



Ensaio: Avaliação e Políticas Públicas em Educação

ISSN: 0104-4036

ensaio@cesgranrio.org.br

Fundação Cesgranrio  
Brasil

Gonzaga Alves, Maria Teresa; Soares, José Francisco; Pereira Xavier, Flavia  
Índice Socioeconômico das Escolas de Educação Básica Brasileiras  
Ensaio: Avaliação e Políticas Públicas em Educação, vol. 22, núm. 84, julio-septiembre, 2014, pp. 733-753  
Fundação Cesgranrio  
Rio de Janeiro, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=399534055007>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica  
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal  
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

# Índice Socioeconômico das Escolas de Educação Básica Brasileiras

Maria Teresa Gonzaga Alves\*

José Francisco Soares\*\*

Flavia Pereira Xavier\*\*\*

## Resumo

Este artigo apresenta a metodologia e os resultados do desenvolvimento de um índice de nível socioeconômico das escolas de educação básica do Brasil. Os dados provêm dos questionários contextuais aos quais os alunos respondam nas avaliações educacionais feitas pelo governo federal nesse nível de ensino. Foram consideradas as respostas válidas de 20.806.062 alunos em 21 bases de dados. Para estimar o índice, itens relacionados às dimensões escolaridade e ocupação dos pais do aluno e a renda familiar foram agregados, empregando-se um modelo da Teoria da Resposta ao Item. Os resultados foram validados mostrando-se fidedignos. Correlações com outros índices semelhantes são positivas e altas. A correlação com a renda *per capita* dos municípios revela que a realidade nacional vista por um indicador econômico e um indicador obtido pelas respostas a um questionário é, no nível macro, a mesma. A avaliação de especialistas em diferentes localidades do país coincide com a retratada pelo indicador.

**Palavras-chave:** Nível socioeconômico das escolas. Educação básica. Desigualdades educacionais.

## 1 Introdução<sup>1</sup>

Este artigo apresenta a metodologia usada e os resultados do desenvolvimento de um índice de nível socioeconômico (NSE) para as escolas de educação básica do Brasil. Esse índice faz parte de uma pesquisa mais ampla que visa a produzir

\* Professora Adjunta da Faculdade de Educação Universidade Federal de Minas Gerais (UFMG) e líder do Núcleo de Pesquisas em Desigualdades Escolares (Nupede). E-mail: mtga@ufmg.br

\*\* Professor Titular aposentado da Faculdade de Educação e pesquisador do Grupo de Avaliação e Medidas Educacionais (Game) da Universidade Federal de Minas Gerais (UFMG). E-mail: francisco-soares@ufmg.br

\*\*\* Professora Adjunta da Faculdade de Educação Universidade Federal de Minas Gerais (UFMG) e vice-líder do Núcleo de Pesquisas em Desigualdades Escolares (Nupede). E-mail: flaviapx@ufmg.br

1 Versões preliminares da pesquisa que deu origem a esse artigo foram apresentadas no Encontro da Associação de Pós-Graduação em Ciências Sociais (Anpocs 2012) e no Encontro da Associação Brasileira de Avaliação Educacional (Abave 2013). O presente trabalho inclui dados mais atualizados, ajustes nos procedimentos metodológicos, além de novas análises e resultados.

indicadores contextuais para pesquisas sobre a educação brasileira. Neste trabalho, foram utilizados os dados dos questionários contextuais aos quais respondem os estudantes que fizeram os testes das avaliações da educação básica organizadas pelo Instituto Nacional de Estudo e Pesquisas Educacionais Anísio Teixeira (Inep), que incluem um amplo espectro de escolas públicas e privadas.

A construção do índice de NSE das escolas é uma contribuição a uma demanda que surge naturalmente, nos últimos anos, a partir da divulgação dos resultados educacionais por estabelecimentos de ensino. As análises desses resultados sem a consideração do NSE geram associações enviesadas, quando se sabe, com base em ampla evidência empírica, que há forte correlação entre esses dois fenômenos (ALVES; SOARES, 2013; COLEMAN et al., 1966; JENCKS et al., 1972; SOARES; COLLARES, 2006). Pela extensão dos dados utilizados, e pela solidez e flexibilidade da metodologia estatística empregadas, o NSE desenvolvido neste trabalho configura-se como uma medida válida e confiável que poderá ser usada em outros âmbitos da pesquisa social.

O artigo está organizado da seguinte forma: na próxima seção será revista, brevemente, a literatura que justifica as dimensões incluídas para mensurar o NSE das escolas. Em seguida, na metodologia, serão apresentadas as bases de dados utilizadas, a preparação das mesmas, os critérios para a seleção e críticas dos itens dos questionários contextuais, a montagem do banco de dados e os procedimentos estatísticos para a estimação do NSE. Na seção dos resultados serão apresentadas as análises descritivas do índice de NSE e a sua validação. Nas considerações finais, serão discutidos as aplicações e os limites do índice de NSE calculado por escola.

## **2 Fundamentação teórica**

O nível socioeconômico é um construto teórico que aloca os indivíduos em classes ou estratos sociais. Na literatura, não há uma definição unívoca desse construto, nem um consenso absoluto sobre quais dimensões devem ser consideradas para a sua operacionalização. As decisões dos pesquisadores dependem de justificativas teóricas, e também da disponibilidade de dados empíricos que expressem as dimensões importantes do construto. No entanto, as dimensões de ocupação, rendimento e nível educacional estão incluídas na maioria das vezes e, por isso, são as consideradas neste estudo.

O estudo das diferenças sociais ocupa posição central na literatura sociológica contemporânea. Embora o interesse pelo tema tenha raízes na sociologia clássica, foi nos Estados Unidos, a partir dos anos 1950, no âmbito das pesquisas em estratificação e mobilidade sociais, que as medidas empíricas de nível socioeconômico passaram a ser operacionalizadas, como, por exemplo, nos importantes trabalhos de Duncan (1963) e Blau e Duncan (1967). Essas pesquisas tiveram impulso graças aos avanços metodológicos propiciados pela aplicação da estatística com o auxílio computacional aos dados populacionais – censitários ou por amostragem –, o que também ocorreu no Brasil nas décadas seguintes (PASTORE, 1979; PASTORE; SILVA, 2000; SANTOS, 2005; SCALON, 1998).

Na pesquisa sociológica, os indicadores utilizados são quase sempre relacionados à posição ocupacional (ERICKSON; GOLDTHORPE, 1992; GAZENBOOM; DE GRAAF; TRAIMAN, 1992). Entretanto, a operacionalização de indicadores relacionados à posição ocupacional dos indivíduos é bastante complexa. A identificação precisa da posição ocupacional depende do conhecimento do título da ocupação (por exemplo, professora) e das atividades executadas na ocupação (por exemplo, ministra aulas no ensino fundamental). Adicionalmente, informações como o nível na hierarquia (direção, supervisão, por exemplo) e o setor econômico (público ou privado) são também requeridas. Assim, o tempo e recursos exigidos para a análise das informações e para a codificação das ocupações restringem esse modelo a poucos estudos (HAUSER; WARREN, 1997).

Um indicador que usa dados de ocupação foi desenvolvido para o Programa Internacional de Avaliação de estudantes, cuja sigla é PISA (*Programme for International Student Assessment*) (GAZENBOOM; DE GRAAF; TRAIMAN, 1992; GAZENBOOM, 2010). Esse indicador, chamado de ISEI (*International Socio-Economic Index of Occupational Status*), é estimado a partir de informações sobre a ocupação dos pais dos alunos, a qual é atribuída um escore escalonado de acordo com a educação e a renda da família medida de forma indireta (posse de bens). O ISEI é utilizado para contextualizar os resultados comparativos entre países, exatamente para evitar análises enviesadas sem o controle do NSE dos alunos que fazem o teste e as diferenças locais. Outra referência interessante sobre o uso da ocupação dos pais no índice de NSE vem da Austrália, onde ele é utilizado para contextualizar a comparação dos resultados educacionais entre as diferentes regiões do país (MARKS et al., 2000).

Nas pesquisas educacionais o debate sobre classes ou estrados sociais é também um tema central, devido à forte correlação entre resultados escolares e o nível socioeconômico e cultural das famílias, comprovada por ampla evidência empírica em vários países do mundo. O Relatório Coleman (COLEMAN et al., 1966), é considerado um dos textos mais importantes sobre a relação entre *background* socioeconômico dos alunos e contexto social das escolas com o desempenho escolar. A dimensão cultural é também reconhecida como um elemento fundamental para caracterização social das famílias, principalmente a partir dos influentes trabalhos de Pierre Bourdieu e seguidores (BOURDIEU; PASSERON, 2011). Este importante construto não será, entretanto, tratado neste texto.

