Estudos longitudinais e pesquisa na educação básica

Alicia Maria Catalano de Bonamino
Lúcia Helena Gazólis de Oliveira
Pontifícia Universidade Católica do Rio de Janeiro

Resumo

Este artigo discute a adequação e o potencial dos dados longitudinais para o desenvolvimento de pesquisas educacionais. Apresenta as principais características tanto da avaliação nacional em larga escala como de um estudo longitudinal, e analisa as duas metodologias de coleta de dados envolvidas nestas iniciativas. Partindo da constatação de que começam a se formar, no Brasil, bancos de dados coletados longitudinalmente, o artigo apresenta e comenta duas pesquisas que utilizam dados desta natureza, com a finalidade de elucidar sua importância, especialmente na geração de evidências para as políticas de melhoria da qualidade na educação básica.

Palavras-chave: Avaliação educacional. Estudos seccionais. Estudos longitudinais.

Longitudinal studies and research in basic education

This article discusses the suitability and potential of longitudinal data for the development of educational research. It presents the main characteristics of large-scale national evaluations and of longitudinal studies, and examines two data collection methodologies involved in these initiatives. Starting from the observation that, in Brazil, databases collected longitudinally are starting to be formed, the article introduces and comments on two surveys that use such data for the purpose of elucidating their importance, especially as empirical data for the development of policies to improve the quality of basic education.

Keywords: *Educational evaluation. Cross-sectional studies. Longitudinal studies.*

Estudios longitudinales y la investigación en la educación básica

Este artículo aborda la adecuación y el potencial de datos longitudinales para el desarrollo de la investigación educativa. Presenta las principales características de la evaluación nacional brasileña y un estudio longitudinal, y examina las dos metodologías de recopilación de datos involucrados en estas iniciativas. A partir de la observación de que se empiezan a formar, en Brasil, bases de datos recogidos longitudinalmente, el artículo presenta y comenta sobre dos investigaciones que utilizan dichos datos con el fin de esclarecer su importancia, especialmente en la generación de evidencias para las políticas de mejora de la calidad en la educación básica.

Palabras clave: *Evaluación educativa. Estudios seccionales. Estudios longitudinales.*

Introdução

Superado o desafio da universalização do acesso ao ensino fundamental, o Brasil enfrenta hoje a tarefa de criar condições para a permanência e a progressão adequada dos alunos no sistema escolar, o que só é possível por meio de um ensino de qualidade. Tão importante quanto o diagnóstico produzido pelas diversas avaliações nacionais e estaduais é a sua utilização, seja para traçar estratégias de ensino que promovam a aprendizagem de todos os alunos nas unidades escolares, seja para obter evidências que informem os formuladores de política educacional sobre fatores que influenciam o aprendizado dos alunos e que devem, portanto, ser priorizados nos investimentos relativos aos sistemas de ensino. Em resposta às demandas por uma educação de qualidade, observa-se que começa a se estabelecer um núcleo de investigações e de publicações baseadas em dados brasileiros – e, vale dizer, uma tendência para a valorização de políticas fundamentadas em evidências. (Franco, 2003; Alves, 2003; Soares, 2005; Alves, 2006; Alves; Soares, 2008)

Sobre os avanços metodológicos que vêm ocorrendo em relação aos estudos em avaliação escolar, Franco (2003) afirma que a década de 90 foi marcante no sentido da produção de alguns critérios consensuais sobre eficácia escolar. O autor ressalta o crescente entendimento de que a adequada investigação da eficácia escolar não poderia prescindir de estudos longitudinais, da abordagem de valor agregado e da modelagem multinível, ou seja, dos modos adequados de levar em consideração as características dos alunos e das escolas na modelagem estatística dos dados.

Nessa perspectiva, o presente artigo tem o objetivo de discutir sobre a necessidade do uso de dados longitudinais e de modelos multiníveis como requisitos metodológicos em pesquisas com dados de *surveys* educacionais, ilustrando suas potencialidades com o relato de duas investigações que adotam esses procedimentos.

Além desta introdução, o trabalho está organizado em quatro seções. A próxima seção contextualiza a emergência da produção de dados longitudinais no Brasil. Em seguida, são analisadas as especificidades e a adequação dos dados seccionais ou longitudinais, dependendo da investigação pretendida pelo pesquisador. Finalmente, são apresentados estudos para cujos objetivos os dados longitudinais e a modelagem multinível são imprescindíveis. Uma seção com as conclusões encerra o artigo.

Avaliação de sistemas de ensino no Brasil e seus desdobramentos

Em 25 anos de existência, a avaliação da educação em larga escala desenvolveu-se, no Brasil, de forma a abranger a grande maioria das escolas públicas de ensino fundamental do país e as escolas de ensino fundamental e médio de mais de dois terços dos estados. Nesse desenvolvimento, a Prova Brasil e as avaliações estaduais consolidaram, junto com o Sistema Nacional de Avaliação da Educação Básica (Saeb), uma abrangente política de avaliação da educação básica.

As avaliações nacionais, como o Saeb e a Prova Brasil, são coordenadas pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep).

O Saeb é o principal sistema de avaliação diagnóstica da qualidade da educação brasileira. Desde 1995, vem sendo aplicado a cada dois anos em uma amostra representativa dos alunos regularmente matriculados na 4ª série / 5º ano e na 8ª série / 9º ano do ensino fundamental e no 3º ano do ensino médio, de escolas públicas e privadas, localizadas em área urbana ou rural, produzindo estimativas para cada um dos 26 estados e para o Distrito Federal.

