

# CONSTRUÇÃO DE INDICADORES PARA DESCREVER DESIGUALDADES DE APRENDIZADO NA PROVA BRASIL

**MARIA TERESA GONZAGA ALVES  
FLAVIA PEREIRA XAVIER**

## **RESUMO**

*O artigo apresenta dois indicadores que sintetizam dimensões complexas das escolas: intervenção para melhoria e currículo na escola. Eles foram construídos com os dados do Sistema de Avaliação da Educação Básica por meio da Teoria da Resposta ao Item. Os indicadores foram utilizados para análise do aprendizado dos alunos das escolas públicas de ensino fundamental e para testar hipóteses sobre desigualdades entre alunos com e sem atraso escolar por meio de modelos estatísticos. Escolas com escores mais altos nos dois indicadores aumentam as chances de seus alunos estarem nos níveis mais altos de aprendizado, mas as práticas sintetizadas nos indicadores contribuem para reduzir diferenças entre alunos com e sem atraso escolar apenas em relação às chances de eles estarem no nível básico. Os indicadores não devem ser tomados de forma determinística, mas têm potencial para indicar problemas ou êxitos nas práticas pedagógicas das escolas.*

**PALAVRAS-CHAVE** INDICADORES EDUCACIONAIS • PROVA BRASIL •  
DESIGUALDADES EDUCACIONAIS • APRENDIZAGEM.

# CONSTRUCCIÓN DE INDICADORES PARA DESCRIBIR DESIGUALDADES DE APRENDIZAJE EN LA PROVA BRASIL

## RESUMEN

*El artículo presenta dos indicadores que sintetizan dimensiones complejas de las escuelas: intervención para mejoras y currículo escolar. Fueron elaborados con los datos del Sistema de Evaluación de la Educación Básica por medio de la Teoría de la Respuesta al Ítem. Los indicadores fueron utilizados para analizar el aprendizaje de los alumnos de las escuelas públicas de educación básica y para testear hipótesis sobre desigualdades entre alumnos con y sin atraso escolar por medio de modelos estadísticos. Escuelas con scores más elevados en los dos indicadores aumentan las posibilidades de que sus alumnos estén en los niveles más altos de aprendizaje, pero las prácticas sintetizadas en los indicadores contribuyen para reducir las diferencias entre alumnos con y sin atraso escolar tan solo en relación a las posibilidades de que estén en el nivel básico. No hay que tomar a los indicadores de forma determinística, pero ellos poseen potencial para indicar problemas o éxitos en las prácticas pedagógicas de las escuelas.*

**PALABRAS CLAVE** INDICADORES EDUCACIONALES • PROVA BRASIL • DESIGUALDADES EDUCACIONALES • APRENDIZAJE.

# CONSTRUCTION OF INDICATORS TO DESCRIBE LEARNING INEQUALITIES IN PROVA BRASIL

## ABSTRACT

*This article presents two indicators that summarize complex dimensions of schools: intervention for improvement and school curriculum. They were constructed with the data provided by the Basic Education Evaluation System by means of the Item Response Theory. The indicators were used to analyze students' learning in elementary education in public schools and to test hypotheses regarding inequality among students with and without age-gap using statistical models. Schools with higher scores in the two indicators increase students' chances to be at the highest levels of learning. However, the practices summarized in the indicators contribute to reduce the differences between students with and without age-gap only in relation to their chances of being at the basic level. The indicators should not be used in a deterministic way, but they have the potential to show problems or successes in pedagogical practices in schools.*

**KEYWORDS** EDUCATIONAL INDICATORS • PROVA BRASIL • EDUCATIONAL INEQUALITIES • LEARNING.

## **INTRODUÇÃO**

As informações geradas pelas avaliações educacionais em larga escala têm sido utilizadas com objetivos muito diversificados. Na gestão educacional, as avaliações, direta ou indiretamente, são empregadas como instrumento para fomentar o avanço de objetivos educacionais e sociais, a fim de determinar a eficácia de políticas, práticas, programas e instituições e indivíduos, para informar o público os resultados educacionais, como fonte de evidências para pesquisas acadêmicas, entre outros fins (HERMAN; BAKER, 2009).

No Brasil, a utilização das avaliações na gestão educacional avançou bastante nas últimas décadas, principalmente nos governos estaduais e municipais que possuem sistemas próprios de avaliação, conforme Brooke e Cunha (2011). Em geral, as iniciativas locais ultrapassam o propósito de monitoramento dos resultados entre séries ou edições, ao desenvolver políticas que tenham objetivos pedagógicos para a melhoria do aprendizado dos alunos. Por exemplo, a identificação de escolas que precisam de alguma estratégia de ação e também a orientação à formação dos professores.

O uso pedagógico das avaliações é considerado como a sua função mais importante, porém, a mais difícil de ser executada (SOARES, 2002). No Plano Nacional da Educação (PNE), uma das metas mais desafiadoras é a de número 7, que prevê a elevação do Índice de Desenvolvimento da Educação Básica (Ideb) até atingir determinados valores que denotariam uma educação de qualidade (BRASIL, 2014). Como o Ideb é calculado pela síntese das médias escolares observadas bianualmente nas avaliações nacionais que compõem o Sistema de Avaliação da Educação Básica (Saeb)<sup>1</sup> e das taxas médias de aprovação em cada etapa escolar, as práticas pedagógicas que possam influir para melhoria desses indicadores nas escolas (notas nos testes e o fluxo escolar) tornam-se cruciais para a meta global.

O Plano Nacional de Educação prevê uma série de estratégias para essa meta, sendo que uma delas, a estratégia 7.3, refere-se ao desenvolvimento de indicadores educacionais baseados no perfil do alunado, de professores, da infraestrutura das escolas, de recursos pedagógicos disponíveis, das características da gestão, etc.

São de grande importância metodologias avaliativas e indicadores confiáveis em educação. O próprio texto do PNE reflete essa preocupação ao estabelecer a estratégia 7.10, que prevê, na vigência do plano, a fixação, o acompanhamento e a divulgação dos resultados pedagógicos dos indicadores do Sistema Nacional de Avaliação da Educação Básica e do Ideb, bem como a contextualização desses resultados.

Existe, portanto, a preocupação de que os resultados educacionais sejam contextualizados. Sem questionar as conclusões sobre o vínculo entre a origem social dos alunos e outros fatores externos à escola com os resultados educacionais, há evidências de que os sistemas de ensino e as escolas variam suficientemente para terem efeitos não negligenciáveis sobre o aprendizado dos alunos (BROOKE; SOARES, 2008; ALVES; FRANCO, 2008).

No último ano, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) divulgou alguns indicadores educacionais baseados no Censo Escolar, tais como: a adequação da formação docente, o esforço docente e

<sup>1</sup> O Sistema de Avaliação da Educação Básica é uma avaliação nacional realizada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, composto de três avaliações nacionais, sendo que duas delas fornecem dados para o cálculo do Ideb: a Avaliação Nacional da Educação Básica (Aneb), e a Avaliação Nacional do Rendimento Escolar (Anresc), também denominada Prova Brasil. Informações sobre o Saeb e o uso das avaliações no cálculo do Ideb podem ser obtidas em: <<http://portal.inep.gov.br/web/saeb/aneb-e-anresc>>. Acesso em: 10 jun. 2016.

a complexidade da gestão. O Inep divulgou também o Indicador de Nível Socioeconômico das famílias dos alunos (Inse), estimado com os dados das avaliações educacionais conduzidas nesse órgão público. Nos questionários contextuais das avaliações realizadas pelo Inep na educação básica, há itens que visam a medir outras dimensões importantes para os processos educacionais. Algumas dimensões são mais objetivas e diretamente observadas pelos itens dos questionários, tais como a formação dos professores e gestores, a qualidade da infraestrutura e os recursos das escolas. Mas há muitos itens relacionados a dimensões complexas alusivas às políticas e práticas pedagógicas, às características da gestão escolar e outras.

O objetivo deste artigo é propor uma forma de construção de indicadores contextuais que expressem algumas dimensões complexas pertinentes às práticas pedagógicas das escolas. Serão apresentados dois indicadores relacionados às características do ensino e ao clima acadêmico da escola: (1) intervenções da escola para a melhoria do fluxo e do aprendizado dos alunos, que será referido no texto simplesmente como intervenção para melhoria; e (2) cumprimento, adequação e forma de desenvolvimento do currículo escolar, referido como currículo na escola. Além da contribuição metodológica, que é da maior relevância para garantir a confiabilidade e o significado desses indicadores, pretende-se demonstrar a utilidade dos mesmos para a caracterização das escolas e das desigualdades de aprendizado, assim como para testar hipóteses sobre desigualdades entre grupos de alunos com e sem atraso escolar. Os dados utilizados são os produzidos pelo Saeb, que é o mais abrangente sistema para esse nível de ensino.

O artigo está estruturado em mais cinco seções após esta introdução. Nas duas próximas seções, apresentamos a abordagem teórica com as definições conceituais que guiaram a operacionalização empírica dos indicadores, e a abordagem analítica, que justifica os procedimentos de validação dos indicadores. Na sequência, destacamos os procedimentos para a construção dos indicadores. Na quinta seção, apresentamos as análises descritivas e os resultados dos modelos

estatísticos concebidos para testar hipóteses sobre a relação entre os indicadores e os níveis de aprendizado dos alunos do ensino fundamental. Nas considerações finais, discutimos as potencialidades e os limites dos indicadores para a pesquisa educacional.

