

Uma medida do nível socioeconômico das escolas brasileiras utilizando indicadores primários e secundários¹

José Francisco Soares² 

Maria Teresa Gonzaga Alves³ 

Este artigo apresenta uma medida do nível socioeconômico (NSE) de quase todas as escolas brasileiras. Indicadores primários – calculados com dados de questionários respondidos pelos estudantes nas avaliações educacionais – e indicadores secundários – que caracterizam o contexto social das escolas – foram agregados em uma escala de NSE por meio de um modelo da Teoria de Resposta ao Item. O NSE foi calculado para mais de 180 mil escolas. A metodologia deu visibilidade às escolas pequenas ou de áreas remotas que, até então, eram invisíveis para as políticas públicas, uma vez que elas não participam das avaliações educacionais. A relação entre resultados educacionais e o NSE representa as desigualdades do país. O NSE das escolas tem utilidade para a pesquisa acadêmica e potencial para contribuir com o aperfeiçoamento das políticas públicas que visam a reduzir as desigualdades educacionais.

Palavras-chave: nível socioeconômico das escolas; educação básica; desigualdades educacionais; escalas e medidas sociais; Teoria de Resposta ao Item

Introdução⁴

O nível socioeconômico (NSE) é uma variável-chave nas pesquisas sociais. De forma especial, tem grande importância para a compreensão da desigualdade na educação, tendo em vista a associação entre resultados educacionais e a origem social dos estudantes (Coleman et al., 1966; Bourdieu; Passeron, 1994; Sirin, 2005; Soares; Colares, 2007; Alves et al., 2013; Alves; Soares; Xavier, 2016; Broer; Bai; Fonseca, 2019). No nível das

¹ Uma versão preliminar deste artigo foi publicada no servidor *preprint* SSRS: <<http://dx.doi.org/10.2139/ssrn.4325674>>. Acesso em: 22 set. 2023.

² Universidade Federal de Minas Gerais. Belo Horizonte (MG), Brasil. E-mail: <francisco.soares.ufmg@gmail.com>.

³ Universidade Federal de Minas Gerais, Departamento de Ciências Aplicadas à Educação, Programa de Pós-Graduação em Educação: conhecimento e inclusão social. Belo Horizonte (MG), Brasil. E-mail: <mtga@ufmg.br>.

⁴ Agradecemos à Fundação Itaú Social pelo financiamento que nos permitiu contar com os auxiliares de pesquisa Túlio Silva de Paula e José Aguinaldo Fonseca na elaboração de códigos em R para a preparação dos bancos de dados e estimação de indicadores primários utilizados neste artigo.

escolas, essa associação costuma ser ainda mais forte e, portanto, não faz sentido analisar a eficácia escolar sem a consideração da composição social de seus estudantes (Willms, 2010; Xavier; Alves, 2015; Sciffer; Perry; McConney, 2020, 2022).

Desde a publicação do clássico Relatório Coleman, há mais de 50 anos, a redução dessa associação permanece como um desafio para as políticas educacionais (Coleman et al., 1966). Persistem, também, desafios na medição do NSE (Buchmann, 2002; Rutkowski; Rutkowski, 2013; Broer; Bai; Fonseca, 2019; Avvisati, 2020) e na identificação de seu impacto nos diferentes contextos educacionais (Palardy, 2008; Willms, 2010; Armor; Marks; Malatinszky, 2018).

O cálculo do NSE dos estudantes e das escolas já foi feito muitas vezes no Brasil utilizando-se dados dos questionários contextuais do Sistema de Avaliação da Educação Básica (Saeb), do Exame Nacional do Ensino Médio (Enem) e outras avaliações ou exames educacionais (Soares, 2005; Soares; Andrade, 2006; Alves; Soares, 2009; Alves; Gouveia; Viana, 2014; Alves; Soares; Xavier, 2014; Barros et al., 2019). Esses questionários incluem itens que podem ser tomados como expressão do NSE dos estudantes, e o NSE de uma escola é, usualmente, a média do NSE dos seus estudantes.

No entanto, essa metodologia restringe a possibilidade de cálculo do NSE às escolas que participam dessas avaliações, o que, geralmente, não alcança nem a metade das escolas registradas no Censo Escolar. Escolas para as quais essas informações não estão disponíveis se tornam quase invisíveis para as políticas públicas que visam à redução das desigualdades educacionais. As lacunas são mais proeminentes entre os estabelecimentos de ensino pequenos, localizados em cidades menores, nas áreas rurais, indígenas e quilombolas, nas regiões Norte e Nordeste, justamente as mais pobres do país. Também há ausência de informações sobre escolas privadas que não ofertam o ensino médio e, por isso, não têm estudantes inscritos no Enem.

O objetivo deste artigo é apresentar uma medida do NSE de praticamente todas as escolas ativas de educação básica no Brasil. Para isso, utilizamos, além da média do NSE dos estudantes por escola, dados públicos que caracterizam o contexto social dos estabelecimentos de ensino, mas que ainda não haviam sido utilizados para esse fim.

Embora a metodologia utilizada para expandir o indicador de NSE para um número maior de escolas tenha relevância científica, é a sua importância nas políticas públicas que será mais enfatizada neste artigo. No Brasil, o Plano Nacional da Educação (PNE) estabelece que os resultados do Índice de Desenvolvimento da Educação Básica (Ideb) – o indicador que monitora a qualidade da educação – devem ser contextualizados em relação a indicadores sociais, dentre eles o nível socioeconômico (Brasil, 2014). Em decorrência dessa normativa, o Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (Inep) passou a calcular o Índice de Nível Socioeconômico (Inse) das escolas de educação básica. Mais recentemente, a lei do “Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação” (Fundeb) introduziu a

redução das desigualdades socioeconômicas no conjunto de condicionalidades para fins de distribuição dos seus recursos entre as redes públicas de ensino (Brasil, 2020).

Entretanto, o Inep publicou, em novembro de 2022, uma nota técnica que aponta dificuldades para aplicar essa condicionalidade do novo Fundeb e recomenda a sua revisão para essa finalidade (Inep, 2022). O Inse foi planejado para contextualizar os resultados do Saeb, não para orientar a distribuição de recursos do Fundeb para as redes de ensino. Quando este artigo foi elaborado, a última versão oficial do Inse disponível era a calculada com dados do Saeb 2019 para 68.868 escolas públicas (Inep, 2020). Esse número corresponde a 49% das escolas públicas e 38% do total de escolas de educação básica.

A inexistência de informações sobre a composição socioeconômica de muitas escolas para orientar as decisões de políticas educacionais pode ter alto impacto para as redes de ensino cujas escolas públicas estão sub-representadas. Este artigo contribui para o enfrentamento desse problema da seguinte forma: após esta introdução, revisamos a literatura que aborda o conceito e a mensuração do NSE no nível do estudante e das escolas. Em seguida, apresentamos a metodologia utilizada para expandir o número de escolas com NSE calculado e, depois, mostramos a descrição dessa medida e a sua associação com resultados educacionais. Por último, discutimos as implicações dessa metodologia para as pesquisas e as políticas educacionais.

Nível socioeconômico: conceito e medida

O nível socioeconômico (NSE) dos estudantes é uma expressão sintética produzida nas pesquisas sociais para situá-los em uma hierarquia social. Já o NSE das escolas é uma medida de sua composição social utilizada para descrever as desigualdades entre a aprendizagem e os contextos educacionais em que ocorre o processo de ensino. São conceitos distintos, mas intrinsecamente vinculados a operacionalização empírica, como explicaremos na sequência.

NSE de estudantes

A definição do NSE dos estudantes se fundamenta nos estudos sociológicos da estratificação social que explicam as hierarquias sociais pelo prestígio ocupacional, a riqueza, o nível de instrução, o poder e outros ativos que conferem vantagens para indivíduos ou grupos nas sociedades (Boudon; Bourricaud, 1982; Harwell; Lebeau, 2010). Do ponto de vista empírico, o NSE dos estudantes é um traço latente que se manifesta por variáveis observáveis sobre o contexto familiar, usualmente a escolaridade, a renda e a ocupação dos pais do estudante (Buchmann, 2002; Sirin, 2005; Alves; Soares, 2009; Avvisati, 2020).

As informações para o cálculo do NSE são geralmente obtidas por meio de questionários. No entanto, em muitas avaliações educacionais, os questionários não

coletam diretamente a ocupação e a renda dos pais, em razão de privacidade ou porque os estudantes, dependendo da idade ou do contexto familiar em que vivem, não conseguem fornecer com precisão essas informações (Broer; Bai; Fonseca, 2019). Além disso, há o desafio de lidar com dados incompletos, comuns quando se coleta dados com questionários (Ferrão; Prata; Alves, 2020). No entanto, as avaliações educacionais avançaram bastante para superar essas limitações e produzir medidas robustas de NSE dos estudantes.

A avaliação internacional PISA (sigla em inglês para *Programme for International Student Assessment*) desenvolveu uma medida composta de status socioeconômico e cultural (ESCS, sigla em inglês para *Index of Economic, Cultural and Social Status*) que influencia as metodologias das avaliações educacionais ao redor do mundo (Alves; Soares, 2009; Rutkowski; Rutkowski, 2013; Avvisati, 2020). O ESCS do PISA é calculado por uma média ponderada de três índices: o nível educacional dos pais de acordo com a Classificação Padronizada Internacional da Educação da Unesco (ISCED, sigla em inglês); o status profissional dos pais na escala do Índice Socioeconômico Internacional (ISEI, sigla em inglês); e a posse ou consumo de bens duráveis para medir a renda familiar, uma vez que indica o seu componente permanente. Esses índices são sintetizados em uma escala estimada por intermédio de um modelo da Teoria de Resposta ao Item (TRI). O emprego da TRI é um diferencial em relação aos métodos tradicionais pela forma como os dados ausentes são tratados. A TRI considera apenas os itens para os quais os estudantes forneceram respostas para estimar os escores. Desse modo, todos os estudantes têm uma medida de ESCS, mesmo que tenham deixado de responder algum item do questionário.

O Brasil possui vasta experiência em produzir medidas de NSE nas pesquisas acadêmicas que utilizam dados produzidos pelos questionários das avaliações e exames educacionais, também empregando modelos da TRI (Soares, 2005; Soares; Andrade, 2006; Alves; Soares, 2009; Alves; Soares; Xavier, 2014; Barros et al., 2019). O Inse do Inep, aplicado aos dados do Saeb e/ou do Enem, foi desenvolvido a partir desse conhecimento acumulado.

No Saeb, os questionários respondidos por estudantes incluem itens sobre a educação dos pais (anos de estudos e/ou condição de alfabetização), a posse de bens duráveis (geladeira, televisão, automóveis etc.) e a contratação de serviço doméstico como medidas indiretas da renda familiar. Entretanto, a ocupação dos pais não é perguntada diretamente devido à dificuldade de se coletar essas informações com o desenho dessa avaliação (Franco et al., 2003). Assume-se que o índice não observado (prestígio ocupacional) é correlacionado com os índices observados (educação e renda indireta) e, por isso, pode ser usado como um dos indicadores que contribuem para o cálculo do NSE.