Os testes de desempenho, geralmente de língua portuguesa e de matemática, são conjugados com questionários sobre fatores escolares, com foco na escola, no professor e no diretor. Até 2011, foram realizados onze ciclos de avaliação, que vêm contribuindo para a produção de um diagnóstico e para o monitoramento da qualidade da educação básica. Em 1995, o Saeb introduziu várias mudanças metodológicas que possibilitam a comparabilidade do desempenho dos alunos entre anos e séries. Estudo encomendado pelo Inep apontou, por exemplo, questões em relação às séries avaliadas. Entre os sistemas estaduais, por exemplo, havia muitas diferenças com relação aos conteúdos dos currículos da 1ª, 3ª, 5ª e 7ª séries, que eram avaliados até 1995 pelo Saeb, e essas diferenças diminuíam significativamente quando se consideravam os currículos das 4ª e 8ª séries. A partir da terceira edição, o Saeb passou a avaliar essas séries do ensino fundamental e incorporou os alunos da 3ª série do ensino médio. A inovação mais importante, no entanto, foi a adoção da Teoria da Resposta ao Item, que permite pôr os itens dos testes e os alunos em uma mesma escala, o que possibilita a comparação dos resultados dos alunos entre as diferentes séries e áreas de conhecimento, em diferentes edições do Saeb. Os dados dessas avaliações são comparáveis ao longo do tempo, podendo-se, assim, acompanhar a evolução dos desempenhos das escolas, das redes e do sistema como um todo. No ciclo de 1997, o Inep elaborou as Matrizes Curriculares de Referência, que serviriam como base para a confecção dos itens dos testes. A partir dessa edição foi construída uma escala de proficiência única para cada disciplina avaliada (língua portuguesa e matemática) que possibilita a comparação entre os resultados obtidos no ciclo anterior, em 1995, e em todos os ciclos subsequentes.

A Prova Brasil foi criada pelo Inep em 2005. A sua criação é justificada em razão das limitações do plano amostral do Saeb para retratar as especificidades de

municípios e de escolas e para auxiliar os responsáveis pela política educacional no estabelecimento de metas e de ações pedagógicas e administrativas visando à melhoria da qualidade do ensino. A Prova Brasil é aplicada a cada dois anos e produz informações sobre o ensino oferecido pelos municípios e pelas escolas públicas de ensino fundamental. Nas três primeiras edições, em 2005, 2007 e 2009, foram introduzidas importantes modificações. Em 2005, esta avaliação foi censitária para as escolas urbanas. Em 2007, foi alterado o número mínimo de alunos na série avaliada, de 30 para 20, de modo a possibilitar que aproximadamente 400 municípios não incluídos na primeira edição pudessem participar da avaliação. Na terceira edição, em 2009, foram incluídas as escolas rurais que tivessem, no mínimo, 20 alunos nas séries avaliadas. A Prova Brasil avalia as mesmas áreas de conhecimento, os mesmos anos escolares do ensino fundamental e produz resultados na mesma escala. A principal diferença entre essas duas avaliações nacionais é a condição censitária da Prova Brasil.

A partir de 2007 os resultados da Prova Brasil passaram a integrar o Índice de Desenvolvimento da Educação Básica (Ideb), referência para a definição de metas bianuais com valores estimados a serem alcançados, gradualmente, pelas escolas e sistemas de ensino até 2021.

Por sua vez, as avaliações estaduais têm como característica comum a tendência a adotar um modelo censitário. Afora isso, a maioria delas se concentra em determinados anos escolares, tipicamente 5º e 9º anos do ensino fundamental e 3º ano do ensino médio, sendo os alunos diferentes a cada aplicação, como ocorre também nas avaliações nacionais (Saeb e Prova Brasil). A partir de 2000, a maioria das avaliações estaduais fez uma nova aproximação com as avaliações nacionais da educação básica, ao incorporar a utilização da escala de proficiência do Saeb.

Paralelamente ao desenvolvimento das grandes avaliações houve, a partir dos anos 90, um aumento considerável das pesquisas sobre a importância das escolas para o desempenho dos alunos (Gatti, 1994; Bomeny, 1998). Isto foi estimulado, por um lado, pelas avaliações nacionais e estaduais que publicaram seus resultados e disponibilizaram seus dados, e, por outro lado, pela demanda dos formuladores e executores da política educacional por diagnósticos e subsídios que embasassem suas decisões. Como consequência, além das mudanças na abrangência espacial, as avaliações em larga escala da educação básica vêm passando também por mudanças nos desenhos metodológicos que abrem a possibilidade de interações significativamente novas com as políticas e a pesquisa educacional.

Especificidades da avaliação com desenho seccional

Atualmente, as grandes avaliações educacionais, como o Saeb, a Prova Brasil e a maioria das avaliações estaduais, adotam desenhos seccionais, ou seja, o mesmo teste é aplicado a grupos de alunos da série avaliada e estes alunos são diferentes a cada ano de avaliação. No mesmo período de avaliação são, ainda, aplicados questionários contextuais sobre a escola, os diretores e os professores. Tipicamente, a captação das informações sobre todas as variáveis (cognitivas e contextuais, do aluno e da escola) é simultânea, ou seja, os dados contextuais também são obtidos de profissionais diferentes a cada ano, e esta característica demarca uma das suas principais limitações.

Essa limitação se traduz na falta de sintonia na relação temporal entre o aprendizado do aluno e as medidas contextuais da escola, o que inviabiliza alguns dos objetivos das avaliações educacionais em vigor, como a oferta de explicações confiáveis sobre o efeito dos fatores escolares no aprendizado dos alunos. Isto acontece porque os instrumentos cognitivos tipicamente testam cada aluno apenas uma vez. Como a medida de desempenho é o resultado do aprendizado dos alunos ao longo de muitos anos, essa medida não pode ser explicada em função do passado recente. Ou seja, o desempenho do aluno no 5ª ano do ensino fundamental retrata não apenas o que ele aprendeu no 5º ano, mas também no 4º, no 3º, no 2º e no 1º ano. Como as variáveis do professor e da escola são observadas apenas no 5º ano, há falta de sintonia temporal entre a medida de desempenho e as medidas de contexto aferidas pelas avaliações seccionais.