### **ABORDAGEM TEÓRICA**

O uso de indicadores tem ampla aplicação na pesquisa social. Um indicador

[...] é uma medida em geral quantitativa, dotada de significado social substantivo, usado para substituir, quantificar ou operacionalizar um conceito social abstrato, de interesse teórico. (JANNUZZI, 2002, p. 55)

Muitos conceitos sociais possuem significados complexos e variados, o que torna bastante difícil produzir medidas simples que consigam captá-los. Geralmente são observados vários itens relativos às dimensões de interesse. Esses itens, quando correlacionados, podem ser combinados em um indicador composto, como os que serão apresentados neste artigo.

Os indicadores propostos são medidas obtidas por meio das análises estatísticas utilizadas para testar os construtos teóricos. Estes, por sua vez, referem-se a traços latentes não diretamente observados, mas que são importantes para a análise dos fenômenos educacionais. Na literatura, há vários trabalhos que apresentam indicadores ou fatores construídos com base em dados educacionais e que se relacionam aos propósitos deste artigo (ALVES; SOARES; XAVIER, 2014; KARINO; VINHA; LAROS, 2014; SOARES NETO et al., 2013).

O uso desse tipo de medida para expressar conceitos complexos, apesar de não ser novidade, permanece como um tema controverso. Saltelli e colaboradores (2004) argumentam que um indicador composto tem a vantagem de produzir uma visão mais ampla do fenômeno analisado, ao reduzir questões complexas e multidimensionais a uma forma sintética e mais comunicável para o público em geral. O indicador pode ser mais eficaz na descrição da realidade e na comparação entre lugares e ao longo do tempo, do que

quando se busca encontrar tendências em variáveis isoladas. Esses autores alertam, entretanto, que um indicador pode ser mal construído e mal interpretado, produzindo uma visão geral muito simplista (ou equivocada) da realidade, o que pode subsidiar políticas públicas ruins. Jannuzzi (2004) alerta para o risco do mau uso ou do abuso de indicadores, da banalização dos conceitos que se pretende expressar com a medida, risco maior quando não se explicita o modelo teórico ou o modelo de intervenção social.

Na área da educação, há uma tradição em lidar com indicadores complexos, para descrever uma realidade que está longe de ser simples. Como argumenta Goldstein (2001, p. 85),

Nas ciências sociais, e principalmente na Educação, as descrições dotadas de uma simplicidade elegante, muitas vezes, mostram-se errôneas. [...] a fim de descrever a complexa realidade que constitui os sistemas educacionais, precisamos de instrumentos de modelagem que envolvam um nível comparável de complexidade.

Um exemplo é a medida de nível socioeconômico. Nas avaliações educacionais em larga escala, os itens que permitem estimar esse indicador são planejados para fazerem parte dos questionários contextuais dessas avaliações, ainda que as opções metodológicas e os itens envolvidos variem um pouco de um trabalho para outro (ALVES; SOARES, 2009; ALVES; SOARES; XAVIER, 2014; SOARES, 2005; SOARES; ANDRADE, 2006). De uma forma geral, o nível socioeconômico é definido conceitualmente com base nas teorias sociológicas da estratificação social. Do ponto de vista empírico, essa medida é operacionalizada como um traço latente descrito por meio dos itens dos questionários contextuais respondidos pelos alunos. Os resultados obtidos nas pesquisas têm revelado alto grau de validade e confiabilidade para descrever as desigualdades sociais e educacionais.

Os questionários contextuais das avaliações incluem também itens para mensurar dimensões relacionadas às políticas e práticas pedagógicas das escolas. Entretanto, os referenciais teóricos para orientar a construção de indicadores sobre esses temas não são tão convergentes, como é o caso

do nível socioeconômico, ainda que alguns esforços para justificativa dos itens incluídos nos questionários tenham sido feitos (FRANCO et al., 2003).

Os dois indicadores propostos neste artigo foram orientados por estudos anteriores que demonstraram a relação entre as práticas pedagógicas das escolas e o desempenho dos alunos (REYNOLDS; TEDDLIE, 2008). As ações da escola para a melhoria do fluxo e do aprendizado dos alunos dizem respeito às políticas de intervenção pedagógicas planejadas e direcionadas para os alunos com baixo desempenho, risco de reprovação ou abandono. O sentido da intervenção é o de garantir o direito ao aprendizado e de prevenir o atraso escolar, visto que esses fenômenos são relacionados. A literatura internacional e nacional acumula inúmeras evidências de que a repetência é ineficaz para melhorar o aprendizado dos alunos com dificuldade e pode contribuir para o abandono precoce da escola (CRAHAY; 2007; RIBEIRO, 1991). Uma escola que promova esse tipo de ação de intervenção pode contribuir para a redução do fracasso escolar nos contextos em que o risco desse evento é maior.

O conjunto de ações da escola voltado para a recuperação e melhoria do aprendizado é uma das dimensões do clima escolar. Esse conceito tem sido caracterizado desde a década de 1970 a partir do trabalho seminal de Rutter e colegas (1979), que o caracterizou como o *éthos* escolar, envolvendo valores e normas de comportamento compartilhados na escola. Desde então, o conceito foi definido de muitas maneiras: como o ambiente organizacional da escola – filosofia educacional, procedimentos de enturmação, alocação de professores, envolvimento dos professores, etc. (MORTIMORE et al., 1988) –; como os valores da escola que a transformam em um ambiente propício ao aprendizado – ordenado, calmo, orientado para as tarefas e que estimula o autocontrole (SAMMONS, 1999) –; como a percepção dos alunos sobre as expectativas escolares, a pressão dos professores e as normas (MADAUS; AIRASIAN; KELLAGHAN, 1980). Todas essas definições são bastante amplas e podem ser operacionalizadas por mais de um indicador.

A opção metodológica deste trabalho foi construir um indicador que pudesse medir uma das dimensões do clima

2 BROOKOVER, W. B. et al.  
*School social systems and student  
achievement: schools make a  
difference.* New York: Praeger, 1979.

escolar: cultura escolar voltada para o aprendizado. O indicador intervenção para melhorias foi inspirado em um dos itens para medir o clima escolar elaborado por Brookover e colegas (1979<sup>2</sup> apud TEDDLIE; STRINGFIELD, 1993). Contudo, vale notar que as escolas com mais ações voltadas à recuperação e melhoria do aprendizado são aquelas que possuem mais alunos na condição de atraso escolar e dificuldades de aprendizado. O indicador não pode ser tomado como uma medida de altas expectativas da escola. Outras dimensões do clima foram operacionalizadas em outro trabalho das autoras (ALVES; XAVIER, no prelo).

Para a operacionalização empírica desse indicador, foram testados itens do questionário contextual do Saeb respondido pelos diretores sobre os programas desenvolvidos na escola para redução das taxas de abandono, reprovação e promoção ou garantia do direito à aprendizagem por parte dos alunos. O indicador expressa a ênfase acadêmica da escola e a preocupação com o desempenho dos alunos (aprendizado e rendimento) por parte da comunidade acadêmica.

O currículo escolar é um tema controverso na pesquisa educacional. Nas discussões acadêmicas, entram em disputa visões divergentes sobre as formas de seleção e organização dos conhecimentos nos currículos escolares e temas como diferença, identidade, multiculturalismo, pluralidade cultural, experiência, dominação cultural entre outros (MOREIRA, 2010; YOUNG, 2011). Ou seja, não é possível pensar em uma definição unívoca ou consensual sobre currículo.

Atualmente, a disputa em torno desse tema repercute nas discussões sobre a Base Nacional Comum dos Currículos (BNCC) prevista no PNE (BRASIL, 2014). Segundo esse documento, a BNCC deverá definir direitos e objetivos de aprendizagem para o currículo de cada área de conhecimento e ano do ensino fundamental e médio, respeitando a diversidade regional, estadual e local, até o final de 2016.

A noção de direitos e objetivos de aprendizagem encontra eco nas discussões sobre desigualdades e justiça social presentes nos escritos de Dubet (2004) e Young (2011). Dubet argumenta que as escolas devem garantir um nível de conteúdos da cultura escolar para todos os alunos, sem abando-

nar nenhum. Mas ele reconhece que a existência de programas educacionais comuns e obrigatórios não impede a desigualdade, porque os melhores alunos poderão tirar mais proveito e progredir mais depressa. Young considera simplista a ideia de que o currículo serve apenas para manter desigualdades, mas também julga ingenuidade esperar que desigualdades geradas em outras esferas da vida social serão superadas pelo currículo. Para dar conta desse paradoxo, ele defende que o currículo deve ser centrado em disciplinas escolares, que traduzem conhecimentos universalistas, para garantir a oportunidade de aprendizado dos estudantes mais desfavorecidos. Numa sociedade desigual, o currículo pode ser visto como uma política para equalização das oportunidades educacionais.