No Enem, o questionário é respondido por estudantes concluintes do ensino médio ou indivíduos que já terminaram a educação básica. Ele inclui itens sobre o nível de escolaridade do pai e da mãe, a renda familiar direta (faixas de salários-mínimos) e indireta (posse de bens e contratação de serviços no domicílio). Em edições mais recentes, com a

adoção do questionário online, foram apresentadas questões fechadas sobre a ocupação do pai e da mãe, com a adequada explicação para o seu preenchimento, o que não é viável no formulário em papel. Os respondentes devem escolher uma opção entre grupos de ocupações organizadas em uma escala ordinal, que se parece com uma versão simplificada do Índice Socioeconômico Internacional (ISEI), utilizado no PISA. Esse modelo de aplicação também reduz os dados ausentes, que ocorrem somente quando o estudante responde “não sei” para algum item.

A condição de pobreza das famílias é também utilizada como *proxy* para a origem socioeconômica dos estudantes em pesquisas sociais. Os índices de pobreza desenvolvidos com dados populacionais são correlacionados com os índices de NSE das pesquisas educacionais (Stubbe; Buddenberg, 2010). Por isso, nos Estados Unidos e em outros países, é comum o uso de dados sobre os estudantes elegíveis para receber almoço gratuito ou a baixo custo na escola, devido à condição de pobreza de sua família, como *proxy* para o seu NSE (Harwell; Lebeau, 2010; Stubbe; Buddeberg, 2010; Domina et al., 2018). Essa é uma informação administrativa acessível e sem dados ausentes, que aparece em cerca de um quarto das pesquisas educacionais norte-americanas (Harwell, 2019).

Entretanto, há controvérsias na literatura sobre a adequação de se tratar a pobreza e o NSE como medidas intercambiáveis. Dentre elas, se destacam: (i) a pobreza é quase sempre definida pelo nível de renda, enquanto o NSE é uma medida composta que mensura outros ativos que conferem vantagens econômicas, sociais e culturais aos indivíduos ou grupos na sociedade; (ii) a condição de pobreza, que define os elegíveis e não elegíveis para o benefício na alimentação escolar, praticamente divide os estudantes em dois grupos, mas a distinção dentro dos grupos é fraca; (iii) as categorias dos grupos elegíveis capturam relativamente pouca variação nos rendimentos familiares; (iv) há pouca participação no programa de estudantes do ensino médio, especialmente da última série, por razões que não são muito claras; e (v) há uma taxa média de erro na classificação dos estudantes em torno de 20%, parte disso porque a solicitação é feita por algum funcionário da escola e não pela família (Harwell; Lebeau, 2010; Domina et al., 2018; Harwell, 2019).

Apesar disso, a medida agregada dos beneficiários do programa de alimentação por escola fornece um índice útil sobre as necessidades socioeconômicas de uma escola em relação às outras (Harwell, 2019). Em vez do uso isolado sobre a posição socioeconômica do estudante, essa variável pode contribuir para o índice composto do NSE, agregando às variáveis estudantis outras informações sobre o contexto escolar e o seu entorno (Stubbe; Buddeberg, 2014; Harwell, 2019).

No Brasil, por meio do Cadastro único (CadÚnico), é possível saber se o estudante é beneficiário do Programa Bolsa Família – programa federal para as famílias em condição de extrema pobreza – que tem, entre as suas condicionalidades, a frequência regular à escola. Essa variável não foi ainda utilizada para o cálculo do NSE de estudantes, possivelmente devido às dificuldades de se associar o código do estudante no Saeb com o código dos indivíduos matriculados em escolas na base de dados do CadÚnico. Existem

trabalhos que utilizaram essa informação como variável de controle em modelos estatísticos com dados educacionais, com outras variáveis de origem social (Santarrosa, 2011; Camargo; Pazello, 2014; Bartholo; Costa, 2016; Bartholo et al., 2020). O percentual de estudantes no cadastro do Programa Bolsa Família por escola já foi utilizado pelo Inep para estimar o Inse de escolas sem dados de questionários contextuais, conforme explicaremos na seção “NSE de escolas” (Inep, s/data).

As críticas ao uso de dados sobre pobreza como único indicador da posição socioeconômica dos estudantes se aplicam ao Brasil. A elegibilidade para o programa Bolsa Família, baseada na renda per capita das famílias, produz uma descrição unidimensional e dicotômica da hierarquia social subjacente, com foco nos estudantes de famílias em situação de extrema pobreza, sem distinguir os demais. A participação da população estudantil de escolas privadas e parte das públicas urbanas nesse programa é menor, ainda que essas escolas recebam matrículas de estudantes de origem social desfavorecida, alguns bem próximos do limiar de extrema pobreza. A possibilidade de erro na classificação dos programas de transferência de renda foi discutida por Firpo e colaboradores (2014) a partir de simulações feitas com dados da Pesquisa Nacional por Amostra de Domicílios (Pnad) de 2006, porém pode ter diminuído ao longo dos anos, com o aperfeiçoamento do CadÚnico (WWP, 2014).

Em termos pragmáticos, a produção de medidas de NSE depende da disponibilidade de dados que tenham relevância conceitual/teórica (Sirin, 2005; Alves; Soares, 2009; Harwell, 2019; Avvisati, 2020). A forma como os itens dos questionários são sintetizados em uma medida composta de NSE – ou se os itens serão utilizados separadamente – depende de escolhas metodológicas dos pesquisadores, mas também da qualidade das informações (Alves; Soares; Xavier, 2014; Broer; Bai; Fonseca, 2019). Os dados ausentes nos questionários são um desafio não trivial (Broer; Bai; Fonseca, 2019; Ferrão; Prata; Alves, 2020), que requer metodologias adequadas para sua superação (Rose; von Davier; Xu, 2010).

NSE de escolas

O NSE da escola é uma caracterização quantitativa de sua composição social, frequentemente calculado pela média aritmética do NSE de seus estudantes. Essa medida ganhou destaque a partir da publicação do Relatório Coleman, o qual revelou que a composição social das escolas explica mais a variação no desempenho dos estudantes do que os atributos escolares (Coleman et al., 1966). Na época, isso expôs as limitações das políticas liberais para promover a igualdade de oportunidades educacionais e influenciou o desenvolvimento das pesquisas sobre eficácia escolar nas décadas seguintes (Brooke; Soares, 2008).

Desde então, o NSE da escola se tornou, talvez, a variável contextual mais popular nos estudos educacionais devido ao seu efeito significativo e persistente nos resultados

escolares (Thrupp; Lauder; Robinson, 2002; Harker; Tymms, 2004; Willms, 2010; Soares; Alves; Xavier, 2016; Sciffer; Perry; McConney, 2020, 2022). Nas pesquisas sobre eficácia escolar, há evidências de que as práticas escolares podem ter efeitos diferenciados dependendo da composição socioeconômica da escola (Palardy, 2008; Timmermans; Thomas, 2015), de forma que o controle dessa variável se tornou crucial para a análise das políticas e práticas escolares (Lee, 2008).

Contudo, a metodologia convencional para produzir essa variável de composição não se aplica ao problema desta pesquisa. Há muitas escolas que, por motivos variados, não participam das avaliações educacionais que fornecem os dados para estimar o NSE dos estudantes. Dessa forma, não é possível produzir diretamente uma medida agregada do NSE dessas escolas.

Os dados administrativos sobre estudantes elegíveis por programas sociais são um indicador sociológico de pobreza e acrescentam informação preciosa aos outros indicadores que compõem o NSE, mas não os substituem. Como discutido na literatura internacional (Stubbe; Buddeberg, 2014; Harwell, 2019), eles podem melhorar as medidas compostas de NSE para descrever as necessidades socioeconômicas das escolas.

O Inep utilizou informação sobre estudantes beneficiários do Programa Bolsa Família para estimar o Inse de escolas públicas que não tinham informações de questionários respondidos por estudante por meio de uma técnica de imputação de dados ausentes (Inep, s/data). Para isso, foi ajustado um modelo de regressão linear múltipla, considerando o Inse calculado pelo Inep com dados do Saeb/Prova Brasil e Enem de 2011, o percentual de estudantes que recebiam Bolsa Família por escola, a dependência administrativa da escola, sua localização urbana ou rural e a unidade da Federação. Essa metodologia foi empregada para contextualizar os resultados da Avaliação Nacional da Alfabetização (Ana), em 2013, e das escolas que participaram da Prova Brasil/Saeb, no mesmo ano, mas não tiveram o Inse calculado na edição anterior. Entretanto, a imputação só pôde ser feita para escolas que tinham estudantes beneficiários do Programa.

Apresentar uma metodologia adequada que, uma vez implementada, supera essa limitação, é a principal contribuição deste artigo.

Metodologia: dados e procedimentos

Os indicadores usados neste artigo para o cálculo do NSE das escolas são oriundos da discretização de variáveis que pertencem a três categorias. A primeira, que chamaremos de NSE primário da escola, é a média do NSE dos estudantes calculado a partir das respostas aos itens dos questionários contextuais do Saeb e/ou Enem, entre 2007 e 2019. O NSE dos estudantes foi calculado para cada edição e cada etapa escolar separadamente. Considerando as edições do Saeb (e etapas avaliadas) e as edições do Enem, foram calculados 26 NSE primários. Cada escola pode ter estimativas para vários desses NSE primários, dependendo da participação de seus estudantes no Saeb e no Enem.

A segunda categoria é constituída por indicadores de características sociais, administrativas e territoriais das escolas. Inclui a variável que informa o percentual de estudantes na escola que estão na base do Cadastro Único (CadÚnico) por serem beneficiários do programa Bolsa Família, a localização da escola (urbana ou rural), a rede (pública ou privada) e uma variável que informa o nível de renda do município (IDH-Renda).

A terceira categoria é constituída por apenas um indicador obtido usando-se a informação gerada pelos estudantes que se transferiram entre escolas, durante sua trajetória educacional, entre 2007 e 2019, identificada no Painel Longitudinal do Censo Escolar do Inep. A grande maioria dos estudantes frequentou, em algum momento de sua trajetória escolar, escolas para as quais o NSE primário é conhecido. Esse fato é usado, como explicado abaixo, para a construção de uma medida que será referida como NSE secundário da escola.

Detalhes sobre a construção desses indicadores e os procedimentos empregados para a construção do NSE final estão disponíveis em nota técnica⁵.

NSE primário: indicadores provenientes do Saeb e do Enem

Neste artigo, utilizamos os microdados de questionários contextuais do Saeb e do Enem, baixados diretamente do site do Inep entre setembro e dezembro de 2021. Estão incluídos os estudantes que fizeram o Saeb nos anos ímpares entre 2007 e 2019 e os participantes do Enem das edições realizadas nesses mesmos anos.

