobre a mobilidade dos alunos, recursos humanos, finanças da escola, saúde dos estudantes e educação pós-ensino médio. O Secretário da Educação de Obama, Arne Duncan, chegou ao cargo após dirigir as escolas públicas de Chicago por muitos anos. Não deve ser nenhuma surpresa que Chicago tenha um sistema de dados quase exatamente igual ao descrito na legislação. Os melhores sistemas de dados (como o de Chicago) permitiriam acompanhar os alunos desde programas de educação infantil até a educação pós-ensino médio. As propostas para serem selecionadas deveriam explicar como esses sistemas de dados seriam usados para melhorar a aprendizagem dos alunos.

Existem fortes orientações que encorajam os Estados a criar sistemas que vinculem dados sobre o desempenho dos alunos a determinados professores. A legislação reza:

Propomos que, para que seja elegível no âmbito deste programa, o Estado não deve ter quaisquer barreiras legais, estatutárias ou regulamentares que impeçam vincular dados sobre o desempenho ou crescimento do aluno a professores para efeitos de avaliação de professores e diretores.

Usando o “incentivo” de um financiamento muito significativo para a educação pública nos Estados Unidos – num momento em que os orçamentos do país estão restritos – o programa *Race to the Top* tenta alterar radicalmente a maneira como é desenvolvida a educação pública. Um objetivo importante é a mudança fundamental na forma como os professores são avaliados, envolvendo tanto sua compensação como se são ou não mantidos na profissão. Na realidade, qualquer Estado com leis que proíbam a possibilidade de vincular o desempenho ou progresso do aluno à avaliação do pessoal não são elegíveis ao financiamento do *Race to the Top*, e os montantes de financiamento possíveis dessas concessões são bastante substanciais. Num momento da nossa história, em que as finanças estão muito limitadas para todos os programas públicos, o governo Obama – sob a gestão do Secretário

Duncan – está tentando direcionar uma reforma educacional num determinado sentido. Os Estados estão sendo pressionados a desenvolver sistemas abrangentes de dados longitudinais, com base no desempenho e progresso individual dos alunos. O progresso acadêmico de seus alunos deve ser um fator importante na avaliação dos professores. Isso nunca aconteceu antes nos Estados Unidos.

O Departamento de Educação americano também publicou recentemente padrões curriculares nacionais sobre o que os alunos devem saber e quando. Isso também causou um alvoroço, principalmente porque, historicamente, os Estados sempre tiveram o direito de desenvolver seus próprios padrões. Para serem elegíveis para a segunda fase do *Race to the Top*, os Estados tinham de concordar em adotar essas normas. Meu próprio Estado – Massachusetts – relutou em aprovar as normas nacionais, pois alegou que seus padrões eram mais rigorosos. No entanto, acabou adotando as normas nacionais e está na disputa para o *Race to the Top*. Alguns acreditam que, uma vez implantados os padrões curriculares nacionais, um exame nacional não pode estar longe.

REAÇÃO DOS SINDICATOS DOS PROFESSORES

A educação pública americana está bastante sindicalizada, com exceção de alguns Estados do sul e das áreas rurais. Embora os professores tenham conseguido significativos ganhos por meio dos sindicatos, os sindicatos dos professores atualmente não são muito populares entre o público americano (em especial, em razão de sua resistência à reforma proposta). Dois grandes sindicatos representam praticamente todos os professores: a Associação Nacional de Educação (*National Education Association – NEA*) e a Federação Americana de Professores (*American Federation of Teachers – AFT*). A AFT representa os professores de cidades grandes (por ex., Nova York, Chicago, Washington, D.C., Detroit) e é considerada a mais militante das duas.

Em janeiro deste ano, o presidente nacional da AFT anunciou o apoio do sindicato à forma como está se propondo que os professores sejam avaliados, incluindo a análise de “valor agregado” do progresso acadêmico de seus alunos. Isso causou um alvoroço em toda a comunidade acadêmica, mas a NEA manteve-se “estranhamente silenciosa”. Vários Estados – além dos que disputam o financiamento do *Race to the Top* – adotaram uma legislação que exige que a avaliação dos professores seja vinculada, em pelo menos 50%, aos dados dos testes dos alunos (por ex., Colorado, Luisiana, Kentucky). Vários outros Estados recusaram-se a participar do *Race to the Top*, em razão desse requisito (por ex., Flórida, Minnesota). Nas eleições para gover-

nador e para o congresso em novembro de 2010, assim como nas próximas eleições presidenciais em 2012, a questão de vincular a avaliação dos professores aos dados dos testes dos alunos será um fator político. Além disso, fala-se na possibilidade e disposição dos professores de processar os distritos escolares caso sejam demitidos, em virtude dos dados dos testes dos seus alunos.