A dimensão do currículo proposta neste artigo – cumprimento, adequação e desenvolvimento do currículo escolar – é alinhada com a noção do direito ao acesso ao conhecimento e não ao conteúdo do currículo, que não é possível mensurar empiricamente com os dados utilizados. Entretanto, no contexto das avaliações externas, é lícito supor que essas iniciativas possam direta ou indiretamente se relacionar com os conteúdos curriculares (BROOKE; CUNHA, 2011). Para Bonamino e Souza (2012, p. 15), as avaliações têm potencial de

[...] propiciar uma discussão mais informada sobre o currículo escolar, em termos das habilidades fundamentais de leitura e matemática que ainda não têm sido garantidas a todos os alunos.

Para a operacionalização, foram considerados alguns itens do questionário contextual relacionados ao cumprimento e planejamento do currículo no ano letivo e à adequação do mesmo ao perfil dos alunos. O indicador expressa a ênfase da escola no direito dos alunos ao aprendizado, sem desconsiderar as condições contextuais de cada escola.

Cabe, ainda, justificar qual a pertinência de se propor indicadores para essas dimensões em vez de analisar os itens separadamente. Afinal, quando se trata de indicadores referidos a conceitos que carecem de um modelo teórico mais forte, não haveria o risco de uma “reificação” da medida? (JANNUZZI, 2002).

Nossa resposta considera, em primeiro lugar, a potencialidade do uso dos indicadores nos modelos de análise sobre desempenho escolar dos alunos. A pesquisa em avaliação educacional visa a encontrar evidências sobre a eficácia das escolas. Na educação, os indicadores podem ser empregados como estatísticas que refletem condições sociais ou escolares, isto é, que “indicam” alguma coisa sobre a natureza e a qualidade de uma situação. Nessa perspectiva, os indicadores apresentados neste artigo têm o potencial de indicar problemas ou êxitos nas práticas pedagógicas das escolas e ajudar no planejamento, gestão e monitoramento das políticas e projetos nessa área.

Em segundo lugar, nossas análises são constrangidas pelos limites dos dados. Os itens utilizados para a construção dos indicadores foram obtidos em questionários contextuais que possuem dados ausentes. A metodologia utilizada, conforme será descrita a seguir, permite manter todos os casos (alunos e escolas) nos modelos estatísticos de análise mesmo com esse limite. Mas como os indicadores são construções empíricas com dados limitados, as suas aplicações devem ser bem criteriosas.

### **ABORDAGEM ANALÍTICA**

Para a validação dos indicadores, apresentamos dois tipos de análises: a distribuição dos respectivos escores segundo níveis de aprendizado dos alunos na Prova Brasil; e modelos estatísticos multivariados nos quais os indicadores são variáveis explicativas.

Os níveis de aprendizado são faixas de valores calculadas a partir da escala de proficiência do Saeb, que possui intervalo de 0 a 500 pontos. Construir uma análise por níveis de aprendizado em vez da análise convencional das médias de proficiência tem a vantagem de evidenciar com mais nitidez os padrões de desigualdades entre grupos sociais, uma vez que a média é uma medida sensível a valores extremos. Por exemplo, é possível que o desempenho médio de uma escola melhore pelo aumento da proficiência de poucos alunos que já apresentam bons resultados e não necessariamente pela diminuição das desigualdades entre os alunos.

Neste trabalho, adotamos a interpretação pedagógica proposta por Soares (2009), que consiste na divisão dessa escala em quatro pontos de cortes aos quais são atribuídos rótulos com significado normativo: nível abaixo do básico, básico, adequado e avançado. Os dois níveis mais altos foram unificados em razão do baixo percentual de alunos no nível avançado (sobretudo no 9º ano) e porque os alunos nesse nível já atingiram a expectativa de aprendizado, assim como os alunos no nível adequado, conforme discutido por Soares (2009).

Os modelos estatísticos multivariados visam a estimar, com o controle de variáveis de características do aluno e da escola, a influência dos indicadores intervenção e currículo nas chances de aprendizado dos alunos. Esses modelos buscam analisar as múltiplas influências sobre as desigualdades de aprendizado e, ao mesmo tempo, os efeitos separados dessas influências (HAIR et al., 2005).

A hipótese subjacente a esse trabalho é de que as escolas com os escores mais altos dos indicadores são mais eficazes para garantir o aprendizado dos seus alunos. Além de estimar a influência dos indicadores, testamos também a hipótese de que os indicadores têm efeitos diferenciados para alunos com e sem atraso escolar. A variável atraso escolar foi escolhida porque, no Brasil, a prática da “Pedagogia da repetência”, apontada por Sérgio Costa Ribeiro (1991), permanece com índices alarmantes no século XXI. E também porque, neste estudo, observamos que os percentuais de alunos com atraso escolar são bem maiores no nível abaixo do básico em todas as edições da Prova Brasil, no 5º e 9º anos, conforme análise descritiva apresentada na Tabela A1 do Apêndice.

Se as práticas educacionais sintetizadas nos indicadores tiverem eficácia, esperamos que, nas escolas com escores mais altos dos indicadores, os alunos com atraso escolar tenham reduzidas suas chances de estarem no nível abaixo do básico e aumentadas as chances de estarem no nível adequado.

Para essas análises, ajustamos um conjunto de modelos para o desempenho dos alunos em Leitura. Nesses modelos, a variável resposta é o nível de aprendizado do aluno, uma variável categórica. Essa característica requer o uso de modelos estatísticos não lineares; especificamente, utilizamos

os modelos de regressão múltipla hierárquica multinomial, que serão descritos junto aos resultados.

### **CONSTRUÇÃO DOS INDICADORES**

Os indicadores intervenção para melhoria e currículo foram criados por meio de uma técnica estatística multivariada. Especificamente, utilizamos a Teoria da Resposta ao Item (TRI), que engloba uma série de modelos cujo objetivo principal é a obtenção de medidas de construtos latentes (fatores), baseados em itens dicotômicos e/ou ordinais (HAMBLETON et al., 1993). A TRI é semelhante à análise fatorial porque ambas estimam variáveis latentes, embora a TRI trate melhor os dados ausentes. A TRI é mais conhecida para a avaliação da qualidade de itens de testes e estimação da habilidade dos alunos, mas é também empregada para estimar indicadores, fatores ou índices (ALVES; SOARES; XAVIER, 2014; SOARES NETO et al., 2013). Para os fins deste trabalho, não distinguimos os termos indicador, fator ou índice, ainda que na literatura técnica isso seja enfatizado.

### **DADOS**

Para a construção dos indicadores, foram utilizadas duas bases de dados que fazem parte do Saeb: a Prova Brasil, de 2007 a 2013, e a Avaliação Nacional da Educação Básica (Aneb), de 2011 e 2013.<sup>3</sup> A Prova Brasil contém as proficiências em Língua Portuguesa e Matemática dos alunos matriculados no 5º e 9º anos do ensino fundamental de todas as escolas públicas que tenham pelo menos 20 alunos no ano escolar avaliado. Além disso, incluem informações obtidas por meio dos questionários respondidos pelos alunos, professores, diretores e um questionário sobre a escola preenchido pelo aplicador. A Aneb utiliza os mesmos instrumentos da Prova Brasil, mas com uma cobertura amostral das escolas públicas não elegíveis pela Prova Brasil (escolas com menos de 20 alunos nas séries avaliadas), escolas privadas e a 3ª série do ensino médio. A utilização de todas essas bases de dados permite que os indicadores construídos sejam representativos de todos os tipos de escolas e comparáveis ao longo do tempo.

<sup>3</sup> Os microdados estão disponíveis em: <<http://portal.inep.gov.br/basicalevantamentos-acessar>>. Acesso em: 18 maio 2016.

O Quadro 1 mostra o número de casos em cada uma das bases de dados.

**QUADRO 1 - Número de casos em cada uma das avaliações consideradas no estudo**

AVALIAÇÃO	ESCOLAS	PROFESSORES	DIRETORES
Prova Brasil 2007	48.149	292.828	47.596
Prova Brasil 2009	58.374	216.495	58.374
Saeb 2011	58.960	316.668	58.960
Saeb 2013	59.251	237.186	37.136

Fonte: Elaboração das autoras com dados do Saeb/Prova Brasil de 2007 a 2013.

Nota: O Saeb 2011 e o Saeb 2013 reúnem, na mesma base de dados, a Prova Brasil (escolas públicas) e a Aneb - a subamostra de escolas não elegíveis para a Prova Brasil representativa das escolas privadas e das públicas com menos de 20 alunos.

### **PREPARAÇÃO DAS BASES DE DADOS**

Os itens que compõem o indicador intervenção para melhoria são originários do questionário do diretor e os itens do indicador currículo na escola são do questionário do professor. No caso dos dados dos professores, os itens foram agregados por escola segundo a média das respostas dos professores. O resultado da agregação é assumido como uma medida do contexto da escola. Como consequência, cada escola possui a informação do seu diretor e a média das respostas dos seus professores. Os itens relacionados a cada um dos indicadores foram compatibilizados em todas as bases de dados. Para o tratamento das variáveis, foi utilizado o *software* SPSS/IBM 20.0.