Em relação ao Saeb, houve alterações na cobertura dos dados disponíveis nesse período e na denominação da avaliação⁶. Porém, para simplificar a exposição, utilizaremos neste artigo somente a sigla Saeb. Os questionários dessa avaliação são respondidos pelos estudantes do último ano de cada etapa do ensino fundamental e do ensino médio. Entretanto, há escolas cujos estudantes participam do Saeb, mas o seu código não é divulgado. Essas escolas são denominadas de “mascaradas”. Isso ocorre para as escolas privadas, que participam somente da parte amostral do Saeb. No entanto, a partir de 2013, uma parte das escolas públicas também teve os seus códigos mascarados por decisões administrativas diversas, conforme registradas na documentação que acompanha os

⁵ A nota técnica com detalhes dos procedimentos metodológicos deste artigo está disponível em: <<https://drive.google.com/drive/folders/125IS42t0EcNdDnDsuGIwKSqGcmK95JLU?usp=sharing>>. Acesso em: 04 dez. 2023.

⁶ Em 2007 e 2009, os dados se referem à Prova Brasil, uma avaliação quase censitária de escolas públicas de ensino fundamental vinculada ao Saeb e que utiliza os mesmos instrumentos (testes e questionários). Em 2011, o Inep disponibilizou os dados da Prova Brasil e da amostra do Saeb, que inclui escolas públicas não elegíveis para a Prova Brasil, escolas privadas, além da avaliação do ensino médio em escolas públicas e privadas. A partir de 2013, os dados da Prova Brasil e do Saeb passaram a ser divulgados conjuntamente e o banco de dados possui uma variável que identifica se a escola é elegível para a Prova Brasil. Em 2019, o nome Prova Brasil deixou de ser utilizado e um sufixo que identifica a “série” distingue as avaliações do Saeb. Informações sobre o Saeb estão disponíveis em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb/historico>>. Acesso em: 04 jan. 2023.

microdados. Neste artigo, usamos as informações fornecidas pelos estudantes das escolas mesmo com código mascarado, pois elas diversificam os perfis de escolas consideradas, melhorando a robustez das estimativas.

No Enem, o preenchimento do questionário contextual faz parte do processo de inscrição. Nas edições mais recentes, o processo é todo realizado online. Dentre os inscritos no Enem, selecionamos os casos de concluintes do ensino médio no mesmo ano da realização do exame.

Inicialmente, foi calculado o NSE dos estudantes, em cada uma dessas bases de dados, utilizando-se o modelo Samejima adequado a respostas politômicas (Samejima, 1969). Os procedimentos foram semelhantes aos empregados em outros trabalhos com dados educacionais (Soares; Andrade, 2006; Alves; Soares; Xavier, 2014). É plausível supor que um mesmo estudante possa estar presente em mais de uma dessas avaliações, na mesma escola ou em outra. Entretanto, como o objetivo é estimar o NSE primário da escola em cada avaliação separadamente, isso não enviesou as análises.

Em seguida, produzimos uma média do NSE dos estudantes de cada escola em cada pesquisa (Saeb e Enem), etapa (5º ano, 9º ano do ensino fundamental e 3ª/4ª série do ensino médio; ou concluintes) e ano de aplicação. Essa média será referida como o NSE primário da escola. O cálculo do NSE em cada avaliação separadamente permite preservar ao máximo a capacidade de discriminação dos itens e a sua relevância em cada situação. Por exemplo, as variáveis sobre o nível de escolaridade do pai e da mãe aparecem, nos diferentes questionários, com categorias de respostas distintas e a padronização resulta em uma simplificação que pode reduzir o seu poder de discriminação em algumas etapas de ensino ou anos de aplicação.

A estimação do NSE com dados de várias edições do Saeb e/ou Enem conjuntamente tem a vantagem de produzir uma medida equalizada ao longo do tempo, como já foi feito (Alves; Soares; Xavier, 2014). Entretanto, essa solução não é adequada para lidar com o problema da expansão do número de escolas para as quais o NSE está disponível tratado neste artigo. O trabalho de Paula (2020), que analisou a estabilidade do NSE dos estudantes e escolas com dados longitudinais de cinco anos, permite inferir que é mínima a diferença nos resultados dos escores estimados pelas duas metodologias. Ele mostrou que o NSE dos estudantes e das escolas é relativamente estável, a despeito da mobilidade estudantil e da metodologia de estimação – se ano a ano ou em conjunto para todos os anos.

Ao final, há 95.463 escolas com códigos Inep identificados e com NSE primários calculados⁷. Como cada NSE primário foi calculado de forma independente, foi necessário padronizá-los para que variassem no mesmo intervalo.

⁷ Calculamos também os escores de NSE primários para 53.227 escolas com código mascarados. A informação produzida para essas escolas será utilizada na última etapa deste artigo.

Características das escolas

Há características sociais, administrativas e territoriais das escolas que informam sobre o seu contexto socioeconômico e que foram utilizadas para estimar o NSE das escolas cujos NSE primários não puderam ser calculados.

A primeira característica é o nível de renda do município e reflete o contexto macrossocial no qual a escola está situada. A variável do Índice de Desenvolvimento Humano-Renda (IDH-Renda), referente a 2010, apesar de desatualizada e sem informação para municípios criados após o Censo Demográfico de 2010, descreve relativamente bem as diferenças macrossociais do país, principalmente fora dos grandes centros urbanos, onde há mais escolas sem o NSE primário. É razoável assumir que os estudantes mais pobres de um município pobre são mais pobres do que os pobres de municípios mais ricos. Por isso, espera-se que as escolas situadas em municípios com alto IDH tenham, tipicamente, maior valor do NSE.

Em seguida, consideramos a localização e a dependência administrativa da escola, conforme registradas nas respectivas bases de dados. As escolas rurais têm tipicamente menor NSE do que as urbanas, enquanto as escolas privadas, em geral, têm NSE mais alto do que as públicas.

Finalmente, a terceira característica é o percentual de estudantes com Bolsa Família na escola. Essa variável é particularmente importante para se caracterizar as escolas pequenas e rurais, que são aquelas que menos participam das avaliações educacionais por não atingirem os critérios de elegibilidade. Portanto, elas não possuem dados para a estimação de seu NSE e se tornam invisíveis para as políticas de redução das desigualdades no nível das escolas. Os dados se referem ao ano de 2018 e foram cedidos pela Secretaria Nacional de Renda de Cidadania para uso em pesquisa⁸.

Cada uma dessas informações, isoladamente, informa pouco sobre o NSE da escola, mas, somada às outras informações, acrescenta solidez para o cálculo do NSE.

NSE secundário: informação das transferências de estudantes

Os diferentes NSE primários das escolas são obtidos com a informação fornecida pelos respectivos estudantes. Já os indicadores da segunda categoria são características estruturais da escola, obtidos sem a interferência de seus estudantes.

Esses dois conjuntos de indicadores estão conectados pela variável referente ao programa Bolsa Família, disponível tanto para a maioria das escolas que tiveram seu NSE primário calculado quanto para as outras. No entanto, essa vinculação tem alcance limitado, principalmente entre as escolas privadas que, tipicamente, não têm estudantes atendidos pelo programa Bolsa Família. Uma ligação mais sólida entre esses dois conjuntos

⁸ Agradecemos o apoio de Tiago Falcão Silva, da Secretaria Nacional de Renda de Cidadania, que nos orientou para obter a autorização de uso desses dados.

de informações foi obtida com aquelas fornecidas pelos estudantes que se transferiram entre escolas, dado disponível no painel longitudinal do Censo Escolar preparado pelo Inep e já usado em outros trabalhos (Rigotti; Hadad, 2018; Soares; Alves; Fonseca, 2021; Fonseca et al., 2022)⁹. A metodologia usada é explicada aqui sucintamente. Informações detalhadas estão disponíveis na nota técnica.

O Inep realiza anualmente o Censo Escolar da Educação Básica, que abrange todas as escolas públicas e privadas de todos os municípios do país. Desde 2007, o Censo registra a matrícula de cada estudante individualmente em cada escola. Após nove anos de coleta de dados individualizados, o Inep preparou um painel longitudinal, que acompanha, ano a ano, a situação dos estudantes matriculados, e que vem sendo atualizado anualmente.

Os estudantes mudam de escolas principalmente por razões estruturais, que são explicadas pela divisão das competências entre municípios, estados e união, o que leva à necessidade de mudança de escolas entre níveis de ensino. Há também redes que organizam as etapas do ensino fundamental em escolas distintas e as transferências ocorrem dentro do nível. Malaguth (2022) analisou dados do painel longitudinal do Censo Escolar em Minas Gerais e observou que, entre 2008 e 2019, mais de 90% dos estudantes que ingressaram no sistema de ensino com seis anos de idade mudaram de escola pelo menos uma vez nesse período. Além disso, todos os anos, muitas escolas deixam de funcionar e os seus estudantes são transferidos para outras no ano seguinte.

O painel longitudinal acompanhou cada estudante durante treze anos. Tomando-se o NSE da escola em que o estudante esteve matriculado como *proxy* de seu NSE, o painel fornece, portanto, até 13 *proxies* para o NSE de cada estudante. O valor máximo ocorre quando o estudante esteve matriculado em diferentes escolas, com NSE primário calculado em cada um dos anos-calendário. A média desses valores foi tomada como o NSE secundário dos estudantes. Esse valor depende do número de escolas nas quais o estudante esteve matriculado e que tenham o NSE primário calculado.

O NSE secundário de cada escola é tomado como a média dos NSE secundários dos seus estudantes. Esse valor é muito próximo ao NSE primário das escolas com muitos alunos, uma vez que as transferências impactam pouco para essas escolas. No entanto, as escolas que não tinham NSE primário calculado passam a contar com uma estimativa dele, possibilitada pelo NSE secundário dos seus estudantes que se transferiram. Um exemplo ilustra a situação. Os alunos de uma escola privada que oferece somente o ensino fundamental e não tem NSE primário se transferem para outras escolas privadas do município, que têm NSE primário devido à participação de seus estudantes no Enem. No outro extremo estão as escolas rurais. Embora elas tenham a informação do programa

⁹ Os dados do painel longitudinal foram obtidos junto ao Centro de Desenvolvimento e Planejamento Regional (Cedeplar) da Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais (UFMG), que, por meio de um acordo de cooperação entre o Inep e a UFMG, possui o acesso a esses dados para fins de pesquisa. Agradecemos ao professor José Irineu Rangel Rigotti, do Cedeplar, por preparar a base de dados sem o código de estudante utilizada neste artigo.

Bolsa Família, seus estudantes vão também para outras escolas, com o NSE primário conhecido, o que permite melhorar a estimativa de seu NSE.