As pesquisas aplicadas também têm contribuições relevantes para a definição de índices de nível socioeconômicos dos indivíduos ou das famílias. Em estudo recentemente publicado, o economista Marcelo Neri (2012) analisa as mudanças sociais do Brasil, descrevendo a evolução das classes econômicas, denominadas como classes A, B, C, D ou E. A classificação é feita pela agregação dos indivíduos segundo a renda oriunda de todas as fontes (salários, aluguéis, pensões, etc.) e o uso que se faz dela, por meio do consumo de bens, serviços e o acesso a alguns ativos de produção, como a educação, a carteira de trabalho, a internet e outros. Estas informações são obtidas na Pesquisa Nacional por Amostra de Domicílios (Pnad) realizada pelo Instituto Nacional de Geografia e Estatística (IBGE).

O Critério de Classificação Econômica Brasil (ou simplesmente Critério Brasil), adotado pela Associação Brasileira de Empresas de Pesquisa (Abep) para estimar o poder de compra das famílias urbanas, tem princípios bastante semelhantes, mas operacionalmente, é bem mais simples<sup>2</sup>. A sua aplicação depende de informações que a maioria das pessoas consegue responder, tais como os bens e serviços disponíveis no domicílio e a escolaridade do chefe da família. A sua fórmula de cálculo é bem fácil, pois envolve apenas uma soma de pontos para situar os indivíduos nas classes A1, A2, B1, B2, C1, C2, D e E. Os resultados são considerados bastante fidedignos quanto ao poder de compra das famílias urbanas com rendimento de até cerca de

2 O Critério Brasil é desenvolvido a partir das informações coletadas pelo Levantamento Socioeconômico realizado anualmente pelo IBOPE, por meio de uma pesquisa probabilística de domicílios nas dez principais regiões metropolitanas do país. Informações em: <http://www.abep.org/novo/Content.aspx?SectionID=84>, consultado em 2 de abril de 2013.

trinta salários mínimos, o que compreende a maioria absoluta da população. Por esses motivos, o Critério Brasil tem ampla aplicação nas pesquisas de mercado e é adotado também por outros estudos que necessitam segmentar os indivíduos para estudar temas variados, como o comportamento eleitoral, o acesso aos serviços públicos, etc.

Um ponto específico da pesquisa educacional é que os dados para a construção de indicadores são obtidos com os questionários aos quais geralmente os próprios alunos responderam. Por isso, a definição dos indicadores importantes, a formulação dos itens com clareza e precisão, a extensão dos questionários, a adequação à faixa etária e a possibilidade de itens total ou parcialmente não terem tido respostas, são elementos críticos a serem considerados. Os pesquisadores necessitam de informações relevantes e fidedignas, o que nem sempre estão disponíveis da forma ideal (ALVES; SOARES, 2009; FRANCO et al, 2003; SOARES; ANDRADE, 2006).

Tendo em vista os referenciais teóricos, os dados educacionais disponíveis, assume-se, neste trabalho, que o nível socioeconômico é um construto latente, ou seja, não diretamente observável, e que sintetiza unidimensionalmente a relação entre a escolaridade, a ocupação e a renda das famílias. Entretanto, estas variáveis não estão diretamente contempladas em todos os questionários contextuais das avaliações educacionais do Inep. Assim, o desenvolvimento do índice de NSE exigiu decisões metodológicas para operacionalizar o construto que serão descritas a seguir.

## 3 Metodologia

### 3.1 Os dados

Neste trabalho foram utilizados os microdados das avaliações educacionais da educação conduzidas pelo Inep<sup>3</sup>. Foram consideradas as seguintes bases de dados: a Prova Brasil dos ciclos referentes aos anos de 2005, 2007, 2009 e 2011; o Sistema de Avaliação da Educação Básica (Saeb) de 2003, 2005 e 2011; e Exame Nacional do Ensino Médio (Enem) realizado nos anos de 2007, 2008, 2009 e 2011. Os estudantes que respondem aos testes dessas avaliações preenchem questionários contextuais que permitem a construção de indicadores relacionados ao construto de interesse desta pesquisa.

3 Os microdados foram obtidos via download no site do Inep, exceto os dados do Saeb e da Prova Brasil 2005, que foram cedidos ao Grupo de Avaliação e Medidas Educacionais (Game) para propósitos de pesquisa.

Estas bases de dados, embora não constituindo amostras aleatórias, têm representatividade bastante ampla das escolas de educação básica brasileiras, pois compreendem um amplo espectro de perfis de alunos. Cada uma das modalidades de avaliação contribui com informações específicas que, no conjunto, compõem um quadro bem rico dos tipos de escolas espalhadas pelo país.

O Saeb avalia os sistemas educacionais por meio de uma amostra representativa de todas as regiões brasileiras e unidades da federação, bem como das escolas urbanas e rurais e das redes pública e privada<sup>4</sup>. Os testes e questionários contextuais são aplicados ao final das etapas da educação básica, ou seja, no 5º e 9º anos (antes nomeadas, respectivamente, 4ª e 8ª série) do ensino fundamental e 3º ano do ensino médio.

A inclusão do Saeb é importante por que as suas respectivas amostras constituem evidências estatisticamente representativas de todos os tipos de escolas. Os dados do Saeb são especialmente importantes no processo de estimação dos parâmetros dos itens pois incluem dados de escolas públicas e privadas, o que permite colocar as escolas destas duas redes em uma mesma escala. No entanto, as escolas do Saeb não são identificadas e, assim sendo, a medida do nível socioeconômico da escola não é calculada, na etapa final deste trabalho para as escolas incluídas nessa avaliação.

A Prova Brasil foi implementada em 2005 para avaliar, a cada dois anos, as escolas públicas que oferecem o Ensino Fundamental e que tenham pelo menos 20 alunos na série avaliada (5º e 9º anos). Inicialmente participavam apenas as escolas urbanas, mas, desde a edição de 2009, as escolas rurais que atendem ao critério do número de alunos mínimos passaram a fazer parte da Prova Brasil. As bases de dados da Prova Brasil contêm os códigos de identificação das escolas e, assim sendo, os nomes e a localização de cada escola puderam ser consultados no sistema do Inep na etapa final deste trabalho. A cobertura da Prova Brasil ultrapassa a 85% das escolas que oferecem o Ensino Fundamental.

O Enem é um exame anual e voluntário, cujo foco é o aluno concluinte do ensino médio. Mas qualquer pessoa pode inscrever-se, uma vez que o

---

4 Desde 2005, a sigla Saeb se refere a dois processos de avaliação: a Avaliação Nacional da Educação Básica (Aneb), que mantém a mesma característica e objetivos do Saeb, e a Avaliação Nacional do Rendimento Escolar (Anresc), mais conhecida como Prova Brasil. Neste trabalho, optou-se por manter as denominações mais conhecidas dessas avaliações, visto que as novas, introduzidas em 2005, não são utilizadas nem mesmo no site do Inep.

resultado no Enem tem várias aplicações para o examinado, como participar de processos seletivos para universidades públicas e privadas, certificação do ensino médio e outras<sup>5</sup>. Isso tem estimulado a adesão crescente ao Enem, tornando-o bastante representativo das escolas de ensino médio das redes pública e privada em todo o país.

Como o Enem atrai um público muito diversificado, neste trabalho, cujo objetivo é criar um indicador de nível socioeconômico para escolas de educação básica, foram considerados somente os dados dos alunos que terminaram o ensino médio no ano de aplicação do Enem. A base de dados do Enem 2010 não foi utilizada porque houve uma mudança profunda no questionário contextual aplicado nessa edição, o que o tornou completamente atípico em comparação a todas as outras bases de dados. Em 2011, o questionário retornou ao formato anterior em relação aos itens utilizados neste estudo.

No total, foram selecionadas vinte e uma bases de dados, visto que cada edição do Saeb e da Prova Brasil está organizada em bancos de dados separados, que correspondem a três e duas séries escolares, respectivamente. Os microdados do Enem estão em bancos de dados únicos, dos quais são excluídos os alunos que não são concluintes no ano da avaliação.

Em avaliações educacionais é comum uma parte dos alunos que fazem o teste deixarem o questionário totalmente em branco. Esses casos foram excluídos, assim como aqueles que não responderam à maioria dos itens de interesse em cada uma das bases de dados. O ponto de corte, isto é, o mínimo de respostas em cada base de dados, foi definido empiricamente, observando-se o número de itens considerados e o percentual de casos que faltaram em cada situação.

O Quadro 1, no Apêndice, apresenta os números de alunos e de escolas considerados em cada uma das bases de dados. No conjunto, os bancos de dados contêm 20.806.062 de casos, correspondentes aos alunos registrados nas bases de dados utilizadas neste estudo.