O indicador intervenção para melhoria foi estimado a partir de quatro itens: 1) se a escola desenvolve programas para: redução do abandono; 2) [idem]: redução da reprovação; 3) [idem]: para apoio ou reforço de aprendizagem dos alunos; e 4) frequência com que o diretor discute medidas com o objetivo de melhorar o ensino e a aprendizagem dos alunos com os professores. O indicador currículo na escola foi estimado a partir de três itens: 1) se os conteúdos curriculares são inadequados às necessidades dos alunos; 2) se os problemas de aprendizado estão relacionados ao não cumprimento do conteúdo curricular; e 3) o percentual dos conteúdos previstos que o professor conseguiu desenvolver.

Todos os itens estão presentes nas edições do Saeb/Prova Brasil analisadas, exceto o item 4 do indicador intervenção, presente apenas na edição 2013.

## ESTIMAÇÃO

Para a estimação dos indicadores, empregou-se um modelo da TRI (HAMBLETON, 1993). Especificamente, utilizou-se o modelo de Samejima (1969), apropriado para itens com respostas graduadas (ordinais). O modelo tem como suposição a unidimensionalidade, isto é, a existência de um único construto latente dominante no conjunto dos dados. Ou seja, assume-se que os itens possuem uma relação causal com o indicador. Por exemplo, para o indicador currículo na escola, quanto maior o percentual dos conteúdos previstos desenvolvidos, maior o escore do indicador. A mesma relação é esperada nos outros itens. Essa suposição precisa ser testada para a validação dos construtos antes da estimação do indicador.

Isso é usualmente realizado via análise da matriz de correlação policórica do construto. As variáveis estão associadas a um único construto quando todas são positivamente correlacionadas. Isso pode ser confirmado para os dois indicadores, uma vez que todas as variáveis estão positivamente correlacionadas. Em análises prévias, outras variáveis foram testadas, mas não se justificaram nos construtos após esse teste. Essas análises foram realizadas com auxílio do *software* IRTPRO 3.<sup>4</sup>

O modelo TRI foi ajustado em duas fases. Na primeira fase, foram estimados os parâmetros dos itens.<sup>5</sup> Na segunda, a partir dos parâmetros dos itens e da distribuição das respostas, foram estimados os escores do indicador. Pelo método empregado, foram considerados apenas os itens para os quais os indivíduos forneceram respostas para se estimar o valor do escore de interesse. Essa é uma vantagem da TRI em relação aos métodos convencionais e bastante adequada para o problema desta pesquisa, que contém dados incompletos, seja porque o item não foi apresentado em alguma das edições do Saeb/Prova Brasil, seja porque os indivíduos não responderam o item apresentado em alguma das edições. Os casos sem resposta em todos os itens foram excluídos.

<sup>4</sup> Além da matriz de correlação policórica, comumente são calculados os autovalores e autovetores do construto, sendo que o primeiro testa a suposição da unidimensionalidade e o segundo, a força da associação entre as variáveis originais e a variável latente (o indicador). As tabelas com os autovalores e os autovetores indicam que os indicadores estimados possuem uma única dimensão latente. Por razões de espaço, as tabelas com essas análises não estão reportadas neste texto, mas estão disponíveis sob requisição.

<sup>5</sup> Os parâmetros estimados dos itens são utilizados para plotar o gráfico da curva característica do item (CCI) e o gráfico da curva de informação do item (CII), para cada item testado. Esses gráficos também são importantes para se avaliar a qualidade dos itens. Os gráficos CCI sugerem a relação entre a probabilidade de um indivíduo dar uma determinada resposta a um item e seu traço latente (o fator estimado).

O CII mostra em qual intervalo da escala do construto latente de interesse um item específico fornece a maior quantidade de informação na estimação dessa escala. Os gráficos não estão apresentados no texto por questão de espaço, mas estão disponíveis sob requisição.

Cabe também ressaltar que eventualmente esses indicadores não são capazes de mensurar o traço latente com a riqueza devida, porque os itens dos questionários não necessariamente foram planejados para medi-los. Ocorre que a mensuração de conceitos complexos é difícil e as escolhas sobre como fazê-la dependem dos referenciais dos pesquisadores e dos dados disponíveis. Além disso, como a seleção dos itens se deu por um julgamento *a posteriori*, formas alternativas de agrupamentos podem ser propostas.

Os escores estimados para os indicadores variavam, originalmente, numa escala de desvios padrão. Para facilitar a interpretação, os valores foram transformados para um escala com intervalo de [0;10] pontos.<sup>6</sup> A Tabela 1 apresenta as distribuições dos valores de cada um dos indicadores. Os números de casos para cada indicador se referem aos que foram utilizados na estimação de cada um deles, considerando os dados das edições do Saeb/Prova Brasil de 2007 a 2013 (escolas públicas e privadas). Nota-se que as médias dos escores tendem para valores altos, especialmente o indicador currículo na escola.

<sup>6</sup> Para a transformação da escala, é feito o seguinte cálculo (indicador - limite inferior/limite superior - limite inferior)\*10.

**TABELA 1 - Medidas descritivas e Intervalo de Confiança para a média dos indicadores propostos**

INDICADOR	NÚMERO DE CASOS	VALOR MÍNIMO	VALOR MÁXIMO	MÉDIA	DESVIO PADRÃO	INTERVALO DE CONFIANÇA (95%)
Intervenção	219.221	0,00	10,00	5,92	1,43	[5,91 ; 5,93]
Currículo	216.527	0,00	10,00	8,15	1,70	[8,14 ; 8,16]

Fonte: Elaboração das autoras com dados do Saeb/Prova Brasil de 2007 a 2013.

## **APLICAÇÃO DOS INDICADORES PARA DESCREVER DESIGUALDADES DE APRENDIZADO**

Nesta seção, apresentamos os dois conjuntos de análises para validação dos indicadores. Primeiro, a distribuição dos escores dos indicadores segundo níveis de aprendizado dos alunos das escolas públicas de ensino fundamental (municipal e estadual) de 2007 a 2013. Em seguida, os modelos estatísticos multivariados ajustados a esses dados. As estatísticas descritivas têm como unidade de análise os alunos e os modelos de

regressão; pela sua natureza multinível, tomam como unidade de análise alunos e escolas, ao mesmo tempo.

### DISTRIBUIÇÃO DOS INDICADORES POR NÍVEIS DE APRENDIZADO

Nas análises descritivas, serão apresentadas as médias dos indicadores, de 2007 a 2013, e as diferenças médias desses indicadores: entre os alunos que estão no nível abaixo do básico (AB) e os que estão no nível adequado/avançado (AD); e entre os alunos no nível básico (B) e os que estão no nível adequado/avançado (AD). Com isso, pretende-se analisar se há tendência de equalização das condições de escolarização, medidas pelos indicadores, considerando os níveis de desempenho. Como os dados da Prova Brasil têm cobertura censitária das escolas públicas de ensino fundamental, as diferenças observadas são diferenças reais.

### INTERVENÇÃO PARA MELHORIA

Nas tabelas 2 e 3, observamos que as médias do indicador intervenção para melhoria são maiores para o nível de aprendizado adequado e com valores crescentes de 2007 a 2013. Porém, as diferenças entre os níveis se reduziram, o que sugere uma tendência para a equalização desse indicador entre as escolas.

**TABELA 2 - Média do indicador intervenção para melhoria por níveis de aprendizado dos alunos em Leitura segundo o ano escolar e a edição da Prova Brasil**

ANO ESCOLAR	EDIÇÃO	NÍVEIS DE APRENDIZADO EM LEITURA			DIFERENÇAS ENTRE OS NÍVEIS	
		ABAIXO DO BÁSICO (AB)	BÁSICO (B)	ADEQUADO/AVANÇADO (AD)	DIFERENÇA AD - AB	DIFERENÇA AD - B
5º ano	2007	5,83	6,01	6,33	0,50	0,33
	2009	5,75	5,97	6,29	0,55	0,32
	2011	5,84	6,09	6,43	0,59	0,34
	2013	6,04	6,17	6,34	0,30	0,18
9º ano	2007	5,47	5,62	5,87	0,39	0,25
	2009	5,55	5,67	5,87	0,32	0,20
	2011	5,59	5,75	5,98	0,39	0,23
	2013	5,96	6,02	6,12	0,15	0,10

Fonte: Elaboração das autoras a partir dos dados da Prova Brasil 2007 a 2013.

**TABELA 3 – Média do indicador intervenção para melhoria por níveis de aprendizado dos alunos em Matemática segundo o ano escolar e a edição da Prova Brasil**

ANO ESCOLAR	EDIÇÃO	NÍVEIS DE APRENDIZADO EM MATEMÁTICA			DIFERENÇAS ENTRE OS NÍVEIS	
		ABAIXO DO BÁSICO (AB)	BÁSICO (B)	ADEQUADO/AVANÇADO (AD)	DIFERENÇA AD - AB	DIFERENÇA AD - B
5º ano	2007	5,85	6,04	6,36	0,51	0,32
	2009	5,74	5,99	6,33	0,59	0,34
	2011	5,82	6,11	6,49	0,67	0,38
	2013	6,05	6,19	6,36	0,31	0,17
9º ano	2007	5,47	5,67	5,95	0,48	0,28
	2009	5,57	5,73	5,92	0,35	0,19
	2011	5,62	5,81	6,02	0,40	0,21
	2013	5,97	6,05	6,16	0,19	0,11

Fonte: Elaboração das autoras a partir dos dados da Prova Brasil 2007 a 2013.