O NSE secundário pode ser calculado para cada um dos 26 NSE primários de cada escola, mas utilizar a estimativa do NSE secundário de uma escola, obtida a partir de poucos estudantes transferidos, seria um indicador muito frágil. Assim, o NSE secundário da escola foi definido como a média dos 26 NSE secundários em cada avaliação e ano. Esse indicador de NSE secundário, ao associar de forma sólida os indicadores da primeira e da segunda categorias, permitiu que a potência das informações de ambas fosse considerada para todas as escolas no cálculo do NSE final.

Banco de dados para cálculo do NSE final

O banco de dados utilizado para cálculo do NSE é constituído dos indicadores gerados pela discretização das variáveis descritas nas seções anteriores. Ele contém as escolas cujos códigos são conhecidos e as escolas mascaradas, mantidas porque os seus respectivos estudantes acrescentam informações para o cálculo do NSE primário e sabemos a sua localização (urbana ou rural), a rede (pública ou privada) e o IDH-Renda, quando o município não está mascarado. Mas elas não são mantidas no banco de dados final com a estimação do NSE de todas as escolas.

As primeiras 26 colunas do banco de dados referem-se aos NSE primários das escolas. O número de escolas com dados ausentes para alguma dessas 26 variáveis gira em torno de 80%, mas a maior parte deles são lacunas estruturais: a escola não oferece a etapa avaliada; a escola não participou da avaliação (em geral, escola privada que atende somente ao ensino fundamental e/ou ao ensino infantil); ou a escola não tem informação para uma etapa do ensino fundamental porque os escores primários, calculados para cada etapa, estão apresentados em colunas (variáveis) distintas no banco de dados. As escalas dessas variáveis estão em desvio-padrão.

A variável IDH-Renda, referente ao município onde a escola está localizada, está ausente para apenas 0,4% das escolas, que são aquelas localizadas em municípios criados após 2010. A escala original, no intervalo de $[0; 1]$, foi transformada em escores padronizados (desvio-padrão).

Duas variáveis estruturais da escola associadas ao construto, a localização (rural ou urbana) e a rede (pública ou privada), não possuem dados ausentes.

Uma variável que informa o percentual de estudantes da escola cuja família recebe o benefício do programa Bolsa Família pela condição de extrema pobreza: quanto maior o percentual, mais desfavorecido é o contexto social da escola. Por isso, a escala original foi invertida e depois transformada em escores padronizados. Essa variável possui 24,8% de dados ausentes, que se referem às escolas que não estavam funcionando no ano de referência (2018).

A variável com o NSE secundário das escolas calculado possui dados ausentes apenas para 1,8% das escolas com código identificado. Isso reflete o fato de que a grande maioria dos estudantes muda de escola durante sua trajetória na educação básica, conforme observado nas pesquisas sobre trajetórias escolares com dados administrativos (Rigotti; Hadad, 2018; Soares; Alves; Fonseca, 2021; Malaguth, 2022). Essa variável também está em uma escala padronizada.

Todas essas variáveis foram discretizadas usando classes definidas pelo desvio-padrão, com exceção das variáveis rede e localização, que foram tratadas como indicadoras. A primeira classe agrega os valores abaixo de -1,5 desvios-padrão. Os outros intervalos têm tamanhos iguais de meio desvio-padrão. A última classe engloba as escolas com desvios-padrão acima de 1,5 ou 2, dependendo da amplitude da variável. O número de classes varia entre 7 e 9.

Modelo de análise

Para calcular o NSE primário dos estudantes, utilizamos um modelo da Teoria de 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 uma medida do NSE mesmo com dados incompletos (Rose; von Davier; Xu, 2010). Nas avaliações educacionais, esse é um problema comum. Por razões diversas, os estudantes deixam sem respostas alguns itens dos questionários. Mesmo assim, a TRI consegue calibrar os itens com respostas faltantes com as respostas observadas para estimar o traço latente. Essa propriedade da TRI funciona como um eficiente método de imputação de dados omissos, que é apropriado para problemas que envolvem a estimação de um traço latente (May, 2006; Paniagua et al., 2017).

O modelo Samejima da TRI, especialmente construído para situações em que os indicadores são variáveis ordinais, foi empregado para estimar tanto o NSE dos estudantes (Samejima, 1969) como o NSE final das escolas. Essa classe de modelos tem como pressuposto a unidimensionalidade, isto é, a existência de um único construto latente dominante no conjunto de dados, o que deve ser verificado antes do ajuste do modelo.

A heurística para a estimação dessa medida envolveu três fases. Para estimar o NSE dos estudantes, inicialmente, testamos o pressuposto de unidimensionalidade do modelo em cada conjunto de itens de cada edição/avaliação. Depois, na aplicação da TRI, foram estimados os parâmetros dos itens. Finalmente, os escores individuais foram estimados para cada estudante, empregando-se os parâmetros estimados na fase anterior. Esses procedimentos foram conduzidos utilizando-se o pacote *mirt* do software R (Chalmers, 2012).

Para produzir os NSE primários das escolas, os NSE dos estudantes foram agregados pela média em escola, etapa e edição da avaliação. Em seguida, os diferentes indicadores (os 26 NSE primários, as quatro variáveis contextuais das escolas e o NSE secundário) foram agregados em um NSE final. Empregamos novamente o mesmo modelo Samejima da TRI com auxílio do pacote *mirt*, tal como fizemos para estimar o NSE primário dos estudantes.

O banco de dados final contém 185.616 códigos das escolas que tiveram o seu NSE final calculado e o respectivo erro-padrão¹⁰. A escala do NSE, em escores padronizados, foi mantida, mas ela poderá ser transformada para um intervalo de valores positivos [0; 10] para facilitar sua utilização pelo público não acadêmico.

Além disso, o banco de dados contém as variáveis: região, UF, município, localização e dependência administrativa da escola. Criamos uma variável chamada "grupo", que categoriza a qualidade da informação empregada na estimativa, isto é, se a escola teve algum escore primário calculado ou somente o escore secundário. Além disso, entre 2007 e 2019, muitas escolas deixaram de funcionar. A variável "situação da escola no Censo Escolar 2020" informa se a escola está em funcionamento regular, paralisada (escola sem matrícula no ano), em funcionamento para atendimento especializado ou ausente (escola fechada), mas com o NSE que pode ser útil para pesquisas longitudinais.

As análises descritivas (tabelas e gráficos) apresentadas na seção "Resultados" foram produzidas com auxílio do software SPSS, versão 22. O mapa foi produzido por meio do Geopandas, uma biblioteca *open source* do Python.

Resultados

Nesta seção, descrevemos os resultados obtidos e mostramos a associação do NSE das escolas com variáveis de resultados no Saeb e segundo o Ideb.

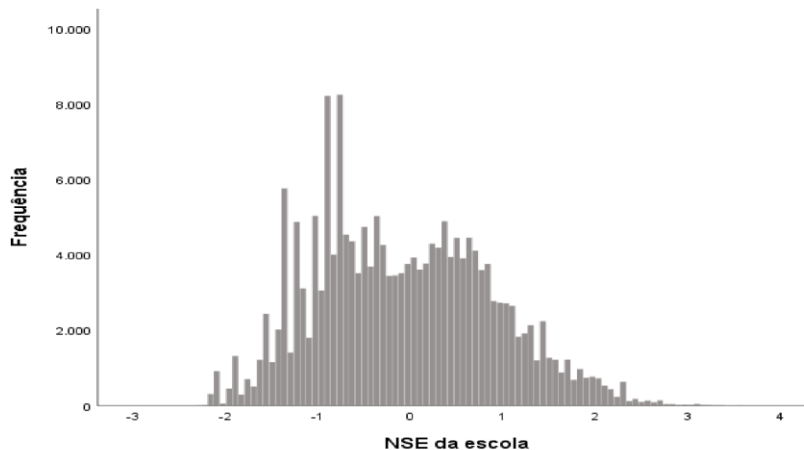
Antes, no entanto, é preciso uma explicação sobre o impacto do mascaramento de escolas. Como explicado, o primeiro passo da metodologia introduzida neste artigo é o cálculo de uma estimativa do NSE dos estudantes de cada escola. Nessa etapa, todos os estudantes que preencheram os questionários são considerados e a informação de mascaramento ou não do código da respectiva escola não é relevante. No entanto, esses estudantes não são considerados para o cálculo do NSE de nenhuma escola, pois não se sabe a qual escola eles pertencem. Assim sendo, se uma escola teve seu código divulgado em uma edição do Saeb e mascarado em uma edição posterior, seu NSE final será fruto de seus NSE primários parcialmente disponíveis, das informações contextuais e do NSE secundário, ou seja, diminui-se a precisão da estimativa para essas escolas. Essa limitação pode ser superada se o procedimento for replicado com bases de dados sem códigos de escolas com máscara.

¹⁰ O banco de dados com o NSE das escolas está disponível em: <https://drive.google.com/file/d/1-ViCfscVYycp2sbpG2LJCz5KpP5Mymus/view?usp=drive_link>. Acesso em 04 dez. 2023.

Descrição

O histograma do NSE final das escolas é apresentado na Figura 1. A média do NSE final, calculado para 185.616 escolas, é de -0,083 e o desvio-padrão de 0,96. Deve-se observar que a cauda direita é mais longa do que a esquerda, refletindo o fato conhecido de que as escolas cujos estudantes têm NSE mais alto são em menor número, mas com valores mais heterogêneos. Por outro lado, há muitas escolas que atendem a estudantes com NSE mais baixo.

Figura 1 - Histograma do NSE das escolas



Fonte: elaboração própria com dados do Inep.