Mais especificamente, enquanto a medida de desempenho em língua portuguesa e em matemática é um agregado do aprendizado dos alunos ao longo de muitos anos, as medidas escolares referem-se às condições da escola no ano da coleta de dados e isto inviabiliza o cálculo do ganho atribuível aos fatores escolares e, portanto, a realização de inferências confiáveis sobre a eficácia ou o efeito das escolas. (Franco; Brooke; Alves, 2008)

Segundo Soares (2005) e Alves (2006), a inexistência ou precariedade de medidas longitudinais levou à utilização de dados seccionais, ainda que este tipo de dado não constitua uma base adequada para a investigação da eficácia escolar e do efeito escola. Observa-se, ainda, que os campos de investigação sobre efeito-escola e sobre escola eficaz às vezes se confundem, fazendo com que nem sempre seja possível classificar as pesquisas e os pesquisadores claramente em um ou outro campo.

Teddlie e Reynolds (2000, apud Alves, 2006, p. 12-13) sinalizam pontos de distinção entre esses dois campos. Para os autores, o objetivo das pesquisas sobre o efeito-escola é analisar o impacto das escolas, através de modelos estatísticos, o que, na primeira fase do desenvolvimento da pesquisa, correspondeu à utilização dos modelos *input-output* e atualmente, corresponde ao uso de modelos multiníveis. Por sua vez, as pesquisas sobre a escola eficaz analisam os processos

organizacionais e pedagógicos de escolas consideradas "atípicas" porque são frequentadas por alunos desfavorecidos economicamente, mas apresentam bons resultados. Esses estudos utilizam tanto de metodologias quantitativas, tais como modelos de correlação, de regressão e multiníveis, quanto de metodologias qualitativas, como estudo de caso e etnografia, que envolvem a análise simultânea das escolas e das salas de aula.

Voltando ao tema das grandes avaliações nacionais e estaduais, deve-se ressaltar que as mesmas têm sido extremamente úteis, porque criaram condições para pesquisas sobre o desempenho dos alunos. Tais pesquisas, vale dizer, demandam grandes investimentos financeiros e, dessa forma, os pesquisadores precisam contar com iniciativas governamentais para conduzirem seus estudos. Nessa conjuntura, as avaliações nacionais e estaduais têm coletado medidas cognitivas e contextuais que concretizam um campo fértil para a pesquisa educacional e vêm fomentando o debate sobre características de boas escolas. No entanto, esses dados apresentam limites consideráveis, conforme mencionado anteriormente, dependendo dos objetivos das pesquisas. Em resumo, a coleta de dados seccionais mede o desempenho dos alunos em determinado momento, mas não é apropriada para medir o seu aprendizado juntamente às condições específicas em que esse aprendizado ocorreu, uma vez que não permite medir o valor agregado pelo estabelecimento de ensino, aos mesmos alunos, conforme ocorre em avaliações longitudinais. Veremos, a seguir, como os estudos longitudinais contribuem para a solução desta questão.

Especificidades da avaliação com desenho longitudinal

O entendimento dessa limitação e o desenvolvimento de novas metodologias de análise para dados educacionais vêm levando pesquisadores e gestores à adoção de desenhos longitudinais de coleta de dados. Nesse desenho, o mesmo aluno é testado mais de uma vez, com o objetivo de obter uma dupla medida de proficiência que permita filtrar o aprendizado do aluno em uma escola num determinado período.

Desenhos longitudinais têm como pressuposto a distinção entre desempenho escolar e aprendizado. Desempenho diz respeito a uma medida pontual, a uma medida estática, ao passo que o aprendizado diz respeito ao progresso do aluno, ou seja, à aquisição de conhecimentos e de habilidades e às mudanças que ocorrem na sua trajetória escolar. (Alves; Soares, 2008)

Os desenhos longitudinais fornecem maior precisão nas estimativas de mudanças temporais do que os estudos seccionais, pois focam a mudança do desempenho dos alunos de um ano para outro e permitem calcular o crescimento do aprendizado do aluno dentro de determinada escola. Este tipo de desenho

vem sendo considerado a melhor estratégia de observação do processo de aprendizagem. Alguns autores usam o termo "valor agregado" para designar esse aprendizado. (Lee, 2010)

Os dados longitudinais permitem, assim, analisar o quanto as práticas escolares contribuem para o aprendizado dos alunos durante o ano em que eles frequentaram uma determinada escola e, por essa razão, são considerados um elemento chave nas pesquisas sobre o efeito escola e também sobre a eficácia escolar.

Atualmente, cinco estados conduzem avaliações com desenho longitudinal em seus sistemas de ensino. Ceará, Pernambuco e Rio de Janeiro coletam dados longitudinais com alunos do 1º, 2º e 3º ano do ensino médio. No outro extremo da educação básica, os Estados de Minas Gerais e Espírito Santo coletam dados longitudinais no período da alfabetização e da aquisição de habilidades básicas de leitura e escrita. O 1º, 2º e 3º ano do ensino fundamental são avaliados no Espírito Santo e 2º, 3º, 4º e 5º anos em Minas Gerais. Essas iniciativas visam apoiar políticas de responsabilização com base no "valor agregado", medida que é considerada mais adequada em relação ao aprendizado do aluno, durante o tempo em que ele frequentou determinada escola ou sala de aula, do que a medida pontual de desempenho. O valor agregado é assim considerado uma medida mais justa do efeito da escola e do professor.