5 O Enem é utilizado como critério de seleção para concorrer à bolsa de estudos no programa federal Universidade para Todos (ProUni), serve para credenciar alunos interessados em específicos das universidades, e também é utilizado como uma forma de certificação de alunos concluintes do ensino médio em modalidades diferentes do ensino regular. Além disto, o Enem serve como um instrumento para verificação do aprendizado dos alunos que terminam o ensino médio, e os seus escores médios produzem uma "nota" das escolas, que é amplamente divulgada em forma de ranking pela imprensa.



### 3.2 Seleção e preparação das variáveis

Em cada banco de dados, todas as variáveis relacionadas de forma direta ou indireta à posição ocupacional e ao nível de instrução dos pais dos alunos e ao padrão de rendimento de suas famílias foram selecionadas para compor um banco de dados único para a estimação do nível socioeconômico das escolas.

Os questionários contextuais das várias avaliações são diferentes. Eles divergem em relação aos itens dos questionários, ao formato das perguntas e às opções de respostas, mesmo entre questionários que pertencem à mesma modalidade de avaliação. As variáveis dos bancos de dados tiveram, então, que ser recodificadas, garantido as mesmas categorias de respostas ao longo das diferentes bases.

O passo seguinte foi incluir, em cada um dos bancos de dados, variáveis correspondentes aos itens não contemplados nos respectivos questionários, que foram tratadas como “dado ausente” para todos os casos. Por exemplo, o item “renda familiar” está presente somente nos questionários do Enem. Para criar um banco de dados único com dados de todos os alunos, criou-se esta variável nos outros bancos de dados (Saeb e Prova Brasil) a qual foi atribuída o código de “dado ausente” para todos os casos. A ausência de informação em parte dos casos não inviabiliza o processo de estimação do NSE devido à flexibilidade do modelo de análise escolhido.

Esse tipo de tratamento foi adequado para a recodificação e padronização das variáveis em quase todas as situações, exceto para três itens da Prova Brasil de 2005. No questionário contextual desse ano, nos itens sobre o número de automóveis, banheiros, rádios e televisores há uma categoria ordinal a menos do que nos outros questionários. Na primeira versão desse trabalho, o questionário da Prova Brasil 2005 foi tomado como padrão para a recodificação dos demais bancos de dados. Isto é, os outros tiveram redução nas suas respectivas categorias. Entretanto, avaliou-se que essa opção penalizou o grau de informação desses itens para discriminar posições mais elevadas na escala estimada, exceto em relação ao item “rádio”. Após vários testes, optou-se por incluir as variáveis “Automóvel” “Banheiro” e “TV” com escala de 4 pontos comuns a todas as bases de dados, exceto para a base de dados da Prova Brasil 2005, na qual essas variáveis foram tratadas como dados ausentes, tal como se fez com as outras variáveis não

presentes em um ou mais bancos de dados. Para não perder as informações dessas variáveis na Prova Brasil 2005, nesse subconjunto de dados as variáveis “Auto3”, “Banheiro3” e “TV3” foram mantidas com escala de 3 pontos.

Ao final, foram selecionadas 24 variáveis – incluindo as subdivisões dos itens automóvel, televisão e banheiro – que estão descritas no Quadro 2. O Quadro 3 sintetiza a presença dessas variáveis nas bases de dados. Ambos estão no Apêndice.

### **3.3 Preparação do banco de dados final: perfis e ponderação**

Dentre os mais de 20 milhões de casos reunidos nas vinte e uma bases de dados, há alunos cujos padrões de resposta aos itens selecionados são absolutamente idênticos. Estes constituem perfis de alunos que podem ser agregados – reduzindo-se o número de casos na base de dados – e as suas ocorrências computadas e analisadas de forma ponderada. Neste processo, um perfil de aluno com muitas ocorrências tem um peso maior na estimação do que um perfil mais raro. A vantagem é tornar o processamento dos dados mais rápido.

Nas bases de dados da Prova Brasil e do Enem para cada perfil observado foi acrescido uma nova variável com a soma dos alunos correspondente àquela ocorrência. No caso do Saeb o processo foi diferente. Como o Saeb é uma avaliação amostral, a representatividade das escolas varia de acordo com o estrato amostral (i.e., escola urbano ou rural; região do país etc.). Nos bancos de dados do Saeb, a representatividade da unidade amostral (aluno) está registrada em uma variável de ponderação (peso). Assim, ao se computar as ocorrências de perfis de alunos que participaram do Saeb estas foram multiplicadas pelo peso dos alunos que estão nos respectivos perfis.

Para evitar distorções que ocorreriam com a agregação simples de bancos de dados que contêm números de alunos muito diferentes, fixou-se em um milhão o total de perfis para todas as bases. Naturalmente os valores obtidos anteriormente foram transformados proporcionalmente para esse valor

A agregação das 24 variáveis reduziu o banco de dados para 3.374.178 perfis únicos. Diante de cada um, registrou-se o número de ocorrências nas respectivas bases de dados pela soma simples ou ponderada do número de casos que se enquadravam em cada um deles. A soma das ocorrências na nova grandeza corresponde aos pesos utilizados por perfil no processo de estimação do índice de NSE.

### 3.4 Modelo de análise

O processo de estimação do construto NSE foi feito com o ajuste de um modelo da Teoria da Resposta ao Item (TRI). A TRI engloba uma série de modelos matemáticos, cujo objetivo principal é a obtenção de medidas de construtos latentes, baseadas na probabilidade de um indivíduo escolher uma determinada resposta a um item dicotômico e/ou ordinal (HAMBLETON, 1993). Uma vantagem da TRI, em relação aos métodos convencionais, é a possibilidade de estimar a medida do NSE mesmo com dados incompletos para algum dos indicadores, capacidade crucial para este estudo, cujos dados provêm de questionários com itens parcialmente em comum.

O modelo da TRI escolhido para a criação do índice de NSE é o especificado para respostas graduadas, desenvolvido por Samejima (1969). Trata-se de um modelo utilizado para estimação de um construto latente através de itens cujas opções de respostas são ordenadas. A expressão matemática deste modelo é:

$$P = (X_{ij} = k | \theta) = \frac{1}{1 + \exp[-a_i(\theta - b_{i,k-1})]} - \frac{1}{1 + \exp[-a_i(\theta - b_{i,k})]} \quad (1)$$

$X_{ij}$  é a resposta fornecida pelo  $j$ -ésimo indivíduo ao  $i$ -ésimo indicador ordinal;

$k = 1, 2, \dots, m$  representa cada um dos níveis de  $X_i$ , sendo  $m$  a categoria associada ao maior valor de  $\theta$ ;

$\theta_j$  é a medida do indicador para o  $j$ -ésimo indivíduo;

$a_i$  é a inclinação e  $b_{i,k}$  é o ponto no eixo  $\theta$  em que a probabilidade de resposta ao  $i$ -ésimo indicador ser maior ou igual a  $k$  é maior do que 50%.

O modelo Samejima pode ser reescrito, de forma sucinta, como:

$$Pr(X_{ij} = K) = P(K) - P(K+1), \quad (2)$$

onde  $P(K)$  se refere à probabilidade de que a resposta ao  $i$ -ésimo indicador seja igual ou superior a  $K$ , para cada valor de  $\theta$ .

O símbolo theta ( $\theta$ ) representa o valor da variável latente, i.e., um escore que corresponde à medida do nível socioeconômico referente a cada perfil. Este escore é estimado pelo modelo e é tão maior quanto maiores forem os valores atribuídos às respostas dos itens. O valor de theta varia tipicamente no intervalo entre -3 a 3, ou seja, expressa desvios-padrão em relação à média.

É importante ressaltar que, devido ao método utilizado, são considerados apenas os itens para os quais os indivíduos forneceram resposta, desconsiderando-se aqueles em branco. Ou seja, considerou-se não apenas a resposta ao item, mas também quantos casos (ou seja, a frequência ponderada dos perfis) responderam daquela forma. Com a correção dos perfis para a ordem de um milhão, realizada na etapa final de preparação da base de dados, alguns perfis muito raros ficaram reduzidos a valores ínfimos. Estes foram mantidos na base de dados, embora tenham contribuído muito pouco para o processo de estimação dos parâmetros do modelo.

### 3.5 Análise dos itens

O poder explicativo das variáveis selecionadas dentro do construto NSE pode ser validado por meio de procedimentos estatísticos. Para isto, foi estimada a matriz policórica e analisados os autovalores e os autovetores da matriz. Em seguida, a partir da estimação dos parâmetros dos itens utilizados no cálculo do construto, foram analisadas as curvas de característica do item e as curvas de informação do item.

A matriz policórica descreve a correção entre variáveis ordinais. Ela foi estimada com algoritmo implementado no software R. O resultado mostra que correlação entre as variáveis escolhidas para compor o NSE é sempre positiva, ainda que algumas com valor muito baixo. A análise de componentes principais indica que há um traço latente dominante, pois o primeiro autovalor é mais de três vezes maior que o segundo (10,97 e 3,08, respectivamente). Esses resultados podem ser solicitados aos autores.