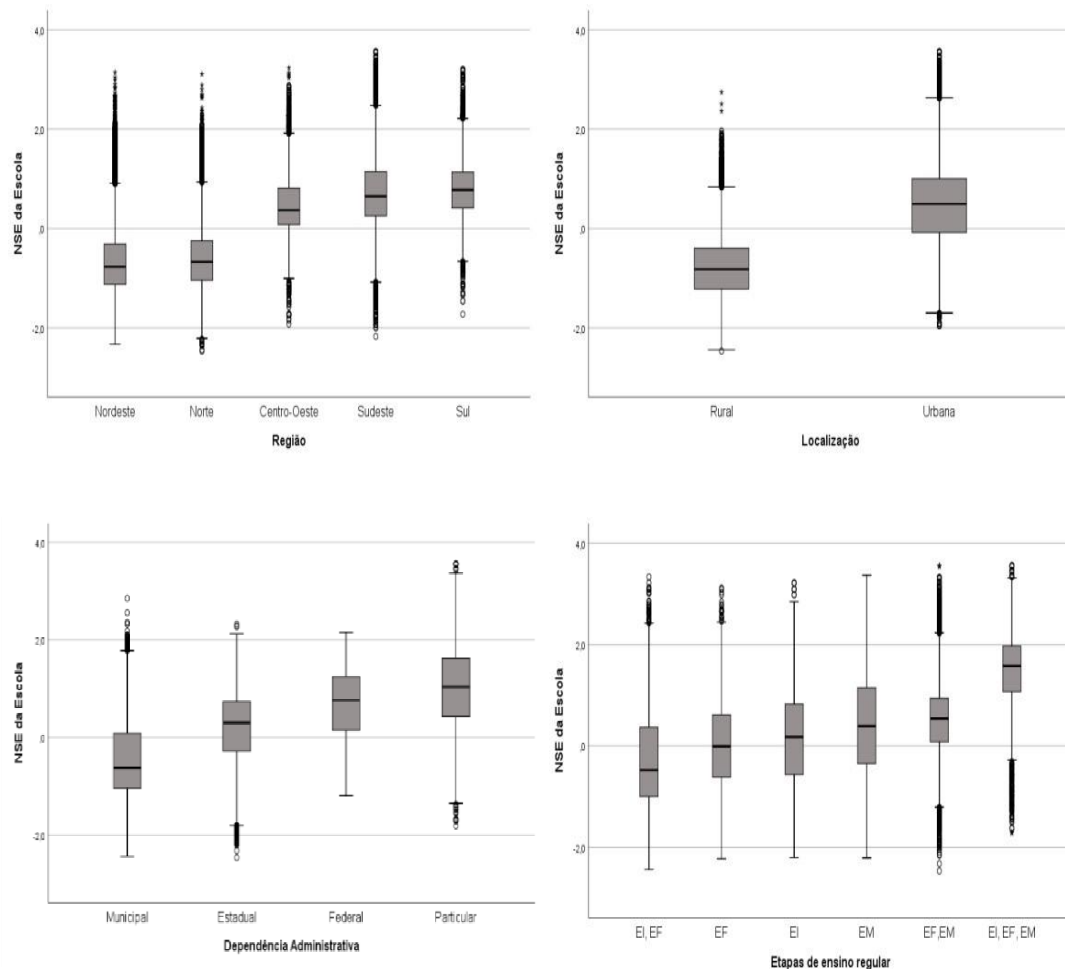
Uma consulta ao arquivo final mostra que os maiores valores de NSE vêm de escolas privadas, amplamente conhecidas nos municípios onde se localizam por atenderem à elite econômica das diferentes cidades. Esse fato pode ser tomado como uma validação idiossincrática dos resultados.

Uma validação estatística foi feita por meio da correlação de Pearson entre o NSE das escolas e o Inse 2019, calculado pelo Inep. Das 68.868 escolas que estão na planilha divulgada pelo Inep, somente para 17 delas os NSE com a metodologia deste artigo não foram calculados. O coeficiente de correlação entre as escolas presentes nas duas bases de dados é 0,93. Esse resultado mostra que as duas metodologias fornecem, essencialmente, a mesma informação.

Também analisamos o NSE das escolas segundo variáveis estruturais e administrativas. Como se observa nos gráficos de caixa reunidos na Figura 2, os padrões de distribuição do NSE nos quatro gráficos confirmam o que é conhecido da literatura educacional: as escolas do Nordeste, do Norte e em localização rural têm NSE mais baixos

em contraste com os valores das escolas das outras regiões e localizadas em áreas urbanas. Há um gradiente crescente entre as redes: as escolas municipais com os escores mais baixos; e as particulares com os mais altos. Em relação aos níveis de ensino, as escolas que ofertam ensinos infantil e fundamental têm NSE mais baixos enquanto os mais altos são das escolas que ofertam as três etapas, o que geralmente ocorre nos colégios privados. Registra-se a presença de NSE *outliers* em quase todas as categorias nos gráficos. Isso reflete o fato de que há escolas com composição socioeconômica mais ou menos privilegiada mesmo em contextos onde isso não é a regra.

Figura 2 - Boxplot com o NSE das escolas, por Região, Localização, Dependência Administrativa e tipo de oferta



*Observação: no gráfico por etapas de ensino, as categorias menos frequentes foram excluídas. As etapas se referem à oferta em 2020.

Fonte: Elaboração própria com dados do Inep.

A interpretação dos valores do NSE das escolas deve considerar que os escores originais foram reduzidos ao intervalo $[-2, 4]$. Assim sendo, ao se verificar o NSE de escolas de um único município, é preciso considerar que os respectivos valores de NSE estão mais próximos entre si, refletindo o fato de que a dispersão das medidas é muito maior quando todas as escolas do país são consideradas conjuntamente. Ou seja, uma escola de uma região mais rica, considerada localmente como de baixo NSE, pode não ter um NSE tão baixo quando todas as escolas do país são consideradas. Isso reflete, também, o fato de que a medida de NSE das escolas é relacional, capta a posição relativa das escolas em um *continuum*.

A precisão das estimativas obtidas varia conforme o número de indicadores disponíveis para cada escola. Os NSE das escolas calculados com indicadores nas três categorias têm maior precisão. A menor precisão é dos NSE das escolas que só podem ser estimados pelo NSE secundário, calculado pela média dos NSE primários das escolas de destino de seus estudantes e pelas características gerais. De uma forma geral, quando há algum NSE primário, a dispersão é menor.

A Tabela 1 mostra as estatísticas descritivas dos desvios-padrão das estimativas, considerando o fato de a escola possuir ou não informação de NSE primário, representadas na tabela como Grupo 1 e Grupo 2. A média dos desvios-padrão do grupo 1 (escolas cujo NSE foi calculado com informação de algum NSE primário) é a metade do valor do grupo 2 (escolas cujos escores foram calculados sem essa informação).

Tabela 1 - Estatísticas descritivas do desvio-padrão do NSE das escolas por grupo de disponibilidade de dados

Grupo	Média	Nº de escolas	Desvio-padrão	Mínimo	Máximo
1 - Escola possui algum NSE primário	0,132	95.533	0,052	0,059	0,634
2 - Escola possui só NSE secundário	0,279	90.083	0,029	0,227	0,683
Total	0,204	185.616	0,085	0,059	0,683

Fonte: Elaboração própria com dados do Inep.

A diferença entre os dois grupos na Tabela 1 se explica, no caso das escolas públicas do grupo 2, pelo fato de que elas são tipicamente rurais ou de pequenos municípios e não possuem o NSE primário; a informação sobre a sua composição social é indireta, vem do percentual de estudantes com Bolsa Família, do NSE secundário e das outras características estruturais (IDH-Renda municipal, rede, localização). No caso das escolas privadas nesse mesmo grupo, geralmente são estabelecimentos que não ofertam o ensino médio e por isso não têm informações do Enem para o cálculo do NSE primário. A variável sobre o Bolsa Família acrescenta pouca informação para essas escolas privadas, pois geralmente elas não têm estudantes com esse perfil. A informação existente é basicamente uma média dos NSE primários das escolas de destino de seus estudantes,

que não capta a eventual presença de estudantes de maior nível socioeconômico, e as características estruturais.

Importante registrar que, antes dos resultados relatados neste artigo, a única opção analítica era a exclusão dessas escolas. Essa decisão é pior do que a consideração de uma medida que, permite classificar essas escolas nas mesmas categorias de análise usadas para as escolas que têm NSE primário, ainda que com menor precisão.

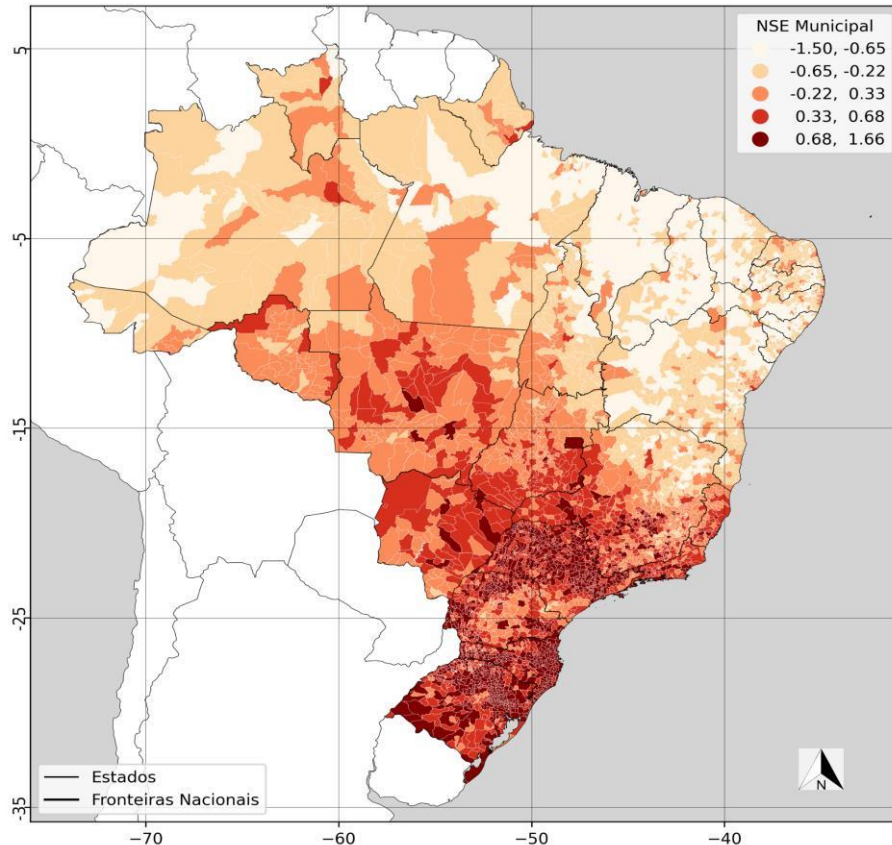
O NSE dos municípios

Definimos o NSE de cada município por meio da média do NSE das escolas ponderada pelo número de estudantes utilizados para o cálculo do NSE primário e do NSE secundário, por município. Com esse procedimento, calculamos o NSE municipal para 5.570 municípios, dois a mais do que o número de unidades com Inse municipal calculado pelo Inep. A correlação entre o NSE municipal, calculado neste artigo, e o Inse oficial do Inep é 0,97, ou seja, ambos mensuram o mesmo fenômeno.

O cartograma da Figura 3 mostra a representação espacial do NSE municipal no território brasileiro. A escala do índice foi dividida em quintis, representados pelo gradiente de tons, do mais claro (menor NSE) ao mais escuro (maior NSE).

O Brasil descrito pelo NSE municipal é muito desigual. As regiões Norte e Nordeste se destacam pelo predomínio de tons mais claros que representam o NSE mais baixo de seus municípios. Na região Norte não há municípios no quintil mais alto e apenas sete estão no quarto quintil. No Nordeste, há apenas um município no quintil mais alto e oito no quarto. O único caso com NSE mais alto nessa região é um *outlier*: a ilha de Fernando de Noronha (não está no mapa - a ilha é um distrito estadual de Pernambuco), que pode ser explicado pela peculiaridade dessa localidade. Por outro lado, nas regiões Sul e Centro-Oeste, não há municípios no primeiro quintil, que agrupa os municípios com NSE mais baixo; e no Sul, apenas quatro municípios figuram no segundo quintil. Na região Sudeste, o estado de Minas Gerais representa uma síntese nacional, pois é o único que possui municípios em todos os quintis; e São Paulo é o estado com a maior concentração de municípios no quintil mais alto - 338, que equivalem a 6% dos municípios do país.