Várias razões foram apresentadas para questionar o uso de dados dos testes dos alunos para a avaliação dos professores – a falta de um processo adequado, a instabilidade temporal das medidas de valor agregado, o fato de alunos serem designados de forma aleatória a professores ou escolas, e vários outros fatores. Por isso, alguns estudiosos, como Diane Ravitch, levantaram a questão de que professores podem ficar sem emprego, sem nenhuma causa real ou sem terem oferecido um ensino de má qualidade. Ou seja, há muitos fatores que afetam os resultados dos testes dos alunos (ou ganhos nos resultados) além do ensino. Conforme recente relatório da Academia Nacional de Ciências dos Estados Unidos, “... técnicas de ‘valor agregado’ ainda não encontraram uma forma adequada de dar conta desses outros elementos”, por exemplo, a motivação ou o apoio dos pais.

O QUE PODEMOS CONCLUIR?

Meu objetivo ao oferecer informações sobre a recente legislação educacional nos Estados Unidos é indicar até onde chegou a noção do “valor agregado” como abordagem à reforma educacional num período relativamente curto, e quão crucial é a disponibilização de dados longitudinais para essa abordagem. No entanto, alguns precedentes e contextos importantes direcionaram nosso país até aqui num tempo tão curto:

- o aumento da aplicação de testes a todos os alunos nos Estados Unidos, com base na legislação já existente;
- um movimento em direção a padrões nacionais sobre o que as crianças deveriam aprender e quando deveriam aprender. Até o momento, esses padrões só haviam sido desenvolvidos no nível local (municipal);
- ampla insatisfação pública em relação à educação pública americana, desencadeada, pelo menos em parte, por nosso desempenho medíocre em comparações internacionais do TIMSS e do PISA;
- uma história de 30 anos de coleta de dados longitudinais de grandes amostras e nacionalmente representativas de alunos, professores e escolas, pelo Departamento de Educação dos Estados Unidos, por meio do

Centro Nacional de Estatísticas Educacionais. Esses dados estão disponíveis gratuitamente a pesquisadores;

- um novo Presidente com uma promessa declarada de melhorar a educação;
- um novo Secretário de Educação com sólida formação educacional, da mesma cidade (Chicago) que o nosso Presidente, amigo pessoal, com excelente conhecimento sobre o que é um sistema de dados de alta qualidade e o que ele pode oferecer;
- a aceitação, pela comunidade acadêmica, de que “crescimento do desempenho” é uma medida bem mais justa do progresso educacional do que “*status* do desempenho” (uma medida pontual), uma atitude que tem crescido com a disseminação de publicações, usando dados longitudinais nacionalmente representativos;
- um bom grau de insatisfação entre o público em geral em relação aos sindicatos dos professores; e
- um substancial estímulo econômico – com o objetivo de tirar nosso país de uma grave recessão econômica – que funciona como motivação para mudanças.

Muitas dessas questões são peculiares aos Estados Unidos e se combinaram para oferecer um contexto que incentiva uma reforma. Mas algumas não são. Entretanto, está bastante claro que em nosso país a ideia de educação com “valor agregado” já se tornou parte da conversa nacional sobre educação, reforma educacional e avaliação docente. A disponibilização de dados longitudinais é absolutamente essencial para essa reforma específica. Além disso, as ideias em relação à forma como as escolas devem mudar – e isso envolve treinamento, contratação e manutenção de professores de alta qualidade – parece finalmente incluir um sólido reconhecimento de que o aprendizado do aluno é a razão de ser das escolas e do ensino. Pode-se refletir sobre como essas ideias podem ser utilizadas no Brasil, uma vez que as noções de dados longitudinais e de valor agregado começam a ser valorizadas.