#### CURRÍCULO NA ESCOLA

As tabelas 4 e 5 mostram que os alunos nos níveis mais baixos de aprendizado estão em escolas em que a média do indicador currículo na escola é menor. As diferenças das médias do indicador entre o nível adequado e abaixo do básico (AD-AB) apresentaram um padrão de aumento ao longo das edições da Prova Brasil em Leitura e Matemática.

**TABELA 4 – Média do indicador currículo na escola por níveis de aprendizado dos alunos em Leitura segundo o ano escolar e a edição da Prova Brasil**

ANO ESCOLAR	EDIÇÃO	NÍVEIS DE APRENDIZADO EM LEITURA			DIFERENÇAS ENTRE OS NÍVEIS	
		ABAIXO DO BÁSICO (AB)	BÁSICO (B)	ADEQUADO/AVANÇADO (AD)	DIFERENÇA AD - AB	DIFERENÇA AD - B
5º ano	2007	7,95	8,15	8,43	0,48	0,28
	2009	7,83	8,09	8,42	0,59	0,33
	2011	8,16	8,41	8,72	0,56	0,31
	2013	7,94	8,27	8,67	0,73	0,39
9º ano	2007	7,98	8,14	8,35	0,37	0,21
	2009	7,88	8,04	8,24	0,36	0,21
	2011	8,15	8,31	8,51	0,36	0,20
	2013	7,96	8,14	8,38	0,42	0,24

Fonte: Elaboração das autoras a partir dos dados da Prova Brasil 2007 a 2013.

**TABELA 5 - Média do indicador currículo na escola por níveis de aprendizado dos alunos em Matemática segundo o ano escolar e a edição da Prova Brasil**

ANO ESCOLAR	EDIÇÃO	NÍVEIS DE APRENDIZADO EM MATEMÁTICA			DIFERENÇAS ENTRE OS NÍVEIS	
		ABAIXO DO BÁSICO (AB)	BÁSICO (B)	ADEQUADO/ AVANÇADO (AD)	DIFERENÇA AD - AB	DIFERENÇA AD - B
5º ano	2007	7,97	8,18	8,47	0,50	0,29
	2009	7,82	8,11	8,47	0,65	0,36
	2011	8,14	8,44	8,78	0,64	0,34
	2013	7,94	8,33	8,72	0,78	0,39
9º ano	2007	7,99	8,18	8,44	0,46	0,26
	2009	7,90	8,10	8,37	0,47	0,27
	2011	8,16	8,36	8,60	0,44	0,25
	2013	7,96	8,21	8,53	0,57	0,32

Fonte: Elaboração das autoras a partir dos dados da Prova Brasil 2007 a 2013.

## MODELOS ESTATÍSTICOS

### DESCRIÇÃO DOS MODELOS MULTINOMIAIS HIERÁRQUICOS

Para analisar a influência dos indicadores intervenção para melhoria e currículo na escola nas chances de aprendizado dos alunos, foram ajustados três modelos multinomiais hierárquicos (RAUDENBUSH et al., 2011, p. 325) a partir dos dados das quatro edições da Prova Brasil, com auxílio do *software* HLM 7.01. Utilizamos informações sobre 66.909 escolas e 16.531.533 alunos do 5º e 9º anos do ensino fundamental de escolas públicas.

O uso de modelos hierárquicos para a estimação dos fatores associados ao desempenho escolar é uma tradição consolidada nas pesquisas educacionais. Como os alunos estão agrupados em escolas, essa família de modelos surge como opção analítica natural (RAUDENBUSH; BRYK, 2002). Neste estudo, as unidades de análise de nível 1, indexadas por (i), são os alunos e as de nível 2, indexadas por (j), são as escolas.

Em todos os modelos, a variável dependente são os níveis de aprendizado em Leitura: abaixo do básico, básico e adequado. O último nível é resultado da soma dos níveis adequado e avançado, e será referido somente como nível adequado, como explicado na seção “Abordagem analítica”. Nos modelos, o nível básico foi tomado como categoria

de referência. Foram estimadas: 1) as chances de um aluno estar no nível abaixo do básico, comparado à chance de estar no nível básico; e 2) as chances de um aluno estar no nível adequado, comparado às chances de estar no nível básico.

O modelo 1 é um modelo básico que inclui as características do aluno (sexo, cor, nível socioeconômico, atraso escolar, série e edições da Prova Brasil) e da escola (média do nível socioeconômico da escola, proporção de alunos atrasados e os indicadores de intervenção para melhoria e currículo). Nesse modelo, são de especial interesse os coeficientes dos indicadores de intervenção para melhoria e currículo na escola. Esses coeficientes mostram o quanto a variação em 1 ponto em cada um dos indicadores afeta as probabilidades de um aluno estar nos níveis abaixo do básico ou adequado, independentemente das características de alunos e escolas que estão controladas no modelo.

As variáveis sexo e raça/cor são itens diretos do questionário contextual respondido pelo aluno. Em relação à raça/cor, optamos por trabalhar com a população amarela e indígena em uma única categoria, chamada “outra cor”, porque, em termos estatísticos, essa população é muito pequena e as inferências seriam menos precisas.

O nível socioeconômico (NSE) do aluno é medido através da síntese de vários itens do questionário do aluno nas avaliações educacionais, via TRI – modelo Samejima (1969) –, relacionados à escolaridade e setor ocupacional de seus pais, à posse de vários itens de conforto e à contratação de empregados domésticos no domicílio.

O atraso escolar mede a diferença entre a idade do aluno e a idade esperada para o ano escolar em que ele se encontrava quando da realização do teste. A medida de atraso foi calculada a partir das informações disponíveis nos questionários do aluno, quais sejam: idade e ocorrência de reprovação e/ou abandono.

O NSE da escola e o percentual de alunos com atraso na escola são medidas obtidas por agregação dos dados individuais. Os procedimentos para o cálculo dos indicadores de intervenção para melhoria e currículo na escola já foram explicados na seção “Construção dos indicadores”.

As variáveis série e edição da Prova Brasil foram incluídas nos modelos como controles porque os dados de todas as edições e das duas séries do ensino fundamental foram analisados em conjunto.

Cabe mencionar que os dados da Prova Brasil são transversais, porém a medida de proficiência é planejada para ser comparável ao longo das edições do teste e entre as séries avaliadas.<sup>7</sup> Outra característica importante desses dados diz respeito à possibilidade de identificação de uma mesma escola ao longo das edições, permitindo uma análise longitudinal ao nível dos estabelecimentos escolares. Contudo, isso não pode ser feito no nível do aluno.

A opção deste estudo foi tratar os dados transversais como uma série temporal. Para tal, é fundamental o controle pela edição do teste. Essa abordagem possui limitações, pois a natureza dos dados não permite que os efeitos das variáveis sejam interpretados como uma causalidade. Entretanto, para os nossos fins, o modelo escolhido fornece informações mais robustas do que um modelo de apenas um ano específico. Isso porque o interesse está centrado mais no comportamento dos indicadores e menos nas flutuações ao longo do tempo.

O modelo 2, além de todas as variáveis que constam do modelo 1, inclui a interação entre o indicador de intervenção para melhoria e o atraso escolar. Esse modelo permite testar se as ações para melhoria do aprendizado, redução do abandono e da reprovação na escola contribuem para diminuir o efeito do atraso escolar do aluno sobre as suas chances de aprendizado. Já o modelo 3, que inclui a interação entre o indicador de currículo na escola e o atraso escolar, visa a testar o mesmo tipo de hipótese em relação às práticas sintetizadas no indicador.

Um efeito de interação existe quando o efeito da variável independente (no nosso caso, o atraso escolar) sobre a dependente (níveis de aprendizado) difere segundo uma terceira variável (no nosso caso, intervenção para melhoria ou currículo na escola) (JACCARD, 2001). Substantivamente, a pergunta que fundamenta esses dois últimos modelos é: as práticas da escola são capazes de diminuir as desigualdades de aprendizado entre alunos com e sem atraso?

**7** Na Prova Brasil, o desempenho dos alunos é estimado por meio de um modelo de três parâmetros da Teoria da Resposta ao Item e os escores resultantes recebem o nome técnico de proficiência. A escala de proficiência originalmente em desvios padrão é transformada para valores de 0 a 500 pontos. Como a escala é a mesma para os diferentes anos, pode-se assumir que as variações na proficiência dos alunos de um determinado ano escolar, ao longo das diferentes edições de aplicação, refletem a melhoria ou piora no aprendizado das coortes avaliadas. Além disso, como a mesma escala é usada para expressar os escores dos alunos dos diferentes anos escolares, os alunos do 5º ano têm proficiências com valores mais baixos do que os alunos do 9º ano.

As variáveis independentes dos modelos, suas métricas e formas de centralização podem ser consultadas no Quadro 2:<sup>8</sup>

<sup>8</sup> As equações dos modelos não estão apresentadas por questão de espaço, mas estão disponíveis sob requisição.