A pesquisa Geres

As grandes avaliações nacionais e estaduais foram relevantes tanto para o diagnóstico da qualidade do ensino e das disparidades sociais e regionais de desempenho escolar como para o desenvolvimento das pesquisas sobre a eficácia da escola. A partir delas, puderam ser percebidas com clareza tanto as potencialidades como as limitações que o desenho metodológico desses *surveys* coloca para a investigação dos fatores escolares, impulsionando a adoção de novas metodologias.

O reconhecimento das limitações da coleta seccional de dados para a investigação do efeito de escolas levou um grupo de pesquisadores de seis universidades (UFMG, Unicamp, UFJF, UFBA, UEMS e PUC-Rio)[1] a desenvolver o projeto Geres — Estudo Longitudinal da Geração Escolar 2005.

Com o objetivo de investigar quais práticas pedagógicas e condições escolares contribuem para a promoção da eficácia e da equidade escolar, o Geres acompanhou a mesma amostra de alunos a partir do 2º ano de escolarização, ao longo de quatro anos, entre 2005 e 2008. Foram aplicados testes cognitivos

1. Universidade Federal de Minas Gerais (UFMG); Universidade Estadual de Campinas (Unicamp); Universidade Federal de Juiz de Fora (UFJF); Universidade Federal da Bahia (UFBA); Universidade Estadual de Mato Grosso do Sul (UEMS), e Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

bem como questionários contextuais para alunos, professores e diretores das escolas participantes em cinco cidades: Belo Horizonte (MG), Rio de Janeiro (RJ), Salvador (BA), Campinas (SP) e Campo Grande (MS). A amostra total do Geres está composta de 303 escolas divididas nessas cinco grandes cidades brasileiras. Além de representarem centros urbanos importantes na maioria das regiões, as 303 escolas representam também as três redes responsáveis pelo ensino fundamental no Brasil: estadual, municipal e privada.

O desenho longitudinal da pesquisa Geres atende ao propósito de se investigar as alterações no desempenho dos alunos acontecidas entre um momento e outro da escolarização, ou seja, o "valor agregado" pelo estabelecimento escolar e, ainda, de se identificar as características das escolas e salas de aulas promotoras desse progresso. Essa decisão metodológica é muito importante, uma vez que permite controlar os resultados dos estudantes ao final de um determinado período pelos seus próprios resultados em teste prévio e aferir o valor agregado nesse período determinado, sob condições específicas de ensino. (Franco; Brooke; Alves, 2008)

Os questionários contextuais, além de utilizar informações relevantes já disponíveis na escola, foram aplicados aos pais dos alunos, aos alunos, aos professores e aos diretores. O questionário dos professores buscou identificar diferentes perfis profissionais e delinear sua prática em sala de aula, enfocando características sobre exigências acadêmicas, modos de organização da sala de aula, uso do tempo, estilos e recursos pedagógicos e práticas de avaliação. O questionário respondido pelos diretores tratou de questões sobre o entorno escolar, a relação com a comunidade, o perfil de gestão e a infraestrutura da escola. Já os alunos e seus responsáveis responderam, por exemplo, itens sobre os recursos econômicos e sobre nível de escolaridade e composição familiar.

Para coletar os dados de caráter cognitivo, o Geres utilizou testes de leitura e de matemática, focalizando habilidades básicas tipicamente demandadas pela escola a alunos das séries iniciais, tendo como orientação uma matriz de referência. Tanto a matriz quanto os testes foram elaborados por especialistas dessas duas áreas de conhecimento escolar. (Brooke; Bonamino, 2011)

A primeira aplicação do Geres ocorreu em março de 2005, com a observação de alunos do 2º ano do ensino fundamental, a segunda, a terceira, a quarta e a quinta ondas de medida ocorreram, respectivamente, em novembro de 2005, 2006, 2007 e 2008, completando o acompanhamento da amostra ao longo dos quatro anos escolares das séries iniciais.

O Geres cuidou da distinção entre 'valor agregado pela escola' e o efeito da 'seleção', o que fez com que alunos retidos continuassem a ser observados, desde que permanecessem na escola ou se transferissem para uma escola da amostra.

Além da escola, nos anos iniciais do ensino fundamental, a sala de aula é particularmente importante. Em razão disso, foi observado o maior número possível de salas de aula dentro de cada escola. Com esse objetivo, o recurso foi escolher um recorte concentrado em cinco municípios. Cada um deles foi

considerado como um estrato e, dentro de cada município, foi selecionada uma amostra probabilística complexa de escolas, turmas e alunos a partir do cadastro do Censo Escolar de 2003, excluídas as escolas que não possuíam 10 alunos ou mais matriculados no 2º ano do ensino fundamental.

Para maximizar a diversidade de escolas selecionadas, foi utilizado um critério adicional de estratificação, variável conforme a disponibilidade de informações em cada estrato. Idealmente, a variável de estratificação foi a medida socioeconômica média das escolas. Nos estratos nos quais esta variável não estava disponível utilizou-se como variável de estratificação o desempenho médio das escolas em avaliação censitária ou, na falta desta, medida de equipamento das escolas, construída a partir de dados do censo escolar.

Selecionadas as escolas, todas as turmas do 2º ano e seus alunos passaram a fazer parte da amostra. Este desenho visou ao aumento da amostra dentro de cada escola e de cada turma, para permitir maiores opções em relação ao estudo do *efeito escola* e do *efeito turma*. Em todas as aplicações, a pesquisa coletou medidas sobre a escola e sobre as salas de aula.