Figura 3 – Cartograma da distribuição territorial da média municipal do NSE das escolas



Fonte: Elaboração própria com dados do Inep.

Importante ressaltar que a descrição da média municipal do NSE não significa ausência de escolas com NSE alto no Norte/Nordeste ou, do contrário, escolas com NSE baixo nas outras regiões. Quando a unidade de análise é a escola, há unidades com valores do NSE nos percentis mais baixos e nos mais altos em todas as regiões, conforme o primeiro gráfico da Figura 2. Uma análise de cada região/unidade da Federação revelaria desigualdades que ficam ocultas na descrição nacional.

Esses resultados são similares a outros indicadores de pobreza no Brasil, como o Índice de Desenvolvimento Humano (IDH) e o Produto Interno Bruto (PIB) municipal per capita, por exemplo. O IDH é um índice composto pela expectativa de vida, a média de escolaridade e renda per capita dos habitantes dos municípios de acordo com dados do Censo Demográfico 2010. A correlação entre o NSE municipal e o IDH dos municípios brasileiros é 0,93, um valor bem alto, mas esperado tendo em vista que a componente de renda do IDH foi empregada como um dos indicadores secundários no cálculo do NSE das

escolas. O PIB per capita, que se refere à soma de todas as riquezas de um município, é considerado um indicador do padrão de vida médio de seus habitantes. A correlação entre o NSE do município e o PIB per capita (ano 2016) é 0,52, um valor médio-alto, que reflete o fato de que é possível ter um PIB per capita alto com uma distribuição desigual da riqueza. Essas similaridades devem ser vistas aqui como uma validação genérica da metodologia apresentada neste artigo.

Associação com resultados de aprendizado

As desigualdades socioeconômicas são uma característica dos sistemas educacionais que recebem cada vez mais a atenção dos pesquisadores. O NSE é uma dimensão essencial, assim como raça/cor e gênero, ao longo da qual a desigualdade educacional deve ser medida (Alves; Soares; Xavier, 2016; Soares; Alves; Xavier, 2016; Soares; Delgado, 2016).

A Tabela 2 mostra o desempenho médio das escolas públicas estaduais e municipais em leitura, calculado pela média das proficiências de seus respectivos estudantes do 5º ano do ensino fundamental (EF), agrupadas por quintis do NSE das escolas em que estavam matriculados no ano de realização da avaliação. Entre 2007 e 2019, as médias das escolas melhoraram em todos os quintis, mas as escolas com NSE mais alto avançaram bem mais. Isso está descrito na última coluna da tabela, que mostra a diferença na média do quintil mais alto em relação ao mais baixo. Em 2007, essa diferença era de 38 pontos e no último ano é de 49 pontos. O máximo foi observado em 2013, quando a diferença atingiu 56 pontos. Essas diferenças correspondem, em média, a aproximadamente três anos de escolaridade. Resultado muito semelhante é observado para matemática.

Tabela 2 - Proficiência média da escola em leitura, no 5º ano do ensino fundamental, por quintis do NSE da escola (escolas estaduais e municipais)

Ano Saeb	Quintil do NSE da escola					
	1º	2º	3º	4º	5º	Dif. 5-1
2007	150,53	154,72	163,61	174,75	188,41	37,9
2009	151,97	160,58	172,69	186,05	198,58	46,6
2011	159,20	167,21	178,83	192,26	206,58	47,4
2013	158,21	168,26	182,09	198,44	214,52	56,3
2015	171,97	181,69	194,63	209,36	223,83	51,9
2017	180,18	190,61	203,39	217,29	230,98	50,8
2019	180,69	191,37	203,18	215,91	230,11	49,4

Fonte: Elaboração própria com dados do Inep.

No 9º ano do EF, as médias das escolas também melhoram em leitura e matemática em todos os quintis nesse período, mas as diferenças praticamente se mantêm. A Tabela 3 ilustra um desses resultados para leitura:

Tabela 3 - Proficiência média da escola em leitura, no 9º ano do EF, por quintis do NSE da escola (escolas estaduais e municipais)

Ano Saeb	Quartil do NSE da escola					
	1º	2º	3º	4º	5º	Dif. 5-1
2007	206,33	211,61	220,58	228,67	240,54	34,2
2009	210,83	221,36	231,47	239,74	250,98	40,1
2011	210,83	221,36	231,33	240,92	252,75	41,9
2013	215,31	224,66	233,98	241,30	252,03	36,7
2015	224,86	233,24	241,51	248,53	260,26	35,4
2017	231,42	241,46	249,73	258,12	270,38	39,0
2019	234,16	243,27	249,74	257,42	270,36	36,2

Fonte: Elaboração própria com dados do Inep.

As Tabelas 2 e 3, no entanto, apresentam comparações entre médias de desempenho das escolas. Há argumentos para se considerar o percentual de estudantes nos diferentes níveis de aprendizado (Soares, 2009). Os estudantes no nível abaixo do básico necessitam de uma intervenção urgente, pois o seu nível de aprendizado pode acarretar abandono e conseqüente evasão.

A Tabela 4 mostra o percentual de estudantes abaixo do nível básico por escola, segundo os quintis do NSE, considerando apenas as escolas estaduais e municipais. Tal como observado na Tabela 2, houve melhora dos resultados entre 2007 e 2019. O percentual de estudantes por escolas no nível abaixo do básico se reduziu em todos os quintis. Entretanto, essa redução foi muito maior nas escolas com NSE mais altos. Em 2007, havia quase três vezes mais estudantes no nível abaixo do básico no primeiro quintil em comparação ao último, das escolas com NSE mais alto, conforme a informação da última coluna da tabela. Em 2019, essa relação passou para seis vezes. Em matemática, os resultados são muito parecidos.

Adotando-se o valor de cinco pontos percentuais como o limite para o percentual de estudantes no nível abaixo do básico por escola (Soares, 2009), pode-se observar que a velocidade da queda entre as escolas com baixos NSE é muito lenta. Esse resultado, mesmo analisado no nível das escolas, é semelhante ao de Soares e Delgado (2016), que apresentaram um indicador de desigualdade educacional aplicado aos dados da Prova Brasil 2013. Os autores utilizaram uma medida "KL" para descrever as desigualdades de aprendizado em relação a uma distribuição de referência e entre estudantes agrupados pelo NSE e outras características. A distância observada entre os grupos de escolas com os menores e os maiores valores de NSE é tão grande que, se o mesmo ritmo for mantido,

os alunos de NSE baixo precisam de até oito vezes mais tempo para chegar à distribuição de referência de qualidade de aprendizado que seus colegas de NSE alto.

Tabela 4 - Percentual de estudantes, por escola, no nível abaixo do básico em leitura, no 5º ano do ensino fundamental, por quintis do NSE da escola (estaduais e municipais)

Ano Saeb	Quartil do NSE da escola					Razão1/5
	1º	2º	3º	4º	5º	
2007	52,11	46,73	37,25	27,72	18,27	2,9
2009	52,76	42,59	30,84	21,02	14,19	3,7
2011	45,61	36,83	26,64	18,11	11,28	4,0
2013	49,63	38,77	26,92	16,70	9,88	5,0
2015	35,24	26,12	16,66	9,84	5,41	6,5
2017	30,31	21,70	13,42	7,56	4,29	7,1
2019	31,41	22,41	14,45	8,91	4,96	6,3

Fonte: Elaboração própria com dados do Inep.

No 9º ano, tal como observado na Tabela 3, ainda que o percentual de estudantes abaixo do básico tenha se reduzido em todos os quintis entre 2007 e 2019, as diferenças entre escolas com NSE mais alto e mais baixo praticamente se mantêm no período, conforme a Tabela 5, que apresenta o resultado para leitura. O resultado para matemática segue o mesmo padrão.

Tabela 5 - Percentual de estudantes, por escola, no nível abaixo do básico em leitura, no 9º ano do ensino fundamental, por quintis do NSE da escola (estaduais e municipais)

Ano Saeb	Quartil do NSE da escola					Razão1/5
	1º	2º	3º	4º	5º	
2007	45,50	40,59	32,37	26,31	19,05	2,4
2009	41,23	31,85	24,50	20,07	14,48	2,8
2011	41,48	31,79	24,48	19,33	13,36	3,1
2013	39,63	31,27	25,14	21,68	16,31	2,4
2015	30,40	23,92	19,33	16,83	11,99	2,5
2017	27,32	20,06	15,50	12,49	8,41	3,2
2019	25,84	19,59	16,20	13,38	9,10	2,8

Fonte: Elaboração própria com dados do Inep.

Em trabalhos recentes, Soares, Alves e Fonseca (2021) e Fonseca et al. (2022) mostraram que, quando se considera como resultado o percentual de trajetórias regulares

e toma-se como unidade de análise o município, se repete o observado nas tabelas anteriores. Ou seja, há uma clara associação entre o NSE e o resultado educacional.

Relação com o Ideb

O Ideb é o indicador oficial para monitorar a qualidade da educação básica brasileira (Brasil, 2007). Há várias maneiras de mostrar a associação desse indicador com o NSE das escolas. A correlação entre o NSE das escolas e o respectivo Ideb em diferentes anos é uma delas.

A Tabela 6 mostra essa análise. Os coeficientes de correlação de Pearson para as escolas com Ideb calculados nos anos de 2005 a 2019 são sempre positivos. Nos anos iniciais do ensino fundamental, que têm o desempenho melhor, a correlação é sempre mais forte. Nos anos finais, o valor do coeficiente é menor, mas ainda muito substancial, assim como no ensino médio, para o qual se passou a calcular o Ideb por escola em 2017.

Tabela 6 - Correlação de Pearson entre o NSE da escola e o Índice de Desenvolvimento da Educação Básica (Ideb), por etapa de ensino e edição

Etapa (*)	Ideb 2005	Ideb 2007	Ideb 2009	Ideb 2011	Ideb 2013	Ideb 2015	Ideb 2017	Ideb 2019
NSE AI	0,726	0,721	0,717	0,694	0,707	0,668	0,643	0,612
NSE AF	0,57	0,575	0,546	0,556	0,502	0,495	0,516	0,491
NSE EM	-	-	-	-	-	-	0,612	0,483

(*) AI: Anos iniciais do ensino fundamental; AF: Anos finais do ensino fundamental; EM: Ensino médio.

Fonte: Inep (Ideb)¹¹; e elaboração própria com dados do Inep.