**QUADRO 2 - Descrição das variáveis independentes utilizadas nos modelos de regressão hierárquica multinomial**

NÍVEL	VARIÁVEL	TIPO	DESCRIÇÃO	CENTRALIZAÇÃO*	
Nível 1 (alunos)	SEXO Feminino	Binária	1 = feminino e 0 = masculino	Métrica original	
	Ausentes SEXO Feminino	Binária	1 = não respondeu e 0 = respondeu	Métrica original	
	COR Parda	Binária	1 = parda e 0 = branca	Métrica original	
	COR Preta	Binária	1 = preta e 0 = branca	Métrica original	
	COR Outra	Binária	1 = outra cor e 0 = branca	Métrica original	
	NSE	Contínua	NSE do aluno		Centralizada na grande média
			Escala de 0 a 10 pontos		
	Atraso escolar	Binária	1 = atraso de 1 ou mais anos e 0 = sem atraso	Métrica original	
	Série	Binária	1 = 5º ano e 0 = 9º ano	Métrica original	
	Ano 2009	Binária	1 = 2009 e 0 = 2007	Métrica original	
	Ano 2011	Binária	1 = 2011 e 0 = 2007	Métrica original	
Ano 2013	Binária	1 = 2013 e 0 = 2007	Métrica original		
Nível 2 (escolas)	NSE da escola	Contínua	Média do NSE do aluno por escola	Centralizada na grande média	
			Escala de 0 a 10 pontos		
	Proporção de alunos com atraso escolar	Contínua	Proporção de alunos com atraso escolar por escola	Centralizada na grande média	
	Intervenção para melhorias	Contínua	Escala de 0 a 10 pontos	Centralizada na grande média	
Currículo	Contínua	Escala de 0 a 10 pontos	Centralizada na grande média		

Fonte: Elaboração das autoras.

Nota: \* Ver Raudenbush e Bryk (2002) sobre centralização em ambos os níveis.

## CHANCES DE APRENDIZADO

A Tabela 6 apresenta os coeficientes e as razões de chances dos três modelos de regressão hierárquica multinomial estimados para os níveis de aprendizado em Leitura.

Os coeficientes das variáveis do modelo 1 podem ser interpretados como o efeito aditivo do aumento de uma unidade da variável X em questão na probabilidade de estar

na categoria 1 (abaixo do básico) ou 2 (adequado) em vez de estar na categoria de referência (nível básico). As razões de chances também são apresentadas na tabela porque são mais interpretáveis. Elas são obtidas pelo  $\text{Exp}(\text{coeficiente})$ , que é interpretado como o efeito multiplicativo do aumento de uma unidade em X sobre as chances de um indivíduo estar na categoria 1 ou 2 em vez de estar na categoria de referência (nível básico).

O modelo 1 permite testar a hipótese de que as escolas com os escores mais altos dos indicadores são mais eficazes para o aprendizado dos seus alunos. Pela Tabela 6, observamos que os coeficientes das variáveis desse modelo estão no sentido esperado. Por exemplo, a média do nível socioeconômico da escola, para a categoria 1: abaixo do básico, possui um coeficiente negativo e, portanto, as razões de chances estão abaixo de 1. Isso significa que cada escore a mais do nível socioeconômico da escola diminui em 0,688 vezes as chances de um aluno estar abaixo do básico, comparado a estar no nível básico. O coeficiente da variável também pode ser lido em termos de incremento percentual,<sup>9</sup> o que equivale a dizer que cada ponto a mais na escala de nível socioeconômico da escola reduz em 31,2% as chances de um aluno estar no nível abaixo do básico. Analisando o coeficiente no nível socioeconômico da escola para a categoria 2: adequado/avançado, notamos que o sinal do coeficiente é positivo; conseqüentemente, as razões de chances estão acima de 1. Cada escore a mais no nível socioeconômico da escola aumenta em 1,580 vezes as chances de um aluno estar no nível adequado/avançado em relação a estar no nível básico. Em termos percentuais, equivale a dizer que ele possui 58% mais chances de estar no nível adequado. A mesma leitura pode ser feita em relação a todos os outros coeficientes.

Mas nosso interesse são os coeficientes de intervenção para melhoria e currículo. Eles são negativos para as chances de os alunos estarem no nível abaixo do básico e positivos para as chances de os alunos estarem no nível adequado. Isso significa que cada escore a mais do indicador de intervenção para melhorias diminui em 0,949 vezes as chances de um aluno estar no nível abaixo do básico, o que equivale a dizer

<sup>9</sup> Para transformar os coeficientes em incremento percentual, aplica-se a seguinte fórmula:  $[\text{Exp}(b)-1] \times 100$ .

que ele possui 5,1% menos chances de estar nesse nível, em relação a estar no nível básico. Além disso, cada ponto a mais na escala do indicador aumenta em 1,125 vezes as chances de um aluno estar no nível adequado/avançado em relação a estar no nível básico. Em termos percentuais, um aluno possui menos 12,5% de chances a mais de estar no nível adequado/avançado. Os escores do indicador de currículo têm valores muito próximos ao indicador de intervenção. Um aluno que estuda em uma escola onde o currículo foi cumprido e considerado adequado tem menos chances (0,925 ou -7,5%) de estar no nível abaixo do básico e mais chances (1,091 ou +9,1%) de estar no nível adequado, em relação a estar no nível básico. Esse resultado mostra que as práticas da escola melhoraram as probabilidades médias de aprendizado em Leitura, independentemente das outras características dos alunos e das escolas controladas no modelo de regressão. O resultado corrobora a relação entre práticas pedagógicas das escolas e o desempenho dos alunos (REYNOLDS; TEDDLIE, 2008).

Os modelos 2 e 3 testam a hipótese de que os indicadores têm efeitos diferenciados para alunos com e sem atraso escolar. No modelo 2, os coeficientes de interação entre intervenções e atraso ( $\gamma_{61(1)}$  e  $\gamma_{61(2)}$ ) são interpretados da seguinte forma: cada ponto a mais na escala do indicador eleva ou reduz (a depender do sinal do coeficiente) o efeito da variável atraso escolar sobre as chances de o aluno estar no nível abaixo do básico ( $\gamma_{61(1)}$ ) ou de estar no nível adequado ( $\gamma_{61(2)}$ ). Os coeficientes da interação entre currículo e atraso do modelo 3 ( $\gamma_{61(1)}$  e  $\gamma_{61(2)}$ ) são interpretados do mesmo modo.

No modelo 2, o coeficiente da interação para a categoria abaixo do básico ( $\gamma_{61(1)}$ ) é negativa, e a interação para a categoria adequado ( $\gamma_{61(2)}$ ) é positiva. Isso significa que cada ponto na escala do indicador de intervenção diminui (em 0,980 vezes ou em -2%) a diferença entre alunos com e sem atrasos em relação às chances de estar no nível abaixo do básico. Por outro lado, o valor positivo da interação ( $\gamma_{61(2)}$ ) para a categoria adequado significa que cada ponto na escala do indicador aumenta (em 1,017 vezes ou em +1,7%) a diferença entre alunos com e sem atraso escolar quanto às chances de estarem no nível adequado.

No modelo 3, a interpretação é a mesma do modelo 2 e os resultados são parecidos. A interação entre currículo e atraso para a categoria abaixo do básico ( $\gamma_{61(1)}$ ) é negativa, e a interação para a categoria adequado ( $\gamma_{61(2)}$ ) é positiva, mas esta última não é estatisticamente significativa. Cada ponto na escala do indicador de currículo diminui as chances (em 0,995 vezes ou em -0,5%) de um aluno com atraso estar no nível abaixo do básico comparativamente ao aluno sem atraso. Para o nível adequado, a interação não afeta a diferença entre os dois grupos. Nesse caso, a ausência de significância pode ser decorrente da distribuição do indicador. Como há muitas escolas com escores altos para o currículo, talvez não tenha sido possível discriminar as diferenças entre esses dois grupos de alunos nas escolas com médias mais altas do indicador.