Foram desenvolvidas duas escalas, uma de leitura e outra de matemática, para descrever o desempenho dos alunos em termos do desenvolvimento das habilidades, da seguinte forma: (i) os alunos encontram-se aquém das habilidades do nível, (ii) os alunos encontram-se no auge das habilidades do nível, (iii) os alunos encontram-se em vias de consolidação das habilidades do nível e (iv) as habilidades foram consolidadas pelos alunos.

Além do desenho longitudinal, outro avanço metodológico adequado à investigação da eficácia escolar que passou a ser adotado pelo Geres diz respeito à abordagem analítica multinível. Em geral, nas análises de um nível só, obtidas por modelos de regressão linear ordinária, os efeitos dos fatores contextuais são sistematicamente subestimados. A utilização dos modelos multiníveis permite medir a variabilidade nos resultados dos alunos, que pode ser explicada por diferentes níveis de análise como, por exemplo, a sala de aula, a escola ou sistema de ensino. Goldstein (2001, p. 96) afirma que a partir dessa metodologia começamos a ver as ações dos individuais como mediadas pelas ações dos outros com os quais eles estão em contato e pelas instituições às quais eles pertencem. Isso não é possível com os modelos de análise tradicionais, porque eles tratam os indivíduos como entidades que atuam de forma independente. A possibilidade de utilização dos modelos hierárquicos solucionou os problemas teóricos e metodológicos que constituíam os principais pontos de crítica em relação aos estudos educacionais produzidos nas décadas anteriores (Alves; Soares, 2007). Na pesquisa sobre o efeito das escolas ou da eficácia escolar, os modelos multiníveis colaboram também com as pesquisas qualitativas, uma vez que propiciam a identificação de escolas que apresentam resultados acima do esperado, depois de controladas as condições sociais dos alunos.

Estudos longitudinais e possibilidades de pesquisas educacionais

Com o objetivo de fomentar a interlocução entre as avaliações longitudinais em curso e as pesquisas em educação, apresentam-se a seguir dois exemplos de investigações realizadas nos últimos anos que têm em conta o valor agregado pelas escolas aos seus alunos.

Dentre as possibilidades de pesquisa com objetivos que demandam dados longitudinais, estão as que se referem ao efeito-professor. Tais pesquisas abordam, por exemplo, a concentração ou ênfase no processo ensino-aprendizagem, o ensino estruturado com propósitos claramente definidos, expectativas elevadas do professor em relação aos alunos e o monitoramento do processo de aprendizagem.

Vários estudos mostram a existência de diferenças bastante significativas de progresso dos alunos em função da sala de aula em que são escolarizados. Cunha (2003, apud Hamada, 2008, p. 51) afirma que a importância do estudo envolvendo o professor enquanto ator significativo no processo de ensino deve-se ao fato de que ele é o agente principal das decisões na sala de aula, quem concretiza a definição didático-pedagógica e aplica dentro de sala de aula um dado conteúdo. Sendo assim, compreender as práticas que diferenciam os professores em termos do quanto de aprendizado conseguem agregar em relação às turmas nas quais lecionam, adicionando evidências que possam reforçar ou contrariar indicações de cunho apenas opinativo sobre o que é importante de ser realizado nas salas de aula, é imprescindível para se construir intervenções e propostas educacionais efetivas.

O primeiro estudo que será comentado foi desenhado especificamente com o propósito acima mencionado, ou seja, investigar questões relativas ao efeito professor no aprendizado dos alunos. O estudo "Habilidades de leitura e práticas pedagógicas associadas ao seu aprendizado" (Oliveira, 2012), utiliza dados cognitivos e contextuais da pesquisa Geres.

Foram investigados 17.628 alunos, de 671 turmas, de 225 escolas, pertencentes às redes municipal, estadual, federal e privada, das cidades do Rio de Janeiro (RJ), de Belo Horizonte (MG), de Campinas (SP) e de Campo Grande (MS) e seus respectivos professores. Os estudantes, participantes do Geres, frequentavam o 2º e o 3º anos do ensino fundamental.

O objetivo da pesquisa de Oliveira (2012) foi averiguar quais práticas pedagógicas estão associadas ao aumento do aprendizado das diferentes habilidades que necessitam ser construídas durante o processo de aprendizagem da língua portuguesa e, ainda, se os impactos dessas práticas se diferenciam conforme o ano escolar frequentado pelos alunos.

Esse é o caso de questões de pesquisa que apenas podem ser respondidas a partir da disponibilidade de avaliações consecutivas realizadas com os mesmos alunos, primeiramente avaliando-se o nível de proficiência com que eles iniciaram

a pesquisa (proficiência prévia) e, posteriormente, verificando-se o conhecimento que agregaram ao final do 2º e do 3º anos de escolaridade. Dessa forma, tendo-se a proficiência prévia como controle, o efeito das variáveis estudadas se dá sobre o aprendizado e não só sobre o desempenho em um dado momento. É importante reforçar esse ponto já que o desempenho tem relação com muitos fatores intrínsecos aos alunos, mas a proficiência medida por avaliações subsequentes dos mesmos alunos permite aferir o valor agregado nesse período e associá-la às condições específicas em que o ensino se deu.

Outra metodologia importante, conforme mencionado anteriormente, possibilitada pelo dado longitudinal, foi a adoção da modelagem multinível, que considera a natureza hierárquica dos dados educacionais. Essa característica precisa ser levada em conta, por exemplo, quando se observam o desempenho, a atitude frente aos estudos ou outra variável individual no nível dos alunos, os quais se agrupam em turmas, que se agrupam em escolas, que, por sua vez, podem se reunir em níveis ainda maiores de agregação, como municípios, estados e países.