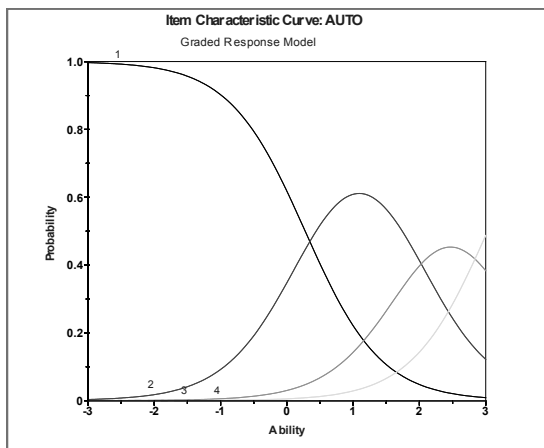
A análise de componentes principais indica que há um traço latente dominante, pois o segundo autovalor é três vezes menos do que o primeiro. Esses resultados, bem como os autovetores, podem ser solicitados aos autores.

Antes de estimar os escores finais, que definem a posição na escala do nível socioeconômico para cada um dos perfis, é preciso ainda analisar a qualidade de itens como componentes da medida característica latente por meio de suas curvas características e curvas de informação.

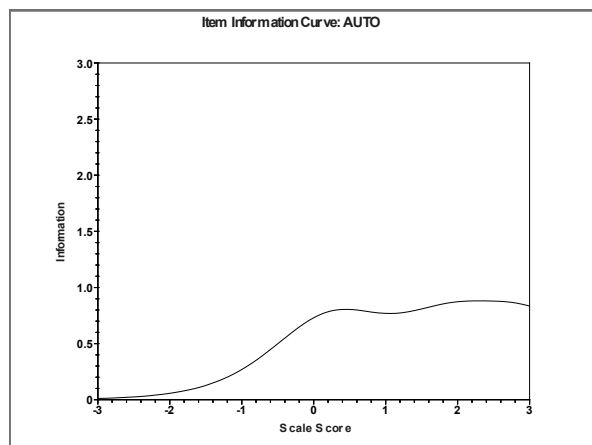
Isto foi feito para todos os itens. Como exemplo, o gráfico 1 apresenta a curva da variável “automóvel” (um item do nível de rendimento da familiar), que possui quatro categorias ordinais. No eixo das abscissas está representada a escala original da medida latente estimada (o escore do NSE) e, no eixo das ordenadas, a probabilidade de um perfil com dado escore está em cada uma das categorias. Pode-se notar que as alternativas 1 e 4 (respectivamente, as opções “não possui” e “tem dois ou mais”) correspondem aos extremos da escala. Além disso, à medida que aumenta o valor do escore, diminui a probabilidade da opção 1 (não tem automóvel). No extremo direito da escala, que corresponde aos escores mais altos, aumenta a probabilidade da opção 4 (mais de dois automóveis). As opções intermediárias têm probabilidades mais elevadas nas posições centrais da escala, como o esperado.

O gráfico 2 apresenta a curva de informação do item “automóvel”. Esta é uma medida de precisão e mostra o quanto cada uma das posições na escala está sendo bem estimada. Nota-se que a precisão máxima está entre os valores positivos (0 a 3). Estes correspondem aos níveis da escala no qual o item contribui com mais informação e o modelo consegue distinguir melhor os perfis, discriminando escores mais altos dos intermediários. Abaixo do ponto zero e especialmente do ponto -1 a informação fornecida para esse item decresce rapidamente. Isso porque, neste ponto da escala, a probabilidade de ocorrência da resposta 1 (não tem automóvel) chega a quase 100%. Ou seja, não há variação entre os perfis, o que torna o item pouco informativo nessa posição da escala. Por isso, é importante incluir itens que contribuam com informações sobre perfis no extremo inferior da escala.

**Gráfico 1- curva característica do item automóvel**

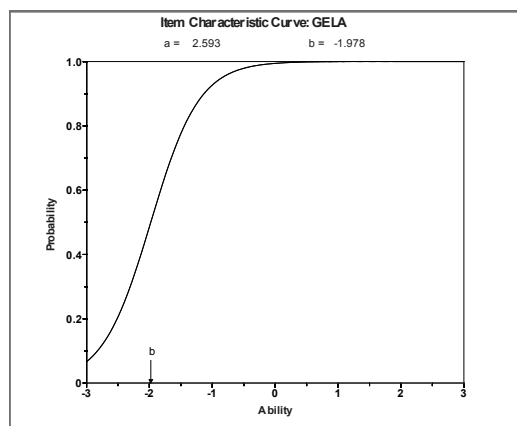


Fonte: Os autores (2014)

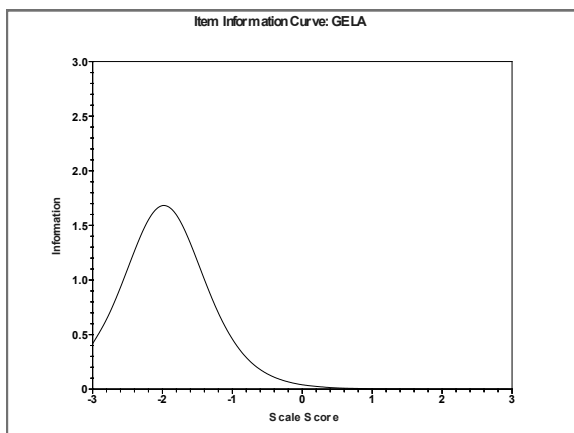
**Gráfico 2 - curva de informação do item automóvel**

Fonte: Os autores (2014).

O item “geladeira” cumpre este papel para o extremo inferior da escala de NSE. Embora um bem de consumo dos mais comuns, há ainda famílias que, no momento da resposta do questionário pelo aluno não o possuíam. Isso se nota no gráfico 3. O índice é dicotômico (“tem geladeira” ou “não tem”), logo a curva característica indica o ponto de inflexão, a partir do qual praticamente todas as respostas são positivas, isto é, “tem geladeira”. Este ponto está situado abaixo do nível -1, na escala original. A curva de informação do item mostra que ele é importante exatamente para distinguir os perfis cujos escores os situam no extremo inferior da escala.

**Gráfico 3 - curva característica do item geladeira**

Fonte: Os autores (2014).

**Gráfico 4 - curva de informação do item geladeira**

Fonte: Os autores (2014).

No conjunto, essa análise produziu resultados bem satisfatórios quanto à relação dos itens selecionados com o construto latente, bem como o grau de informação que eles trazem sobre os perfis das famílias ao longo da escala.

Outra evidência importante sobre a qualidade dos itens é feita pela comparação entre as frequências esperadas com as frequências observadas das categorias de cada um dos itens. Essas duas proporções obtidas, são muito próximas, para todos os itens. A tabela com os valores para os itens pode ser obtida diretamente com os autores.

## 4 Resultados

### 4.1 Resultado parcial: escores dos perfis

Foi estimado um escore (theta) para cada perfil observado que corresponde à medida latente para aquele padrão de resposta aos itens. O passo seguinte foi atribuir a cada aluno o seu NSE que, naturalmente, é igual ao de seu perfil de respostas. Ou seja, há muitos alunos com o mesmo NSE. Finalmente calcularam-se as médias dos escores dos alunos de cada escola, tomando-se esta média como o NSE da escola.

Entretanto, antes de se produzir o resultado final, avaliou-se que os escores apresentados em unidades de desvios-padrão não são de fácil interpretação. Como a intenção deste trabalho é produzir uma medida do nível socioeconômico

das escolas, que pudesse ter uma ampla utilização além do ambiente estritamente acadêmico, optou-se por converter a escala original em um intervalo que contenha valores positivos e que também tenha interpretação substantiva.

A solução mais óbvia de conversão dos valores empíricos observados num intervalo de 0 a 10 não é a mais adequada, porque supõe que existiriam escolas cujo NSE fosse igual a zero. Essa situação tem probabilidade mínima de ocorrências, pois só ocorreria se todos os alunos dessa escola tivessem exatamente o mesmo padrão de respostas igual ao valor mínimo. Pelo mesmo raciocínio, uma escola com NSE médio igual a 10 seria improvável.

É razoável supor que a escala do NSE das escolas varie entre valores teoricamente estabelecidos, considerando que há alunos na base e no topo da hierarquia social. Assim, teoricamente é esperado que os alunos da base da hierarquia não possuam nenhum dos itens de bens de consumo descritos, tenham renda familiar no nível mais baixo, seus pais tenham escolaridade muito baixa (ou nenhuma escolaridade) e exerçam ocupações temporárias, informais ou de baixo prestígio ocupacional (ou ainda estão desempregados). Da mesma forma, teoricamente é esperado que os alunos no topo da hierarquia social possuam todos os itens de consumo nas situações mais vantajosas, tenham renda familiar na categoria mais alta, bem como pais com escolaridade superior e em ocupações de alto prestígio, em posições de comando ou liderança.