Além do Ideb da escola, o Inep divulga o Ideb dos municípios e unidades da Federação, que, em conjunto, contribuem para a meta nacional do indicador. Repetimos essa análise de correlação entre o Ideb municipal e o NSE municipal, calculando apenas pela média do NSE das escolas públicas. As correlações seguem o mesmo padrão e são ainda mais fortes nas etapas do ensino fundamental.

Conclusão

A principal contribuição deste artigo foi a produção de uma medida do NSE para a grande maioria das escolas de educação básica brasileira, possibilitando estudos até aqui inviáveis. O NSE das escolas – construído a partir de informações dos questionários respondidos por estudantes e complementadas por informações do contexto social das escolas – se revelou uma medida válida, que pode ser incorporada nas pesquisas

¹¹ A planilha com os resultados do Ideb e a nota técnica estão disponíveis em: <<https://www.gov.br/inep/ptbr/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb/resultados>>. Acesso em: 20 ago. 2022.

acadêmicas e também contribuir para o aperfeiçoamento das políticas públicas educacionais.

A metodologia empregada deu visibilidade para as escolas pequenas, que são numerosas e até aqui estavam invisíveis, uma vez que elas não são elegíveis para as avaliações educacionais. O número de escolas com NSE estimado é quase três vezes maior que o apresentado na última versão do Inse do Inep – o indicador oficial para monitorar as políticas públicas nacionais. A alta correlação entre o indicador do Inep e o NSE apresentado neste artigo, tanto no nível da escola quanto do município, mostra que as duas informações se equivalem. Contudo, a maior cobertura de escolas possibilita análises mais consistentes das desigualdades educacionais.

O NSE dos municípios permite a análise de associação entre essa medida e resultados educacionais municipais. Dessa forma, é possível monitorar o direito à educação, no estudo das desigualdades educacionais, mas também o dever do Estado de verificar se os recursos – infraestrutura, humanos e pedagógicos – estão distribuídos adequadamente. Isso pode ser útil, por exemplo, para aperfeiçoar a execução do novo Fundeb, que tem a redução das desigualdades de nível socioeconômico como uma de suas condicionalidades. Uma medida mais robusta do NSE dos municípios, calculada com um número muito maior de escolas, permite produzir diagnósticos e prognósticos mais qualificados para orientar as decisões de políticas públicas que têm um alto impacto para os entes federados.

Nas pesquisas acadêmicas, o NSE da escola pode ser usado como variável de controle em estudos de associação entre fatores escolares e sociais com resultados educacionais. Nesse sentido, vale observar que o valor do NSE da escola é muito mais útil do que o respectivo valor no nível do estudante para descrever as desigualdades educacionais no Brasil (Soares; Alves; Oliveira, 2001; Xavier; Alves, 2015; Alves; Soares; Xavier, 2016; Soares; Alves; Xavier, 2016). As controvérsias metodológicas sobre o uso de variáveis de composição nos modelos estatísticos, como viés de agregação e tipo de medida (van Ewijk; Slegers, 2010; Marks, 2015), são solucionadas com o emprego dos modelos hierárquicos bem especificados (Harker; Tymms, 2004; Timmermans; Thomas, 2015). Essa classe de modelos permite verificar os efeitos de um fenômeno similar (por exemplo, o NSE) nos níveis do estudante e da escola, o que os torna interessantes para analisar a iniquidade entre escolas (Lee, 2008).

O NSE de cada escola foi calculado com um painel de dados que cobriu um longo período. Entretanto, essa é uma medida relativamente estável, mesmo quando se comparam as estimativas com dados longitudinais (Paula, 2020). De todo modo, é possível repetir o processo utilizando apenas dados de NSE primários e secundários de uma janela de tempo menor e criar estimativas com médias móveis.

A medida do NSE da escola não capta estruturas internas de segregação, como colocar os estudantes mais pobres em turnos ou turmas específicas. Esses são estudos

importantes que complementam o que foi obtido neste artigo e poderão ser objeto de investigações posteriores.

Um tipo de estudo necessário, na continuidade desse programa de investigação, é verificar as escolas e municípios com NSE *outliers*. Uma investigação de base mais qualitativa pode sugerir melhorias ao apontar quais são as escolas com esses valores extremos em cada município e nas unidades da Federação.

Uma característica importante deste artigo foi o esforço de transparência. Os dados usados foram baixados do site oficial do Inep. São, portanto, dados públicos e abertos, bem como todas as análises utilizadas. Ou seja, a metodologia pode ser verificada, ainda que isso exija expertise computacional e familiaridade com a área. No entanto, entendemos que isso é importante no momento atual.

Embora este artigo tenha utilizado dados que ainda não tinham sido incorporados à produção de estimativas do NSE das escolas, como o percentual de estudantes que recebem Bolsa Família e os dados longitudinais do Censo Escolar, há espaço para aperfeiçoamentos. O cálculo do NSE poderia ter mais precisão se a condição do estudante no CadÚnico fosse divulgada nos microdados do Censo Escolar. Essa informação já é coletada no formulário do estudante e, para esse tipo de pesquisa, ela poderia ser agregada no nível das escolas, isto é, o percentual de estudantes que recebem o benefício. O uso de informações para mais anos sobre o Programa Bolsa Família, além das que tivemos acesso, poderia contribuir para melhorar as estimativas pela redução de dados ausentes das escolas em funcionamento em outros anos. Outro aperfeiçoamento seria o acesso, para fins exclusivos de pesquisas, da informação sobre as escolas mascaradas, que permitiria melhorar as estimativas do NSE primário. Para aperfeiçoar o NSE secundário – que se revelou uma estratégia muito poderosa para os objetivos deste artigo – seria interessante incorporar dados de mais de uma coorte de estudantes cujas transferências ocorrem dentro do período coberto pelos dados do Saeb/Enem.

Além disso, uma recomendação seria a melhoria dos formulários de coleta de dados do Censo Escolar. A inclusão de perguntas sobre a escolaridade dos pais e sobre a ocupação do pai, mãe e/ou responsáveis possibilitariam estimar o NSE dos estudantes e de suas escolas com maior precisão. Essa informação estaria disponível para um número muito maior de estudantes, não apenas para os matriculados nas etapas/escolas elegíveis para as avaliações educacionais. Isso pode ser viável em municípios e estados que já usam sistemas computacionais para realizar a matrícula dos seus estudantes.

Em suma, este artigo apresenta os resultados de uma pesquisa acadêmica desenvolvida com dados públicos, com transparência e que têm claros usos na proposição e monitoramento de políticas públicas educacionais. Mas a implantação dessas políticas deve se basear em dados administrativos completos, cuja produção é de responsabilidade inalienável do Estado.

Referências bibliográficas

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, vol. 15, nº 1, p. 1-30, 2009.

ALVES, M. T. G., et al. "Fatores familiares e desempenho escolar: uma abordagem multidimensional". *Dados*, Rio de Janeiro, vol. 56, nº 3, p. 571-603, 2013.

ALVES, M. T. G.; SOARES, J. F.; XAVIER, F. P. "Índice socioeconômico das escolas de educação básica brasileiras". Rio de Janeiro, *Ensaio: Avaliação e Políticas Públicas em Educação*, vol. 22, nº 84, p. 671-703, set. 2014.

_____.; _____.; _____. "Desigualdades educacionais no ensino fundamental de 2005 a 2013: hiato entre grupos sociais". *Revista Brasileira de Sociologia*, Porto Alegre, vol. 4, nº 7, p. 49-81, 2016.

ALVES, T.; GOUVÊA, M. A.; VIANA, A. B. N. "Proposta de um Indicador Socioeconômico para os alunos das escolas públicas dos municípios brasileiros". *Dados*, Rio de Janeiro, vol. 57, nº 1, p. 129-166, 2014.

ARMOR, D. J.; MARKS, G. N.; MALATINSKY, A. "The Impact of School SES on Student Achievement: Evidence from U.S. Statewide Achievement Data". *Educational Evaluation and Policy Analysis*, Washington, DC, vol. 40, nº 4, p. 613-630, 2018.

AVVISATI, F. "The measure of socio-economic status in PISA: a review and some suggested improvements". *Large-Scale Assessments in Education*, vol. 8, nº 1, p. 1-37, 2020.

BARROS, G. T. F., et al. *Indicador de Nível Socioeconômico dos Inscritos do ENEM: concepção, metodologia e resultados*. Brasília, DF: Inep, 2019. (Texto para Discussão nº 47).

BARTHOLO, T. L.; COSTA, M. "Evidence of a school composition effect in Rio de Janeiro public schools". *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, vol. 24, nº 92, p. 498-521, 2016.

BARTHOLO, T. L., et al. "School segregation and education inequalities at the start of schooling in Brazil". *REICE. Revista Iberoamericana Sobre Calidad, Eficacia y Cambio en Educación*, vol. 18, nº 4, p. 77-96, 2020.

BOUDON, R.; BOURRICAUD, F. Estratificação social. In: BOUDON, R.; BOURRICAUD, F. (Eds.). *Dicionário crítico de Sociologia*. São Paulo: Editora Ática, p. 214-221, 1982.

BOURDIEU, P.; PASSERON, J. C. *A reprodução: elementos para uma teoria do sistema de ensino* (7a Ed.). Petrópolis: Editora Vozes, 1994.

BRASIL. Decreto Presidencial nº. 6.094, de 24 de abril de 2007. Dispõe sobre a implementação do Plano de Metas Compromisso Todos pela Educação, pela União, em regime de colaboração com Municípios, Distrito Federal e Estados. Diário Oficial da União, Poder Executivo, Brasília, DF, 243 de abr. 2007. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/decreto/d6094.htm>. Acesso em: 23 set. 2023.

BRASIL. Lei Federal 13.005, de 25 de junho de 2014. Aprova o Plano Nacional de Educação - PNE e dá outras providências. Brasília: Congresso Nacional, 2014. Disponível em: <<https://www2.camara.leg.br/legin/fed/lei/2014/lei-13005-25-junho2014-778970-publicacaooriginal-144468-pl.html>>. Acesso em: 22 set. 2023.

BRASIL. Lei nº 14.113, de 25 de dezembro de 2020, que dispõe sobre o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação. Brasília:

Congresso Nacional, 2020. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/l14113.htm>. Acesso em: 22 set. 2023.

BROER, M.; BAI, Y.; FONSECA, F. *Socioeconomic inequality and educational outcomes: Evidence from twenty years of TIMSS*. IEA Research for Education/Springer Open, 2019. Disponível em: <<http://www.springer.com/series/14293>>. Acesso em: 03 ago. 2022.

BROOKE, N.; SOARES, J. F. Seção 1-A escola não faz diferença. Comentários. In: BROOKE, N.; SOARES, J. F. (Orgs.). *Pesquisa em eficácia escolar: origens e trajetórias*. Belo Horizonte: Editora da UFMG, p.14-22, 2008.