Essa consideração é imprescindível porque a distribuição dos alunos por turmas e por escolas não ocorre de forma aleatória. Existem características próprias desses contextos (das escolas e das turmas, e não só dos alunos) que influenciam na forma como os alunos aprendem. O fato de alunos pertencerem a uma determinada turma ou a uma determinada escola faz com que eles tendam a se assemelhar mais entre si do que com alunos de outras turmas ou de outras escolas. Tais semelhanças costumam ocorrer no próprio nível dos alunos: é comum que colegas de uma mesma escola, por exemplo, provenham de famílias com um nível socioeconômico e cultural semelhante. Entretanto, as semelhanças obviamente se tornam ainda mais fortes quando se consideram níveis maiores de agregação, como turmas e escolas. No caso das turmas, os alunos compartilham entre si o fato de estarem expostos ao ensino ministrado pelo mesmo professor, que tem, por sua vez, características didáticas e de formação específicas, entre outras. Algo semelhante se passa no nível das escolas, nas quais os alunos de um mesmo estabelecimento educacional normalmente estão submetidos a uma mesma direção escolar, com suas dinâmicas próprias e as mesmas condições de infraestrutura. Diante dessa característica, quando se realiza a regressão da proficiência dos alunos em relação a uma ou mais variáveis explicativas, é natural que os resíduos da regressão dependam fortemente da turma ou da escola especificamente considerada e que tais resíduos se distribuam em torno das retas de regressão com diferentes valores de variância, conforme a turma ou escola analisada.

Machado Soares (2003) afirma que a estruturação hierárquica constitui uma sequência natural de agrupamentos aninhados, de tal forma, que as variáveis representativas das características nos diversos níveis podem interagir com outras variáveis dentro do mesmo nível hierárquico e, também, com variáveis de outro nível. Tendo em vista, principalmente, essa natureza dos dados, é frequente

a utilização de modelos de regressão hierárquicos, que permitem investigar a influência das características de cada nível da hierarquia no desempenho escolar dos alunos e na diferenciação entre as escolas e, ainda, separar a variabilidade nos resultados – associada às escolas – da variabilidade dentro de cada escola – associada aos alunos ou turmas de alunos.

As técnicas de análise multinível trabalham com uma ideia relativamente simples e engenhosa para diminuir os obstáculos característicos dos dados hierarquizados: a determinação simultânea de um grande número de modelos lineares, um para cada unidade de interesse. Desta forma, se o interesse do pesquisador for analisar escolas diferentes ou turmas diferentes estabelece-se uma equação de regressão para cada escola ou para cada turma, conforme o caso.

Retomando-se as questões em estudo na pesquisa apresentada, que busca a existência de associações significativas entre a proficiência dos alunos e determinadas características didáticas dos seus respectivos professores, como os professores, mesmo dentro de uma mesma escola, geralmente variam entre as diferentes turmas, as variáveis que expressam o seu fazer pedagógico influenciam, prioritariamente, o nível da turma. Além disso, variáveis identificadas como importantes para a explicação da proficiência estão associadas aos alunos. Desta forma, os modelos desenhados para a investigação contaram com a modelagem hierárquica em dois níveis: o nível turma e o nível aluno. No nível do aluno foram utilizadas variáveis representativas do nível socioeconômico, da proficiência anterior, do gênero e uma variável adicional da rede (pública ou particular) frequentada pelo aluno. No nível da turma, foram utilizadas as variáveis representativas das práticas pedagógicas utilizadas pelos professores junto aos seus alunos.

Resumidamente, a associação entre as práticas de leitura e a proficiência, realizada pela autora, revelou ganhos diferenciados por parte dos alunos, conforme a habilidade, o ano escolar e as práticas adotadas pelos professores. Os principais resultados mostraram que (i) a ênfase em métodos de alfabetização que abordam as sílabas de forma descontextualiza teve correlação negativa com o aprendizado das habilidades de processamento do código de leitura e de localização de informações no 2º ano; (ii) no 3º ano, a prioridade em atividades de cópia, ditado e caligrafia apresentou correlação negativa com o aprendizado dos alunos em todas as habilidades investigadas (iii) e a leitura realizada pelo professor para os seus alunos, no 3° ano, foi a única modalidade de prática de leitura que se manteve significativa e com potencial para agregar conhecimento aos estudantes, também em todas as habilidades pesquisadas.

O segundo estudo a ser comentado ilustra a importância de se contar com dados longitudinais na mensuração do valor agregado pela escola, bem como uma das possibilidades de colaboração entre abordagens quantitativas e qualitativas de pesquisa na identificação de escolas que promovem resultados acima do

esperado.

Na pesquisa "A relação entre o projeto pedagógico e a aprendizagem dos alunos em escolas participantes do Projeto Geres em Belo Horizonte", Saraiva (2009) investigou a relação entre o Projeto Político Pedagógico (PPP) da escola e o nível de crescimento da aprendizagem dos seus alunos.