Todavia a aplicação deste modelo teórico aos dados empíricos enseja dois problemas. Primeiro, as bases de dados utilizadas neste estudo não foram planejadas para serem analisadas em conjunto. Como registrado anteriormente, a compatibilização dos itens foi uma das etapas cruciais do trabalho. Além disso, os itens considerados para descrever a hierarquia social não estão igualmente incluídos em todos os questionários. Por isso, os casos teóricos acima descritos são impossíveis de ser observados. O segundo problema é que, mesmo que os itens fossem os mesmos em todos os questionários, os casos extremos podem simplesmente não aparecer, haja vista que as avaliações não contemplam todos os alunos brasileiros de educação básica.

A solução adotada para estimar os perfis teóricos posicionados mais à direita e à esquerda na escala original estimada pelo modelo adotado foi adicionar alguns poucos perfis teóricos de alunos à base de dados para a estimação de uma nova escala com amplitude maior e que possa gerar uma interpretação



mais substantiva. Os perfis teóricos são casos extremos simulados que possuem “respostas” para todos os itens dos questionários nas situações menos favoráveis e mais favoráveis referentes, respectivamente, à base e ao topo da hierarquia social. Eles foram adicionados à base de dados com peso equivalente ao dos casos empíricos mais raros observados na base de dados originais, que são aqueles com apenas uma ocorrência.

Feito isto, estimaram-se os escores finais. Eles foram comparados com os escores estimados antes da inclusão dos perfis simulados. O resultado não produziu valores diferentes para os casos reais. Como pode ser observado na Tabela 1, a média e o desvio-padrão do escores (theta) estimados pela TRI para os perfis se mantiveram. Mas, como esperado, com a inclusão dos perfis simulados a amplitude da escala aumentou.

Tabela 1- Estatísticas descritivas do valor de Theta por perfil de alunos empírico e com acréscimo de casos simulados

	N	Mínimo	Máximo	Média	Desvio- Padrão
Theta (perfil original)	3.374.178	-2,37	2,37	0,06	0,58
Theta (com perfis simulados)	3.374.182	-2,71	2,71	0,06	0,58

Fonte: Os autores (2014).

Estes valores teóricos são assumidos como estimativas plausíveis dos valores mínimos e máximos do escore que mede o NSE, na escala original, para a população de referência. A escala usada para o NSE final é obtida com a seguinte equação.

$$NSE10 = \left[ \frac{(\text{valor original do NSE} - \text{valor mínimo teórico})}{(\text{valor máximo teórico} - \text{valor mínimo teórico})} \right] \times 10 \quad (3)$$

Com esse procedimento, o valor mínimo estimado para o perfil simulado mais baixo passou a ter o valor 0, e o valor mais alto, 10.

4.2 Resultado final: NSE das escolas

Ao final deste processo, foram obtidos o NSE médio para 81.140 escolas de educação básica. Para evitar distorções nas médias, optou-se por divulgar o índice de NSE apenas das escolas que possuem informação de pelo menos 15 alunos. Assim, o banco de dados final contém informações sobre 72.018 escolas,

82,4% delas em área urbana. A sinopse estatística da educação básica com base no Censo Escolar 2011 registra a existência, no Brasil, de 61% de escolas em área urbana. A diferença acima registrada é esperada porque os dados analisados neste estudo não têm representatividade censitária, conforme explicado anteriormente.

A Tabela 2 apresenta as estatísticas descritivas do índice de NSE das escolas, considerando somente aquelas que têm informações válidas para, pelo menos, 15 alunos. Observa-se que os novos valores possibilitam uma interpretação mais substantiva dos resultados. Assim, embora a escala teórica varie de 0 a 10, os valores empíricos das escolas estão contidos em uma amplitude menor. Poder-se-á supor que, caso se dispusesse de dados sobre todas as escolas do país, com questionários planejados para essa finalidade e aos quais os alunos responderam por completo, a amplitude aumentaria um pouco mais. Porém, ela só atingiria os limites caso houvesse, por um lado, uma escola onde todos os alunos tivessem o NSE igual ao mínimo observado, ou, por outro lado, se a escola tivesse todos os alunos com o NSE igual ao máximo.

A média observada é inferior ao ponto médio da escala, mas isso é esperado porque 90,8% das escolas para as quais o NSE foi calculado são estaduais ou municipais, nas quais parte dos alunos vêm de famílias com baixo NSE.<sup>6</sup>

Tabela 2 - Estatísticas descritivas do índice de NSE das escolas

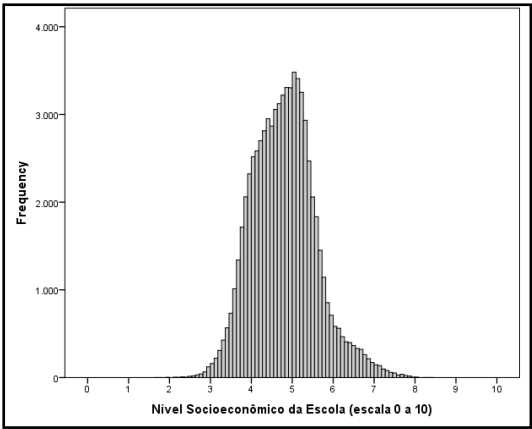
	N	Mínimo	Máximo	Média	Desvio-Padrão
NSE da escola	72.018	1,65	8,38	4,80	0,78

Fonte: Os autores (2014).

O gráfico 5 apresenta a distribuição dos valores do NSE em um histograma, no qual é possível apreciar a maior concentração de valores baixos entre as escolas. O gráfico 6 relaciona o nível socioeconômico das escolas em relação à dependência administrativa. A linha de referência marca a média de todos os valores calculados. Observa-se que escolas municipais e estaduais têm NSE semelhantes e em níveis mais baixos do que escolas federais e privadas. As escolas privadas, por sua vez, agregam os alunos de maior nível socioeconômico.

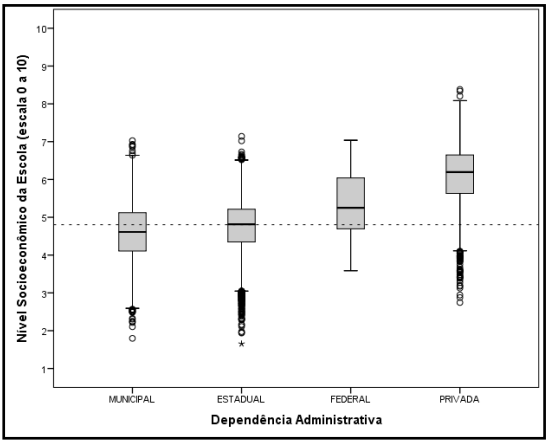
<sup>6</sup> A sinopse estatística da educação básica com base no Censo Escolar 2011 registra a existência, no Brasil, de 81,0% de escolas públicas das redes estaduais e municipais no nível de educação básica, percentual que inclui do infantil ao ensino médio e todas as modalidades de ensino (regular, EJA, especial, profissional).

Gráfico 5 - histograma do NSE escola



Fonte: Os autores (2014).

Gráfico 6 - Box plot NSE por rede de ensino



Fonte: Os autores (2014).

Finalmente, para gerar o banco de dados com o NSE de cada escola, fez-se a identificação dos dados das escolas (nome, localização por município e unidade da federação). Entretanto, ao longo dos anos, escolas podem deixar de existir com o código Inep registrado em nossos bancos de dados. Seja porque elas deixaram de funcionar, ou porque houve transferência entre redes de ensino (exemplo, a municipalização de escolas de ensino fundamental) ou outra razão. Assim, é possível

que algumas escolas registradas no banco de dados cujos dados foram obtidos em anos anteriores a 2011 não existam mais no sistema de informações do Inep.

### 4.3 Criação de grupos de NSE médio das escolas

A caracterização das escolas, a divisão da escala de NSE em grupos de escolas pode tornar o uso dessa informação mais fácil. Os agrupamentos tendem a ser mais estáveis do que a posição específica de uma escola na hierarquia obtida.

A divisão das escolas em grupos por percentis não se mostra a mais adequada. Isto porque analisando-se o histograma do NSE das escolas (Gráfico 6) nota-se que há quase uma justaposição de duas curvas normais. Em geral, as escolas municipais e estaduais são aquelas com menor NSE enquanto as federais e privadas possuem maior NSE. Mas, os intervalos dos valores do NSE das redes de ensino se inter cruzam. Ou seja, há escolas privadas com baixo NSE.