**TABELA 6 - Estimativas dos parâmetros dos Modelos de Regressão Hierárquica Multinomial para os níveis de aprendizado em Leitura**

EFEITOS FIXOS	MODELO 1	MODELO 2	MODELO 3
	RAZÃO DE CHANCES	RAZÃO DE CHANCES	RAZÃO DE CHANCES
<b>PARA A CATEGORIA 1: ABAIXO DO BÁSICO</b>			
<b>NÍVEL 1 - ALUNOS</b>			
SEXO Feminino	0,618	0,618	0,618
Ausentes SEXO Feminino	1,504	1,505	1,504
COR Parda	0,95	0,95	0,95
COR Preta	1,197	1,197	1,197
COR Outra	1,172	1,172	1,172
NSE	0,95	0,951	0,95
Atraso escolar	1,052	1,056	1,053
Série	1,357	1,354	1,356
Ano 2009	0,788	0,789	0,788
Ano 2011	0,764	0,764	0,764
Ano 2013	0,855	0,855	0,855
<b>NÍVEL 2 - ESCOLAS</b>			
Intercepto	0,585	0,583	0,584
NSE da escola	0,688	0,688	0,688
Proporção de alunos com atraso	1,158	1,161	1,158
Intervenção para melhorias	0,949	0,96	0,949
Currículo	0,925	0,925	0,928
Intervenção para melhorias*Atraso	-	0,98	-
Currículo*Atraso	-	-	0,995
<b>PARA A CATEGORIA 2: ADEQUADO / AVANÇADO</b>			
<b>NÍVEL 1 - ALUNOS</b>			
SEXO Feminino	1,479	1,479	1,479
Ausentes SEXO Feminino	0,656	0,655	0,655
COR Parda	0,867	0,867	0,867
COR Preta	0,867	0,65	0,65
COR Outra	0,747	0,746	0,747
NSE	1,177	1,177	1,177
Atraso escolar	0,971	0,964	0,97
Série	2,292	2,293	2,292
Ano 2009	1,461	1,461	1,461
Ano 2011	1,627	1,627	1,627
Ano 2013	1,963	1,963	1,963
<b>NÍVEL 2 - ESCOLAS</b>			
Intercepto	0,175	0,176	0,175
NSE da escola	1,58	1,58	1,58
Proporção de alunos com atraso	0,947	0,948	0,948
Intervenção para melhorias	1,125	1,117	1,125
Currículo	1,091	1,091	1,09
Intervenção para melhorias*Atraso	-	1,017	-
Currículo*Atraso	-	-	1,001

Fonte: Elaboração das autoras com dados da Prova Brasil de 2007 a 2013.

Nota: Todos os coeficientes são estatisticamente significativos (p-valor inferior a 0,001), exceto os coeficientes de ambas as interações relativas à categoria 2 (adequado/avançado).

Os resultados do modelo 2 corroboram o argumento de Crahay (2007) de que a repetência é ineficaz para ajudar alunos com dificuldade e que a escola deve ter ações para evitar isso. Os nossos achados mostram que as ações sintetizadas no indicador intervenção para melhoria são relevantes para o aluno com atraso no nível abaixo do básico, mas o mesmo não se observa no nível adequado. Dito de outra forma, em situações de muitas desvantagens de aprendizado, as intervenções são capazes de produzir alguma melhoria, mas, nos níveis mais altos de aprendizado, elas não são suficientes para ultrapassar certas barreiras. É como se aluno com atraso tivesse menos prontidão, dado seu *background* e defasagens acumuladas, para se beneficiar dessas práticas que, a princípio, o têm como foco principal.

Esse resultado corrobora achados de outras pesquisas que indicam que em escolas com nível socioeconômico mais baixo, nas quais, em geral, há mais alunos com atraso, as intervenções focam em habilidades básicas, visando a evitar que o quadro se torne irreversível (TEDDLIE; STRINGFIELD, 1993).

Em relação ao modelo 3, os nossos resultados são compatíveis com a discussão de Dubet (2004) sobre justiça social com meritocracia. Quando o currículo é adequado e cumprido durante o ano letivo, a escola evita que os alunos com atraso escolar tenham mais chances de serem abandonados sem acesso à cultura escolar; no nosso caso, a falta de acesso está expressa no nível abaixo do básico. Porém, as práticas reunidas no indicador não apresentam o mesmo efeito equalizador entre alunos com e sem atraso quanto às chances de permanecerem no nível adequado. Ou seja, para o nível mais alto de aprendizado, a vantagem do aluno sem atraso prevalece.

## DISCUSSÃO

Este trabalho teve como objetivo propor dois indicadores (intervenção para melhoria e currículo na escola) construídos a partir dos questionários do Saeb e apresentar a sua utilidade para descrever as desigualdades de aprendizado nas escolas públicas de ensino fundamental. Foram considerados vários

aspectos que devem ser levados em conta para a construção e uso de indicadores, segundo Saltelli e colaboradores (2004), quais sejam: consistência teórica, normalização, seleção dos dados, método de estimação e dados ausentes, ponderação e agregação, robustez e sensibilidade, correlação com outras variáveis e indicadores, visualização e volta aos dados reais (decomposição do indicador para propósitos analíticos).

A construção dos indicadores para mensurar conceitos complexos a partir de instrumentos que não foram planejados para esse fim constitui um grande desafio. Dado esse fato, o ponto de partida foi fundamentar teoricamente os construtos nas pesquisas em eficácia escolar e sociologia da educação visando à delimitação dos conceitos. Em seguida, foram definidos os dados que seriam mais adequados para mensurar os construtos e o modelo para a estimação. Por fim, buscou-se validar os indicadores estimados, relacionando-os aos resultados escolares dos alunos.

Quanto aos dados, as informações dos questionários contextuais do Saeb representam um grande espectro das escolas brasileiras públicas e privadas ao longo de muitos anos. Para esta pesquisa, trabalhou-se com as informações de 2007 a 2013 de escolas públicas e privadas do ensino fundamental. Esses dados representam, em cada ano, milhares de escolas. Esse período de tempo foi escolhido porque os dados da Prova Brasil, que é uma avaliação censitária, estão disponíveis a partir de 2007. Como são dados públicos, eles já foram amplamente utilizados e demonstram bastante consistência para a caracterização das escolas e alunos brasileiros.

Para cálculo dos indicadores, utilizou-se um modelo da TRI adequado para variáveis ordinais. Na estimação dos indicadores pela TRI, são considerados apenas os itens para os quais os indivíduos forneceram respostas para se estimar o valor do escore de interesse. Essa é uma vantagem importante da TRI em relação aos métodos convencionais, tal como a análise fatorial, e bastante adequada para o problema desta pesquisa, que contém muitos dados incompletos. Vale ressaltar também que a metodologia da TRI atribui empiricamente pesos aos itens do indicador segundo os parâmetros dos itens.

Como resultado dessa abordagem, a escala obtida é comparável ao longo das edições do Saeb. Isso é possível porque os questionários foram compatibilizados e há questões iguais que permitem equalizar as edições via TRI.

Para a validação dos indicadores, foram construídos dois tipos de análise. Primeiro, observou-se a distribuição das médias de cada um deles segundo os níveis de aprendizado em Leitura e Matemática por edições da Prova Brasil e série. Os padrões descritos mostraram que os escores mais altos nos indicadores correspondem aos níveis mais elevados de aprendizado.

Em seguida, foram estimados modelos estatísticos para testar hipóteses sobre os efeitos dos indicadores nas chances de aprendizado e a interação entre eles e a variável atraso escolar dos alunos. Os resultados dos modelos mostraram que as escolas que possuem escores mais altos nos dois indicadores contribuem para reduzir a diferença entre alunos com e sem atraso escolar quanto às chances de eles não estarem no nível abaixo do básico de aprendizado, no qual os alunos com atraso constituem a maioria. Não se observou o mesmo efeito equalizador dos indicadores no nível mais elevado de aprendizado.

Importante salientar que os efeitos dos indicadores foram estimados por um modelo rigoroso, com controle de variáveis de *background* e de caracterização das condições da escola. Por essa razão, as magnitudes dos coeficientes não são muito elevadas, mas são significativas e com efeitos não desprezíveis. Por exemplo, um aluno que estuda em uma escola que possui 10 pontos na escala de intervenções teria 51% (5,1% vezes 10 pontos) de chances a menos de estar no nível abaixo do básico e 125% de chances a mais de estar no nível adequado. Também é importante destacar que os efeitos devem ser analisados como fatores associados às chances de aprendizado, porque não se trata de uma pesquisa experimental que visa à verificação de uma relação de causalidade.

Os indicadores estimados são determinados empiricamente, ou seja, dependem dos dados que foram selecionados para o cálculo. Eles são úteis quando aplicados em modelos estatísticos para análise de resultados escolares. Eles não são

indicadores sintéticos amplos, tais como o Índice de Desenvolvimento Humano (IDH) e o Índice de Desenvolvimento da Educação Básica, que servem para comparar lugares e evolução, independentemente do número de casos utilizados.

Os indicadores propostos são adequados para a redução de dimensões complexas. Isso não quer dizer que, observada uma dada relação entre eles e um resultado educacional, não seja possível analisar, de modo mais detalhado, os efeitos dos itens que os compõem sobre o fenômeno em foco.

Por fim, os indicadores não devem ser entendidos como mecanismos determinísticos na produção de bons resultados. Isso porque, muito provavelmente, ações no sentido de mudar um indicador na escola podem forçar a mudança de outros. Mesmo que uma escola adote certas práticas que foram apontadas como de forte influência no aprendizado, isso não quer dizer que ela obterá os mesmos resultados, porque depende da configuração de vários fatores do contexto da escola.

## **AGRADECIMENTOS**

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processo 446638/2014-5), à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) (APQ-02333-14 e APQ-02544-14) e à Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco – Representação Brasil) pelo apoio às pesquisas realizadas no Núcleo de Pesquisa em Desigualdades Escolares.

## **REFERÊNCIAS**

ALVES, M. T. G.; FRANCO, C. A pesquisa em eficácia escolar no Brasil: evidências sobre o efeito das escolas e fatores associados à eficácia escolar. In: BROOKE, N.; SOARES, J. F. (Ed.). *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte: Editora UFMG, 2008. p. 482-500.

ALVES, M. T. G.; SOARES, J. F. Medidas de nível socioeconômico em pesquisas sociais: uma aplicação aos dados de uma pesquisa educacional. *Opinião Pública*, Campinas, SP, v. 15, p. 1-30, 2009.