A pesquisa foi desenvolvida em quatro escolas estaduais de Belo Horizonte, que foram selecionadas em razão dos seus resultados no Geres. A partir da definição das escolas da amostra, utilizando-se o critério de valor agregado, a pesquisa analisou o PPP de cada escola, em relação à elaboração, ao conteúdo e à adequação aos pressupostos teóricos e às diretrizes oficiais, além da capacidade de nortear as ações pedagógicas em sala de aula. Na sequência, foi aplicado um questionário contextual a uma amostra de professoras dos anos iniciais para levantar informações pertinentes à sua participação na elaboração do PPP da escola, seu conhecimento desse projeto e a utilização que faz dele na sala de aula. Essas informações foram relacionadas com o valor agregado da escola, com a finalidade de verificar se o PPP poderia auxiliar na explicação das diferenças observadas entre as escolas e o progresso no nível de aprendizagem dos seus alunos. (Saraiva, 2009, p.6)

Assim como no primeiro estudo apresentado, os dados longitudinais permitiram que Saraiva (2009) utilizasse o valor agregado pelos estabelecimentos escolares e que adotasse os modelos de regressão linear multinível, nesse caso, os níveis aluno e escola. A autora utilizou três variáveis para o cálculo do valor agregado: Proficiência em matemática e língua portuguesa nas ondas 1 e 3, no 2º e no 3º anos do ensino fundamental; Nível Socioeconômico do Aluno (NSE) e Nível Socioeconômico da Escola. Eentre as vinte escolas da rede estadual de Belo Horizonte que participam do projeto Geres e que contavam com informação sobre o valor agregado em matemática e língua portuguesa e sobre o valor referente ao nível socioeconômico médio da escola (NSE), a maioria apresentou valores de NSE negativos. Isso significa que estas escolas se encontram abaixo da média em relação ao conjunto de escolas participantes da pesquisa Geres, que inclui um número equivalente de instituições particulares. Por outro lado, as escolas apresentam valores positivos de valor agregado, revelando que a maioria delas consegue contribuir para a aprendizagem dos seus alunos. No entanto, a existência de outras escolas com valor agregado muito baixo e até negativo, demonstra que existem também instituições escolares nas quais os alunos aprendem pouco ou nada além daquilo que seria esperado deles dada sua condição socioeconômica. Nestes casos, segundo saraiva, as escolas estão simplesmente confirmando a situação socioeconômica do aluno ou até diminuindo as suas chances de mobilidade. (Saraiva, 2009, p. 46)

Como o objetivo da pesquisa era analisar o valor agregado para selecionar escolas que, apesar de atender um público discente muito parecido, fizessem diferença no progresso dos seus alunos em comparação com a média, Saraiva

selecionou escolas contrastantes, duas com valor agregado significativamente acima da média e outras duas com valor agregado significativamente abaixo da média.

A autora procurou escolher escolas nas quais os alunos apresentavam proficiências iniciais parecidas ou pequenas diferenças que não comprometessem a comparação das escolas. Saraiva (2009) mostra que se tivesse sido levado em conta apenas o valor agregado, poderiam ter sido aceitas como iguais escolas que não são comparáveis uma vez que recebem alunos com níveis iniciais de aprendizagem diferentes.

As semelhanças das escolas também foram observadas em relação a outras dimensões, como a estrutura física, os recursos pedagógicos e tecnológicos a disposição das professoras e o número de professores por turma e turno, revelando não existirem grandes diferenças em termos de recursos físicos e humanos nas escolas da rede estadual de Belo Horizonte. Diante dessas semelhanças, a pesquisa buscou confirmar ou rejeitar a hipótese segundo a qual as diferenças no valor agregado das escolas poderiam dever-se ao Projeto Político Pedagógico.

Em linhas gerais, os resultados mostraram uma padronização na estruturação do PPP das escolas e indicativos de que esse documento teria sido elaborado, na maioria dos casos, pelas Diretoras, o que levaria à perda da capacidade esperada de nortear as ações pedagógicas e o trabalho das professoras na sala de aula.

A análise dos quatro PPP e as entrevistas realizadas com as professoras levaram Saraiva (2009) à conclusão de que não é possível estabelecer uma relação entre o PPP da escola e o seu valor agregado, devido ao grau de similaridade entre os projetos. As diferenças no valor agregado parecem residir, segundo Saraiva, no trabalho docente e na ação individual das professoras. Apenas em relação à escola de maior valor agregado, a autora mostra que, apesar de o documento não apresentar diferenças relevantes com os documentos das outras escolas, as professoras relatam uma participação e um conhecimento maior do PPP, o que poderia, segundo Saraiva (2009), significar a existência de outros mecanismos internos de organização e circulação de informações que estariam contribuindo para um melhor funcionamento dessa escola.

Conclusões

Discutimos neste artigo os limites dos dados seccionais e o potencial dos dados longitudinais para a pesquisa educacional e para a produção de evidências sobre a eficácia escolar e sobre o efeito das escolas no aprendizado dos alunos. Consideramos ser esta uma discussão oportuna. Uma vez que um número crescente de estados começa a conduzir avaliações longitudinais no início do ensino fundamental e no ensino médio, a disponibilização de dados coletados

longitudinalmente poderá propiciar o desenvolvimento de estudos sobre valor agregado e de novas pesquisas sobre a importância das escolas para o aprendizado dos alunos.

Discutimos o fato de as avaliações nacionais e estaduais terem um desenho adequado para acompanhar a evolução da qualidade da educação e para o desenvolvimento de pesquisas sobre o desempenho dos alunos e discutimos, também, as limitações desse desenho seccional. Argumentamos que a coleta seccional de dados por essas avaliações limita as possibilidades de pesquisar o aprendizado juntamente com as condições específicas em que o mesmo ocorreu, levando à subestimação do peso dos fatores intraescolares sobre esse aprendizado.

Em nossa argumentação, procuramos colocar em evidência os atuais avanços metodológicos no uso de dados das avaliações educacionais, propiciados pelas análises que utilizam dados longitudinais, a abordagem de valor agregado e a modelagem multinível, discutindo sua adequação para o estudo do efeito da escola no processo temporal de desenvolvimento da aprendizagem do aluno.