Considerando essa distribuição e obedecendo à simetria dos dados, foram criados 7 grupos por meio da análise de conglomerados pelo método de agrupamento não hierárquico designado por k-means. O conglomerado nesse caso, equivale a um conjunto de perfis de escolas no qual cada perfil está mais próximo ao centroide que define o cluster do que dos centroides de quaisquer outros clusters. A partir deste procedimento, para tornar os grupos mais interpretáveis, eles foram denominados como: (1) Mais baixo; (2) Baixo; (3) Médio Baixo; (4) Médio; (5) Médio Alto; (6) Alto; e (7) Mais alto.

A Tabela 3 mostra a frequências dos grupos de escolas por NSE. Como esperado, os grupos extremos possuem percentual menor do que os grupos intermediários.

Tabela 3 - Frequência dos Grupos de Escolas por NSE

Grupos	Frequência	Percentual	Percentual acumulado
Mais baixo	3756	5,2	5,2
Baixo	12698	17,6	22,8
Médio baixo	16942	23,5	46,4
Médio	19219	26,7	73,1
Médio alto	13462	18,7	91,8
Alto	4225	5,9	97,6
Mais alto	1716	2,4	100,0
Total	72018	100,0	

Fonte: Os autores (2014).

A Tabela 4 mostra a distribuição dos grupos de NSE por dependência administrativa. Observa-se que as escolas municipais estão concentradas nos grupos “mais baixo”, “baixo” e “médio baixo”. Já as escolas estaduais estão nos grupos “médio baixo”, “médio” e “médio alto”. As federais estão no “médio”, “médio alto” e “alto”, assim como as escolas privadas, que se destacam nos dois últimos grupos. Vale destacar que no grupo “mais baixo” a maioria das escolas pertence à rede pública, mas há algumas escolas privadas. A pesquisa aos nomes dessas escolas revelou que muitas são escolas rurais, vinculadas a instituições religiosas e oferecem bolsas de estudo. Duas delas foram citadas em um jornal de uma cidade de um estado do Nordeste como aquelas que obtiveram o pior desempenho no Enem do estado. Considerando que o desempenho escolar é fortemente correlacionado ao nível socioeconômico, a observação dessas escolas está dentro de um cenário possível. A análise do jornal é que foi injusta, pois não considerou a origem social desses alunos.

Tabela 4 - Grupos de NSE por Dependência Administrativa da Escola

	Municipal	Estadual	Federal	Privada	Total
Mais baixo	2337	1393	1	25	3756
	5.9%	5.4%	.5%	.4%	5.2%
Baixo	9084	3553	19	42	12698
	23.1%	13.7%	8.7%	.7%	17.6%
Médio baixo	10241	6461	38	202	16942
	26.0%	24.9%	17.4%	3.1%	23.5%
Médio	10148	8453	49	569	19219
	25.8%	32.5%	22.4%	8.8%	26.7%
Médio alto	6751	5314	46	1351	13462
	17.1%	20.5%	21.0%	21.0%	18.7%
Alto	810	797	48	2570	4225
	2.1%	3.1%	21.9%	39.9%	5.9%
Mais alto	13	5	18	1680	1716
	.0%	.0%	8.2%	26.1%	2.4%
Total	39384	25976	219	6439	72018
	100.0%	100.0%	100.0%	100.0%	100.0%

Fonte: Os autores (2014).

## 4.4 Validação dos resultados

Na última etapa deste trabalho, o índice de NSE das escolas foi avaliado por meio de métodos quantitativos e qualitativos.

Primeiramente, testou-se a correlação com o NSE da pesquisa Geres: Geração escolar 2005, um estudo longitudinal com quatro anos de duração que acompanhou, em cinco cidades do Brasil, alunos matriculados em 312 escolas públicas e privadas durante os anos iniciais do ensino fundamental. Mais informações sobre essa pesquisa podem ser consultadas em Franco, Brooke e Alves (2008). Na pesquisa Geres, o NSE dos alunos foi calculado com base nas informações sobre o nível de renda da família (estimado pela posse de bens eletrodomésticos, automóvel e a contratação de serviços domésticos) e a ocupação e a escolaridade dos pais. Há 229 escolas em comum nas duas bases de dados<sup>7</sup>, e a correção entre os índices estimados é de 0,93.

Em um estudo realizado com os dados do polo Belo Horizonte da pesquisa Geres, onde foram acompanhados alunos de 60 escolas, o NSE foi calculado de modo semelhante, porém o rendimento da família foi estimado por meio da renda média do setor censitário de residência dos alunos obtida com base nos microdados do Censo Populacional (ALVES, SOARES, 2009). Há 50 escolas em comum entre as duas bases de dados e a correção observada entre o NSE delas é de 0,89.

Outra medida de NSE disponível provém de uma avaliação estadual realizada pela Secretaria de Estado da Educação do Rio Grande do Sul. Naquele estado, durante a matrícula para o ano letivo de 2009 os pais dos alunos preencheram um questionário informando sobre a sua ocupação, escolaridade e renda. Há 4.015 escolas em comum nas duas bases de dados (somente escolas públicas). A correlação entre os índices de NSE obtidos é também alta, atingindo o valor de 0,95.

Testou-se também, com dados obtidos em avaliações da rede estadual realizadas no estado de São Paulo, nos anos de 2008, 2009 e 2010. Nesse caso, são os pais que preenchem a informação de escolaridade e itens de conforto doméstico. Há 11.538 escolas em comum. A correlação também atingiu um valor também muito alto, de 0,97.

<sup>7</sup> O Geres inclui muitas escolas privadas pequenas e que atendem somente o Ensino Fundamental. Este segmento não está representado no presente estudo uma vez que os alunos dessas escolas não fazem o Enem. Somente escolas privadas com ensino médio (algumas que atendem também ao Ensino Fundamental) estão incluídas neste estudos.

Testou-se ainda a correlação entre o indicador NSE agregado por município com a rendimento mensal *per capita* por município com base no Censo 2010. Há 5.565 municípios em comum nas duas bases de dados, ou seja, todos os municípios brasileiros. O resultado da correlação entre os dois indicadores é muito alto, atingindo 0,90.

Finalmente de forma qualitativa, educadores de várias partes do país foram solicitados a verificar se a classificação obtida pelo índice de NSE é consistente com o conhecimento que cada um deles tem da realidade local. Esta análise também apresentou resultado bastante fiel.

## 5 Considerações finais

Este trabalho reuniu dados de avaliações educacionais realizadas durante os anos de 2003 a 2011 para produzir um índice de nível socioeconômico (NSE) das escolas de ensino fundamental e médio. O maior desafio enfrentado para isso foi produzir um indicador consistente, uma vez que os dados coletados, por meio de questionários contextuais respondidos pelos alunos nessas avaliações, não foram planejados para essa finalidade. Inúmeras decisões metodológicas tiveram que ser tomadas, versões prévias testadas, até se chegar a este resultado. O banco de dados com o resultado final tem um enorme potencial para orientar estudos e pesquisas educacionais. Entretanto, algumas ressalvas para a leitura dos resultados devem ser feitas.

Estão ausentes estabelecimentos de ensino que não se enquadram no perfil das escolas representadas nas bases de dados cujas informações foram agregadas para se produzir o NSE médio das escolas. Por exemplo, não há identificação das escolas privadas que atendem apenas ao ensino fundamental. Este segmento escolar contribuiu para a estimação do NSE porque ele está representado nas bases de dados do Saeb. Mas, como explicado na metodologia, não é possível identificar nominalmente escolas disponíveis na bases de dados do Saeb já que quando do seu aceite para participar dessa avaliação, o INEP comunicou a cada escola que seus dados não seriam divulgados de forma que a escola pudesse ser identificada.

As escolas do setor privado está identificado quando presente nas bases de dados do Enem, que contém informação sobre o estabelecimento onde o aluno

concluiu o ensino médio. Eventualmente, estes podem atender também ao ensino fundamental. Mas o Enem, por ser um teste voluntário, não deve ser tomado como parâmetro para representar o setor privado.

O NSE das escolas localizadas em áreas rurais foi calculado com base em informações das duas últimas edições da Prova Brasil, quando elas passaram a fazer parte do exame (desde que tenham pelo menos 20 alunos na série avaliada) ou então são escolas rurais de ensino médio (possivelmente técnico), que são mais raras. Assim, escolas rurais pequenas não estão representadas no estudo.

Também não foi calculado o NSE de escolas cujo número de alunos que participaram das avaliações realizadas no período considerado é inferior a 15, para não gerar distorções excessivas na estimativa média.

Devido a limitações dos dados, não se considerou diferenças intraescola, isto é, um NSE por turno ou turma dentro da mesma escola. A literatura aponta que essas diferenças não são negligenciáveis. Contudo, estimar um NSE em separado, por exemplo, para cada turno, seria pouco plausível porque os dados não foram planejados para tal finalidade e isso reduziria muito as possibilidades de análise.