ALVES, M. T. G.; SOARES, J. F.; XAVIER, F. P. Índice socioeconômico das escolas de educação básica brasileiras. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 22, n. 84, p. 671-703, 2014.

ALVES, M. T. G.; XAVIER, F. P. *Desigualdades de aprendizado entre alunos das públicas brasileiras: evidências da Prova Brasil (2007 a 2013)*. Brasília, DF: MEC/Inep; Unesco. No prelo.

BONAMINO, A.; SOUSA, S. Z. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. *Educação e Pesquisa*, São Paulo, v. 38, n. 2, p. 373-388, 2012.

BRASIL. *Lei n. 13.005*, de 25 de junho de 2014, que aprova o Plano Nacional de Educação e dá outras providências. 2014. Disponível em: <[http://www.planalto.gov.br/CCIVIL\\_03/\\_Ato2011-2014/2014/Lei/L13005.htm](http://www.planalto.gov.br/CCIVIL_03/_Ato2011-2014/2014/Lei/L13005.htm)>. Acesso em: fev. 2016.

BROOKE, N.; CUNHA, M. A. A avaliação externa como instrumento da gestão educacional nos estados. *Estudos & Pesquisas Educacionais*, São Paulo, v. 2, 2011.

BROOKE, N.; SOARES, J. F. O que faz a diferença? Os métodos e evidências da pesquisa sobre efeito escola: comentários. In: BROOKE, N.; SOARES, J. F. (Ed.). *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte: Editora UFMG, 2008. Seção 3, p. 218-224.

CRAHAY, M. Qual pedagogia para aos alunos em dificuldade escolar? *Cadernos de Pesquisa*, São Paulo, v. 37, n. 130, p. 181-208, jan./abr. 2007.

DUBET, F. O que é uma escola justa? *Cadernos de Pesquisa*, São Paulo, v. 34, p. 539-555, 2004.

FRANCO, C. et al. O referencial teórico na construção dos questionários contextuais do Saeb 2001. *Estudos em Avaliação Educacional*, São Paulo, n. 28, p. 39-74, jul./dez. 2003.

GOLDSTEIN, H. 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.

HAIR, J. F. et al. *Análise multivariada de dados*. Porto Alegre: Bookman, 2005.

HAMBLETON, R. K. Principles and selected applications of Item Response Theory. In: LINN, R. L. (Ed.). *Educational measurement*. 3. ed. Washington, DC: American Council on Education; National Council on Measurement in Education, 1993. p. 147-200.

HERMAN, J. L.; BAKER, E. L. Assessment Policy: making sense of the babel. In: SYKES, G.; SCHENEIDER, B.; PLANK, D. N. *Handbook of educational policy research*. New York: American Educational Research Association; Routledge, 2009. p. 176-190.

JACCARD, J. *Interaction effects in logistic regression*. Thousand Oaks, CA: Sage, 2001.

JANNUZZI, P. M. Considerações sobre o uso, mau uso e abuso dos indicadores sociais na formulação e avaliação de políticas públicas municipais. *Revista de Administração Pública*, Rio de Janeiro, v. 36, n. 1, p. 51-72, 2002.

KARINO, C. A.; VINHA, L. G. A.; LAROS, J. A. Os questionários do SAEB: o que eles realmente medem? *Estudos em Avaliação Educacional*, São Paulo, v. 25, n. 59, p. 270-297, set./dez. 2014.

MADAUS, G. F.; AIRASIAN, P. W.; KELLAGHAN, T. *School effectiveness: a reassessment of the evidence*. New York: McGraw-Hill Book, 1980.

MOREIRA, A. F. B. Currículo: concepções, políticas e teorizações. In: OLIVEIRA, D. A.; DUARTE, A. M. C.; VIEIRA, L. M. F. *Dicionário: trabalho, profissão e condição docente*. Belo Horizonte: UFMG/Faculdade de Educação, 2010. CDROM.

MORTIMORE, P. et al. *School matters*. California: University of California, 1988.

RAUDENBUSH, S. W.; BRYK, A. S. *Hierarchical linear models: applications and data analysis methods*. 2nd ed. Newbury Park, California: Sage, 2002.

RAUDENBUSH, S. W. et al. *HLM 7: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International, 2011.

REYNOLDS, D.; TEDDLIE, C. Os processos da eficácia escolar. In: BROOKE, N.; SOARES, J. F. (Ed.). *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte: Editora UFMG, 2008. p. 297-328.

RIBEIRO, S. C. A pedagogia da repetência. *Estudos Avançados*, São Paulo, v. 12, n. 5, p. 7-21, 1991.

SALTELLI, A. et al. *Composite indicators: the controversy and the way forward*. OCDE. Palermo, 11-14 November, 2004, Disponível em: <[www.oecd.org/site/worldforum/33841312.doc](http://www.oecd.org/site/worldforum/33841312.doc)>. Acesso em: 29 maio 2016.

SAMEJIMA, F. *Estimation of latent ability using a response pattern of graded responses*. Richmond, VA: Psychometric Society, 1969. (Psychometric Monograph, n. 17).

SAMMONS, P. *School effectiveness: coming of age in the twenty-first century*. Lisse: Swets & Zeitlinger, 1999.

SOARES, J. F. Índice de Desenvolvimento da Educação de São Paulo: IDESP. *São Paulo em Perspectiva*, São Paulo, v. 23, n. 1, p. 29-41, 2009.

SOARES, J. F.; ANDRADE, R. J. Nível socioeconômico, qualidade e equidade das escolas de Belo Horizonte. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 14, n. 50, p. 107-126, 2006.

SOARES, S. Avaliação educacional como instrumento pedagógico. *Trabalho e Sociedade*, v. 2, n. 4, p. 23-25, 2002.

SOARES, T. M. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. *Pesquisa Operacional*, Rio de Janeiro, v. 25, n. 1, p. 83-112, 2005.

SOARES NETO, J. J. et al. Uma escala para medir a infraestrutura escolar. *Estudos em Avaliação Educacional*, São Paulo, v. 24, n. 54, p. 78-99, jan./abr. 2013.

TEDDLIE, C.; STRINGFIELD, S. *Schools make a difference: lessons learned from a 10-year study of school effects*. New York: Teachers College, 1993.

YOUNG, M. O futuro da educação em uma sociedade do conhecimento: o argumento radical em defesa de um currículo centrado em disciplinas. *Revista Brasileira de Educação*, Rio de Janeiro, v. 16, n. 48, p. 609-623, 2011.

---

**MARIA TERESA GONZAGA ALVES**

Professora adjunta do Departamento de Ciências Aplicadas à Educação da Faculdade de Educação da Universidade Federal de Minas Gerais (DECAE/FaE/UFMG)  
*mtga@ufmg.br*

**FLAVIA PEREIRA XAVIER**

Professora adjunta do Departamento de Ciências Aplicadas à Educação da Faculdade de Educação da Universidade Federal de Minas Gerais (DECAE/FaE/UFMG)  
*flaviapx@ufmg.br*

## APÊNDICE A

**TABELA A1 – Distribuição de frequência da variável atraso escolar pelos níveis de aprendizado em leitura do 5º e 9º anos, para cada edição da Prova Brasil**

ANO ESCOLAR	EDIÇÃO	NÍVEIS DE APRENDIZADO	ATRASO ESCOLAR		
			SEM ATRASO	1 OU MAIS ANOS DE ATRASO	DIFERENÇA (SEM ATRASO-COM ATRASO)
5º ano	2007	Abaixo do básico	22,5%	41,8%	-19,3%
		Básico	44,8%	45,1%	-0,3%
		Adequado/Avançado	32,7%	13,1%	19,6%
	2009	Abaixo do básico	18,7%	36,9%	-18,2%
		Básico	41,4%	44,5%	-3,1%
		Adequado/Avançado	39,8%	18,5%	21,3%
	2011	Abaixo do básico	17,7%	36,4%	-18,7%
		Básico	39,0%	44,3%	-5,2%
		Adequado/Avançado	43,3%	19,4%	23,9%
	2013	Abaixo do básico	16,3%	38,8%	-22,5%
		Básico	34,1%	41,6%	-7,5%
		Adequado/Avançado	49,6%	19,6%	30,0%
9º ano	2007	Abaixo do básico	21,1%	39,3%	-18,2%
		Básico	58,8%	53,9%	4,9%
		Adequado/Avançado	20,1%	6,9%	13,3%
	2009	Abaixo do básico	16,6%	32,0%	-15,3%
		Básico	56,5%	56,1%	0,4%
		Adequado/Avançado	26,9%	11,9%	14,9%
	2011	Abaixo do básico	16,9%	33,4%	-16,4%
		Básico	56,1%	54,6%	1,5%
		Adequado/Avançado	26,9%	12,0%	14,9%
	2013	Abaixo do básico	19,2%	35,1%	-15,8%
		Básico	52,2%	51,9%	0,2%
		Adequado/Avançado	28,6%	13,0%	15,6%

Fonte: Elaboração das autoras com dados da Prova Brasil de 2007 a 2013.

Recebido em: JUNHO 2016

Aprovado para publicação em: NOVEMBRO 2016