Discutimos que, como os estudos com dados longitudinais focam a mudança na aprendizagem dos alunos entre os anos escolares e permitem calcular o valor agregado ao seu aprendizado dentro de uma determinada escola e analisar o quanto as características dessa escola contribuíram para o aprendizado dos alunos durante o tempo em que os alunos a frequentaram, a coleta longitudinal de dados possibilita uma maior precisão nas estimativas de mudanças temporais no aprendizado dos alunos do que os estudos seccionais. Mais especificamente, apontamos que os dados longitudinais, a abordagem de valor agregado e a modelagem multinível são a forma mais adequada de se obter evidências mais consistentes para induzir conclusões sobre os recursos e as práticas escolares que promovem a aprendizagem dos alunos.

No encerramento do artigo, procuramos ilustrar essa adequação por meio do relato de duas pesquisas que aplicaram a abordagem de valor agregado e a modelagem multinível a dados do Geres, o primeiro estudo longitudinal sobre a aprendizagem de leitura e de matemática nos anos iniciais do ensino fundamental a ser concluído e divulgado no Brasil. Esperamos que o debate, aliado aos exemplos de investigação, possam abrir caminhos para novas pesquisas com as características apontadas e que são tão necessárias à construção de uma educação de qualidade.

Referências

ALVES, Maria Teresa. *Efeito-escola e fatores associados ao progresso acadêmico dos alunos entre o início da 5ª série e o fim da 6ª série do ensino fundamental:* um

estudo longitudinal em escolas públicas no município de Belo Horizonte - MG (190 f.). Tese (Doutorado em Educação) - Programa de Pós-Graduação, Universidade Federal de Minas Gerais, 2006.
ALVES, Maria Teresa; SOARES, José Francisco. As pesquisas sobre o efeito das escolas: contribuições sociológicas para a sociologia da educação. Sociedade e Estado, Brasília, v. 22, n. 2, p.435-473, maio/ago, 2007.
_____. O efeito das escolas no aprendizado dos alunos: um estudo com dados longitudinais no Ensino Fundamental. *Educação e Pesquisa*, São Paulo, v.34, n.3, p. 527-544, set.-dez. 2008.
BOMENY, Helena (Org.). *Avaliação e determinação de padrões na educação latino-americana*: realidades e desafios. Rio de Janeiro: Fundação Getúlio Vargas, 1998.
BROOKE, Nigel; BONAMINO, Alicia. (Orgs.). Geres 2005: Razões e resultados de uma pesquisa longitudinal sobre eficácia escolar. Rio de Janeiro : WalPrint, 2011.
FRANCO, Creso. A investigação sobre a eficácia escolar no Brasil. In: TORRECILLA, Javier Murillo. (Coord.). *La investigación sobre eficacia escolar en Iberoamerica.* Revisão internacional sobre el estado del arte. Bogotá: Convenio Andrés Bello/ Centro de Investigación y Documentación Educativa, 2003. p.191-207.
FRANCO, Creso; BROOKE, Nigel; ALVES, Fátima. Estudo longitudinal sobre qualidade e equidade no ensino fundamental brasileiro: GERES 2005. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 16, n. 61, p. 625-638, out.-dez. 2008.
GATTI, Bernadete. Avaliação educacional no Brasil: experiências, problemas, recomendações. *Estudos em Avaliação Educacional*, n. 10, Fundação Getúlio Vargas, jul.-dez., p. 67 – 80, 1994.
GOLDSTEIN, Harvey. Modelos de realidade: novas abordagens para a compreensão de processos educacionais. In: FRANCO, C. (Ed.). *Avaliação, ciclos e promoção na educação*. Porto Alegre: Artmed, 2001. p.84-99.
HAMADA, Hélio Hiroshi. *Ensino profissional da polícia militar em Minas Gerais*: análise do efeito-professor no curso técnico de segurança pública (145 f.). Dissertação (Mestrado em Educação) - Programa de Pós-Graduação, Universidade Federal de Minas Gerais, 2008.
LEE, Valerie. Dados longitudinais em educação: um componente essencial da abordagem de valor agregado no que se refere à avaliação de desempenho escolar. *Estudos em Avaliação Educacional*, São Paulo, v. 21, n. 47, p. 531-542, set.-dez. 2010.
OLIVEIRA, Lúcia Helena Gazóllis. *Habilidades de leitura e práticas pedagógicas associadas ao seu aprendizado*, 2012 (158 f.). Tese (Doutorado em Educação). Programa de Pós-Graduação, Pontifícia Universidade Católica do Rio de Janeiro, 2012.
SARAIVA, Ana Maria. *A relação entre o Projeto Pedagógico e a aprendizagem dos alunos em escolas participantes do projeto Geres em Belo Horizonte*, 2000 (118 f.). Dissertação (Mestrado em Educação). Programa de Pós-Graduação. Universidade

Federal de Minas Gerais, 2009.
SOARES, José Francisco. O efeito da escola no desempenho cognitivo de seus alunos. In: A. M. Souza (Ed.). *Dimensões da avaliação educacional.* Petrópolis, Editora Vozes, 2005, p.174-204.
SOARES, Tufi Machado. Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no Simave-2002. *Estudos em Avaliação Educacional*, São Paulo: Fundação Carlos Chagas, n. 28, jul-dez, 2003, p. 105.

Recebido em setembro de 2012
Aprovado em novembro de 2012

Alicia Maria Catalano de Bonamino é doutora em educação, professora associada da Pontifícia Universidade Católica do Rio de Janeiro e pesquisadora do Laboratório de Avaliação da Educação (LAEd). Email: alicia@puc-rio.br

Lúcia Helena Gazólis de Oliveira é doutora em ciências sociais-educação, professora do Colégio de Aplicação da UFRJ, pesquisadora do Laboratório de Avaliação da Educação da PUC-Rio (LAEd) e na área de leitura, escrita e tecnologias da informação e da comunicação, na UFRJ. Email: lhgazolis@oi.com.br