Apesar dessas ressalvas, o banco de dados produzidos com o NSE das escolas contempla um largo espectro de escolas brasileiras – públicas e privadas. O cálculo desse indicador possibilita o estudo das desigualdades educacionais entre alunos, escolas e municípios associadas ao nível socioeconômico. Os resultados mostram que o índice de NSE construído a partir de uma base nacional de dados de várias fontes é bastante consistente. As correlações com resultados de outras avaliações mostrou resultados positivos e muito altos. Da mesma forma, a correlação com o indicador do rendimento médio *per capita* dos municípios revelam que a visão da realidade nacional vista por um indicador econômico e um indicador obtido pelas respostas de alunos a um questionário contextual é, no nível macro, a mesma.

A inspeção qualitativa dos estabelecimentos de ensino situados nos extremos superior e inferior coincidiu com as percepções sobre a realidade educacional de várias cidades brasileiras. Conseguiu-se, assim, um bom resultado mesmo junto a públicos muito difíceis de se acessar de forma direta ou indireta, como é o caso das famílias de alto poder aquisitivo, cujos filhos estudam em escolas privadas mais



elitizadas. Com todos os esforços feitos, pode-se afirmar que o índice de NSE produzido reflete uma ordenação em grupos que descreve de forma fidedigna a realidade socioeconômica das escolas incluídas na base de dados.

Em termos práticos, o índice de NSE pode ser usado como variável explicativa ou de controle em estudos educacionais, ou no planejamento de estudos comparativos para a definição dos grupos de escolas. Neste sentido, é importante evitar absolutizar os valores do NSE, porque há uma variação em torno das médias. Se houver inclusão de novos dados ou de novos itens e se o modelo for ajustado novamente a estes novos dados, a posição de escolas próximas na atual ordenação pode mudar. Isso indica, de forma clara, que as hierarquias derivadas dessa escala não são absolutas. Isso limita a utilidade do uso destes dados para a mera produção de *rankings* de escolas.

Os agrupamentos de escolas propostos neste trabalho são mais estáveis do que a posição específica de uma escola na hierarquia obtida. Assim sendo, as comparações de resultados entre escolas dentro dos grupos homogêneos são mais justificativas. Hoje a ausência de um indicador como este induz a comparação entre escolas com alunados muito diferentes, como se vê nos *rankings* de escolas cujos alunos participam do Enem ou das escolas públicas pelo Índice de Desenvolvimento da Educação Básica (Ideb).

Finalmente, esse estudo mostra a importância dos questionários contextuais para a pesquisa educacional. As diferentes versões desses questionários entre as avaliações, que tiveram inclusão e exclusão de itens de um ano para outro, bem como alterações na própria formulação dos itens apontam para a inexistência de referências teóricas mais sólidas na construção desses instrumentos. Naturalmente, não se defende que os questionários se tornem fixos, mas as mudanças precisam ser justificadas, pré-testadas e validadas e as mudanças que, porventura, sejam feitas ocorram sem prejuízo de comparabilidade.

A título de exemplo, nota-se que o questionário da edição do Enem 2011, a última utilizada neste trabalho, excluiu os itens sobre posição ocupacional dos pais dos alunos. Mesmo reconhecendo a dificuldade de se medir essa dimensão em questionários autoaplicados, a literatura sociológica aponta a ocupação com a variável mais importante na estratificação social. Essa dimensão está presente no ISEI, o índice de nível socioeconômico utilizado no PISA e em outros estudos

compararam resultados educacionais entre países ou regiões. Nesse trabalho, as análises desses itens mostraram que eles têm bom poder de discriminação dos perfis familiares e agregam informação ao construto. Argumenta-se, inclusive, que esses itens poderiam ter testados junto aos alunos mais velhos que fazem o Saeb e a Prova Brasil, aqueles que estão no nono ano do ensino fundamental e no ensino médio. Todavia, a decisão adotada no último Enem sugere que o Inep passou a considerar referenciais que privilegiam apenas a discriminação das famílias pelo poder de compra.

## Referências

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

\_\_\_\_\_. Contexto escolar e indicadores educacionais: condições desiguais para a efetivação de uma política de avaliação educacional. *Educação e Pesquisa*, v. 39, n. 1, p. 177-194, 2013.

BLAU, P. M.; DUNCAN, O. D. *The American occupational structure*. New York: Wiley. 1967.

BOURDIEU, P; PASSERON, J-C. *A reprodução*: Elementos para uma teoria do sistema de ensino. 5. ed. Petrópolis: Editora Vozes, 2011.

COLEMAN, J. S. et al. *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office, 1966.

DUNCAN, O. D. A Socioeconomic Index for All Occupations. In: REISS, A. J. (Ed.). *Occupations and Social Status*. New York: Free Press. 1963, p. 109-138.

ERICKSON, R.; GOLDTHORPE, J. H. The CASMIN Project and the American Dream. *European Sociological Review*, [S.l.], v. 8, p. 283-306, 1992.

FRANCO, C. et al. O Referencial teórico na construção dos questionários contextuais do SAEB 2001. *Estudos em Avaliação Educacional*, Rio de Janeiro, v. 28, jul./dez., p. 39-74. 2003.

FRANCO, C.; BROOKE, N.; ALVES, F. T. Estudo longitudinal sobre qualidade e equidade no ensino fundamental brasileiro: GERES 2005. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v.16, p. 625-637, 2008.

GAZENBOOM, H. B. G. A new international socio-economic index [ISEI] of Occupational status for the international standard classification of occupation 2008 [ISCO-08] constructed with data from the ISSP 2002-2007; with an analysis of quality of occupational measurement in ISSP. In: ANNUAL CONFERENCE OF INTERNATIONAL SOCIAL SURVEY PROGRAMME, 1., 2010, Lisbon. *Trabalhos apresentados...* Lisbon, maio 2010. Disponível em: <<http://www.harryganzeboom.nl/Pdf/index.htm>>. Acesso em: 1 dez. 2013.

GAZEBOOM, H. B. G.; DE GRAAF, P.; TREIMAN, D. A Standard International Socio-Economic Index of Occupational Status. *Social Science Research*, [S.l.], v. 21, p. 1-56, 1992.

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

HAUSER, R. M.; WARREN, J.; R. Socioeconomic Index of Occupational Status: A Review, Update, and Critique. In: RAFTERY, A. (Ed.). *Sociological methodology*, Cambridge: Blackwell. 1997, p. 177-298.

JENCKS, C. et al. *Inequality: a reassessment of the effect of family and schooling in America*. New York: Basic Books, 1972.

MARKS, G. N. et al. *The Measurement of Socioeconomic Status for the Reporting of Nationally Comparable Outcomes of Schooling*. National Education Performance Monitoring Taskforce. [Draft Report]. Australian Council for Educational Research & Sociology Program Research School of Social Sciences Australian National University, maio, 2000.

NERI, M. *A nova classe média: o lado brilhante da base da pirâmide*. Rio de Janeiro: FGV: Centro de Políticas Sociais; Editora Saraiva, 2012.

PASTORE, J. *Desigualdade e mobilidade social no Brasil*. São Paulo: Editora da Universidade de São Paulo, 1979.

PASTORE, J.; SILVA, N. V. *Mobilidade Social no Brasil*. São Paulo: Markron, 2000.

SAMEJIMA, F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, New York, v. 17, Monograph Supplement, n. 4, part 2, 1969.

SANTOS, J. A. F. Uma classificação socioeconômica para o Brasil. *Revista Brasileira de Ciências Sociais*, [S.l.], v. 20, n. 58, p. 27-45, 2005.

SCALON, M. C. Mapeando estratos: critérios para escolha de uma classificação. *Dados: Revista de Ciências Sociais*, Rio de Janeiro, v. 41, n. 2, p. 337-375, 1998.

SOARES, J. F.; COLLARES, A. C. M. Recursos Familiares e o Desempenho Cognitivo dos Alunos do Ensino Básico Brasileiro. *Dados: Revista de Ciências Sociais*, Rio de Janeiro, v. 49, n. 3, p.615-650, 2006.

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.

## Apêndice

Quadro 1 - fontes dos dados, ano da avaliação, séries avaliadas, número de alunos e escolas (\*)

Fonte	Ano	Série/Ano	Alunos	Escolas
Saeb	2003	4a/8a/11a	235.272	7.467
Saeb	2005	4a(5º)/8ª(9o)/11a(12o)	217.107	4.851
Prova Brasil	2005	4a(5º)/8ª(9o)	3.275.821	50.097
Prova Brasil	2007	4a(5º)/8ª(9o)	4.109.265	64.864
Prova Brasil	2009	5o/9º	4.535.301	75.545
Prova Brasil	2011	5o/9º	4.218.135	55.932
Enem	2007	12o (**)	979.585	22.985
Enem	2008	12o (**)	1.029.902	24.379
Enem	2009	12o (**)	848.036	28.446
Enem	2011	12o (**)	1.496.991	31.019

Fonte: Os autores (2014).

**(\*) Um mesmo aluno ou escola pode estar presente em diferentes avaliações.**

**(\*\*) Alunos que terminaram o ensino médio no ano da aplicação do Enem e que responderam ao questionário contextual.**

Quadro 2 - Dimensão, itens, descrição, categorias recodificadas

Dimensão	Item	Descrição	Categorias
Nível educacional dos pais	AlfaMãe	Sua mãe ou a mulher responsável por você sabe ler e escrever?	1. Não. 2. Sim
	AlfaPai	Seu pai ou homem responsável por você sabe ler e escrever?	
	EscolaM	Até que série sua mãe ou a mulher responsável por você estudou?	1. Nunca estudou/não completou a 4ª série. 2. Completou a 4ª série, mas não completou a 8ª série. 3. Completou a 8ª série, mas não completou o Ensino Médio. 4. Completou o Ensino Médio, mas não completou a faculdade. 5. Completou a faculdade.
	EscolaP	Até que série seu pai ou o homem responsável por você estudou?	

(continua)

Nível de rendimento da família	Aspirador	Em sua casa tem aspirador de pó?	1. Não. 2. Sim
	Auto	Em sua casa tem automóvel?	1. Não. 2. Tem 1 3. Tem 2 4. Tem 3 ou mais
	Banheiro (*)	Em sua casa tem banheiro?	1. Não. 2. Tem 1 3. Tem 2 4. Tem 3 ou mais
	PC	Em sua casa tem um computador?	1. Não 2. Sim, sem acesso à internet 3. Sim, com acesso à internet.
	Emprega	Em sua casa trabalha empregada doméstica?	1. Não tem 2. Sim, uma diarista (faxineira, etc.) uma ou duas vezes por semana. 3. Uma ou mais, todos os dias.
	Freezer	Em sua casa tem freezer?	1. Não. 2. Sim
	G_ Duplex	Em sua casa tem geladeira com freezer juntos?	1. Não. 2. Sim
	Geladeira	Em sua casa tem geladeira?	1. Não. 2. Sim
	Máquina	Em sua casa tem uma máquina de lavar roupa (não é tanquinho)?	1. Não. 2. Sim
	Rádio	Em sua casa tem rádio?	1. Não. 2. Tem 1 3. Tem 2 ou mais
	RendaF	Renda Familiar (somando a do respondente e a das pessoas que moram com ele)	1. Nenhuma/ Até 1 salário mínimo 2. De 1 a 2 salários mínimos 3. De 2 a 5 salários mínimos 4. De 5 a 10 salários mínimos 5. Mais de 10 salários mínimos
	TV (*)	Em sua casa tem televisão em cores?	1. Não. 2. Tem 1 3. Tem 2 4. Tem 3 ou mais
	Vídeo_ DVD	Em sua casa tem videocassete/ DVD?	1. Não 2. Sim, tem

(continua)

Ocupação dos pais	ÁreaM	Área que a mãe trabalha ou trabalhou?	1. Área A – mais informal/primário [agrupamento dos itens: (A) Na agricultura, no campo, em fazenda ou na pesca. (F) Trabalhador do setor informal (sem carteira assinada). (G) Trabalha em casa em serviços (costura, cozinha, aulas particulares etc.).] 2. Área B – mais formalizado [agrupamento dos itens: (B) Na indústria. (C) No comércio, banco, transporte ou outros serviços. (D) Funcionário público do governo federal, estadual ou municipal. (E) Profissional liberal, professor ou técnico de nível superior.]; 9. No lar/ Não trabalha/ Não sabe/Sem informação
	AreaP	Área que o pai trabalha ou trabalhou?	
	OcupaM	Posição da ocupação da mãe	1. Nível 1 [agrupamento dos itens: (G) Trabalho temporário, informal, sem carteira assinada. (H) Trabalho por conta própria.] 2. Nível 2 [agrupamento dos itens: (D) Empregado no setor privado, com carteira assinada. (E) Funcionário público (federal, estadual ou municipal), sem função de direção. (F) Militar (guarda-civil, polícia estadual ou Forças Armadas), sem posto de comando.
	OcupaP	Posição da ocupação do pai	3. Nível 3 [agrupamento dos itens: (A) Gerente, administrador ou diretor de empresa privada. (B) Funcionário público (federal, estadual, municipal), com funções de direção. (C) Militar (guarda-civil, polícia estadual ou Forças Armadas), com posto de comando. 9. Desempregado Aposentado/ Outra situação/ /Sem informação

Fonte: Os autores (2014).

**(\*) No banco de dados da Prova Brasil, esse item foi tratado com uma categoria ordinal a menos (a última).**

Quadro 3 - Itens selecionados por base de dados

Variáveis do NSE	PB 2005	PB 2007	PB 2009	PB 2011	SAEB 2003	SAEB 2005	SAEB 2011	ENEM M2007	ENEM 2008	ENEM 2009	ENEM 2011
ALFAMAE	-	1	1	1	1	-	1	-	-	-	-
ALFAPAI	-	1	1	1	1	-	1	-	-	-	-
ESCOLAM	1	1	1	1	1	1	1	1	1	1	1
ESCOLAP	1	1	1	1	1	1	1	1	1	1	1
ASPIRA-DOR	1	1	-	-	1	1	-	-	-	-	1
AUTO	-	1	1	1	1	1	1	1	1	1	1
AUTO2005	1	-	-	-	-	-	-	-	-	-	-
BANHEIRO	-	1	1	1	1	1	1	-	-	-	1
BANHO 2005	1	-	-	-	-	-	-	-	-	-	-
PC	1	1	1	1	1	1	1	1	1	1	1
EMPREGA	1	1	1	1	1	1	1	-	-	-	1
FREEZER1	1	1	1	1	1	1	1	-	-	-	1
G_DUPLEX	-	-	1	1	1	1	1	-	-	-	-
GELADEIRA	1	1	1	1	1	1	1	1	1	1	1
MAQUINA	1	1	1	1	1	1	1	1	1	1	1
RADIO	-	1	1	1	1	1	1	1	1	1	1
RADIO 2005	1	-	-	-	-	-	-	-	-	-	-
RENDAP	-	-	-	-	-	-	-	1	1	1	1
TV	-	1	1	1	1	1	1	1	1	1	1
TV2005	1	-	-	-	-	-	-	-	-	-	-
VIDEO_DVD	1	1	1	1	1	1	1	1	1	1	1
AREAM	-	-	-	-	-	-	-	1	1	1	-
AREAP	-	-	-	-	-	-	-	1	1	1	-
OCUPAM	-	-	-	-	-	-	-	1	1	-	-
OCUPAP	-	-	-	-	-	-	-	1	1	-	-
TOTAL	13	15	15	15	16	14	15	14	14	12	14

Fonte: Os autores (2014).



## **Socioeconomic index of Brazilian Basic Education schools**

### **Abstract**

*This article presents the methodology and the results of a new index of socioeconomic status of the schools of basic education in Brazil. The data used are from the contextual questionnaires filled by students at the time they answer the cognitive tests of several educational assessments organized by the federal government. Valid responses from 20,806,062 students in 21 databases were considered. In order to estimate the index, items related to the dimensions of education and occupation of student's parents and family income were aggregated using a model of Item Response Theory. The results were validated showing to be quite reliable. Correlations of the index introduced in this paper with similar indices, especially the per capita income of municipalities are positive and high. This shows that the reality described by the national economic indicator and an indicator obtained by students responses to a questionnaire in an educational evaluation, at the macro level, are basically the same. The assessment of socioeconomic status of schools as perceive by educational specialists in different localities of the country coincides with the one depicted by the indicator.*

**Keywords:** Socioeconomic status of schools. Basic education. Educational inequalities.

## **Índice Socioeconómico de las Escuelas de Educación Básica de Brasil**

### **Resumen**

*Este artículo presenta la metodología y los resultados del desarrollo de un índice de nivel socioeconómico de las escuelas de educación básica en Brasil. Los datos proceden de los cuestionarios contextuales que los alumnos contestaron en las evaluaciones educativas que el gobierno federal realizó para este nivel de educación. Se incluyeron en este estudio respuestas válidas de 20.806.062 estudiantes de 21 bases de datos diferentes. Para estimar el índice se agregan los indicadores sobre las dimensiones de escolaridad, ocupación de los padres y los ingresos de la familia de los estudiantes y se utiliza un modelo de la Teoría de Respuesta al Ítem. Los resultados fueron validados y se pudo demostrar que son confiables. Las correlaciones con índices similares son positivas y altas. La*

*correlación con la renta per cápita de los municipios muestra que la realidad vista por un indicador económico nacional y un indicador obtenido mediante respuestas a un cuestionario contextual contestados por estudiantes, en el nivel macro, es la misma. La evaluación de los especialistas en distintas localidades del país coincide con la retratada por el indicador.*

**Palabras clave:** *Nivel socioeconómico de las escuelas. Educación básica. Desigualdades educacionales.*