BUCHMANN, C. Measuring Family Background in International Studies of Education: Conceptual Issues and Methodological Challenges. In: PORTER, A. C.; GARAMON, A. (Eds.). *Methodological advances in cross-national surveys of educational achievement*. Washington DC: National Research Council, p. 150-197, 2002.

CAMARGO, P. C.; PAZELLO, E. T. "Uma análise do efeito do programa bolsa família sobre o desempenho médio das escolas brasileiras". *Economia Aplicada*, vol. 18, nº 4, p. 623-640, 2014.

CHALMERS, P. "Package 'mirt' - A multidimensional item response theory package for the R environment (Version 1.37.1)". *Journal of Statistical Software*, vol. 48, nº 6, 2012.

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

DOMINA, T., et al. "Is Free and Reduced-Price Lunch a Valid Measure of Educational Disadvantage?". *Educational Researcher*, vol. 47, nº 9, p. 539-555, 2018.

FERRÃO, M. E.; PRATA, P.; ALVES, M. T. G. "Multiple imputation in big identifiable data for educational research: An example from the Brazilian education assessment system". *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, vol. 28, nº AHEAD, p. 599-621, 2020.

FIRPO, S., et al. "Evidence of eligibility manipulation for conditional cash transfer programs". *Economia*, Niterói, vol. 15, nº 2, p. 243-260, 2014.

FONSECA, I. C., et al. "Monitoramento da Permanência na Educação Básica com o Indicador de Regularidade de Trajetórias Educacionais" (preprint), 2022. Disponível em: <<https://doi.org/10.31219/osf.io/k2mb9>>. Acesso em: 22 set. 2023.

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

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, p. 147-200, 1993.

HARKER, R.; TYMMS, P. "The effects of student composition on school outcomes". *School Effectiveness and School Improvement*, vol. 15, nº 2, p. 177-199, 2004.

HARWELL, M. "Don't Expect Too Much: The Limited Usefulness of Common SES Measures". *Journal of Experimental Education*, vol. 87, nº 3, p. 353-366, 2019.

HARWELL, M.; LEBEAU, B. "Student eligibility for a free lunch as an SES measure in education research". *Educational Researcher*, vol. 39, nº 2, p. 120-131, 2010.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Nota Técnica: Indicador de Nível Socioeconômico das Escolas de Educação Básica* (Inse). Brasília, INEP, s/data. Disponível em:

<https://download.inep.gov.br/educacao_basica/prova_brasil_saeb/resultados/2013/nota_tecnica_ondicador_de_nivel_socioeconomico_das_escolas_de_educacao_basica_inse.pdf>. Acesso em: 22 set. 2023.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Nota Técnica: Indicador de Nível Socioeconômico do SAEB 2019*. Brasília, INEP, 2021. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/indicador_nivel_socioeconomico_saeb_2019_nota_tecnica.pdf>. Acesso em: 22 set. 2023.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. "Nota Técnica Conjunta No. 15/2022. Brasília, INEP, novembro de 2022. Disponível em: <https://www.gov.br/mec/pt-br/aceso-ainformacao/pdf/pdf/nota_tecnica_conjunta_0911456.pdf>. Acesso em: 22 set. 2023.

LEE, V. E. Utilização e modelos hierárquicos lineares para estudar contextos sociais: o caso dos efeitos da escola. In: BROOKE, N.; SOARES, J. F. (Orgs.). *Pesquisa em eficácia escolar: origens e trajetórias*. Belo Horizonte: Editora da UFMG, p. 273-296, 2008.

MALAGUTH, T. Z. "Migrações e fluxo escolar da coorte de estudantes de 2008 a 2019, em Minas Gerais". Dissertação de Mestrado em Demografia. Universidade Federal de Minas Gerais, Belo Horizonte, 2022. Disponível em: <<https://repositorio.ufmg.br/handle/1843/47288>>. Acesso em: 22 set. 2023.

MARKS, G. N. "Are school-SES effects statistical artifacts? Evidence from longitudinal population data". *Oxford Review of Education*, vol. 41, nº 1, p. 122-144, 2015.

MAY, H. A. "Multilevel Bayesian Item Response Theory Method for Scaling Socioeconomic Status in International Studies of Education". *Journal of Educational and Behavioral Statistics*, vol. 31, nº 1, p. 63-79, 2006.

PALARDY, G. J. "Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis". *School Effectiveness and School Improvement*, vol. 19, nº 1, p. 21-49, 2008.

PANIAGUA, D., et al. "Comparison of methods for dealing with missing values in the EPV-R". *Psicothema*, vol. 29, nº 3, p. 384-389, 2017.

PAULA, T. S. "Eficácia escolar sobre o aprendizado em matemática: um estudo longitudinal sobre como a desigualdade prejudica a estimação do efeito das escolas para estudantes da rede municipal de Belo Horizonte". Tese de Doutorado em Educação. Universidade Federal de Minas Gerais, Belo Horizonte, 2020. Disponível em: <<https://repositorio.ufmg.br/handle/1843/34711>>. Acesso em: 12 dez. 2022.

RIGOTTI, J. I. R.; HADAD, R. M. "An analysis of the relationship between internal migration and education in Brazil". *Background paper prepared for the 2019 Global Education Monitoring Report*. Unesco, 2018. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000266085/PDF/266085eng.pdf.multi.page=1&zoom=auto,-16,842>>. Acesso em: 22 set. 2023.

ROSE, N.; VON DAVIER, M.; XU, X. "Modeling Nonignorable Missing Data with Item Response Theory (IRT)". ETS - Educational Testing Service, 2010. Disponível em: <<http://www.ets.org/research/contact.html>>. Acesso em: 22 set. 2023.

RUTKOWSKI, D.; RUTKOWSKI, L. "Measuring socioeconomic background in PISA: One size might not fit all". *Research in Comparative and International Education*, vol. 8, nº 3, p. 259-278, 2013.

SAMEJIMA, F. "Estimation of Latent Ability Using a Response Pattern of Graded Responses". *Psychometric Monograph*, Iowa: Psychometric Society, nº 17, p. i-169, 1969.

SANTARROSA, R. B. "Impacto das transferências condicionadas de renda sobre a proficiência dos alunos do ensino fundamental no Brasil". Dissertação de Mestrado em Economia. Fundação Getúlio Vargas, São Paulo, 2011.

SCIFFER, M. G.; PERRY, L. B.; MCCONNEY, A. "Critiques of socio-economic school compositional effects: Are they valid?" *British Journal of Sociology of Education*, v. 41, n. 4, p. 462-475, 2020.

_____.; _____.; _____. "The substantiveness of socioeconomic school compositional effects in Australia: measurement error and the relationship with academic composition". *Large-scale Assessments in Education*, vol. 10, nº 21, p.1-22, 2022.

SIRIN, S. R. "Socioeconomic Status and Student Achievement: a meta-analytic review of research". *Review of Educational Research*, vol. 75, nº 3, p. 417-453, 2005.

SOARES, J. F. "Índice de desenvolvimento da Educação de São Paulo – IDESP: bases metodológicas". *São Paulo em Perspectiva*, São Paulo, vol. 23, nº 1, p. 29-41, 2009.

SOARES, J. F.; ALVES, M. T. G.; FONSECA, J. A. "Trajetórias educacionais como evidência da qualidade da educação básica brasileira". *Revista Brasileira de Estudos de População*, vol. 38, e0167, p.1-21, 2021.

SOARES, J. F.; ALVES, M. T. G.; OLIVEIRA, R. M. "O efeito de 248 escolas de nível médio no vestibular da UFMG nos anos de 1998, 1999 e 2000". *Estudos em Avaliação Educacional*, São Paulo, nº 24, p. 69-118, 2001.

SOARES, J. F.; ALVES, M. T. G.; XAVIER, F. P. "Effects of Brazilian schools on student learning". *Assessment in Education: Principles, Policy and Practice*, vol. 23, nº 1, p. 7597, 2016.

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, vol. 14, nº 50, p. 107-126, 2006.

SOARES, J. F.; COLLARES, A. C. M. "Family resources and cognitive performance by primary school students". *Dados*, Rio de Janeiro, vol. 49, nº 3, p. 615-650. 2007.

SOARES, J. F.; DELGADO, V. M. S. "Medida das desigualdades de aprendizado entre estudantes do ensino fundamental". *Estudos em Avaliação Educacional*, São Paulo, vol. 27, nº 66, p. 754-780, 2016.

SOARES, T. M. "Utilização da Teoria da Resposta ao Item na produção de indicadores socioeconômicos". *Pesquisa Operacional*, Rio de Janeiro, vol. 25, nº 1, p. 83-112, 2005.

STUBBE, T. C.; BUDDEBERG, M. "Poverty and School Achievement: An Additional Indicator for Socio-Economic Status in School Achievement Studies". In: *Anais 4th IEA International Research Conference*, Göteborg: 2010. Disponível em: <<https://www.iea.nl/publications/presentations/poverty-and-school-achievement>>. Acesso em: 02 ago. 2022.

THRUPP, M.; LAUDER, H.; ROBINSON, T. "School composition and peer effects". *International Journal of Educational Research*, vol. 37, nº 5, p. 483-504, 2002.

TIMMERMANS, A. C.; THOMAS, S. M. "The impact of student composition on schools' value-added performance: a comparison of seven empirical studies". *School Effectiveness and School Improvement*, vol. 26, nº 3, p. 487-498, 2015.

VAN EWIJK, R.; SLEEGERS, P. "The effect of peer socioeconomic status on student achievement: A meta-analysis". *Educational Research Review*, vol. 5, p. 134-150, 2010.

WILLMS, J. D. "School composition and contextual effects on student outcomes". *Teachers College Record*, vol. 112, nº 4, p. 1008-1037, 2010.

WWW - WORLD WITHOUT POVERTY. *Breve histórico do Cadastro Único*. Banco Mundial; IPC-IG; Ipea; MDS, 2014. Disponível em <<https://wpp.org.br/publicacao/breve-historicodo-cadastro-unico/>>. Acesso em: 03 set. 2022.

XAVIER, F. P.; ALVES, M. T. G. "A composição social importa para os efeitos das escolas no ensino fundamental?". *Estudos em Avaliação Educacional*, São Paulo, vol. 26, nº 61, p. 216-243, 2015.

Abstract

A measure of the socioeconomic status of Brazilian schools using primary and secondary indicators

The article presents a measure of socioeconomic status (SES) for nearly all Brazilian schools. Primary indicators – calculated with data from questionnaires answered by students in educational assessments – and secondary indicators – which characterize the schools' social contexts – were aggregated into an SES scale through an Item Response Theory model. The SES was calculated for more than 180,000 schools. The methodology gave visibility to small schools or ones in remote areas that, until now were invisible to public policies since they had not participated in educational assessments. The relationship between educational outcomes and SES represents the inequalities in the country. The schools' SES are useful for academic research, and can help improve public policies aimed at reducing educational inequalities.

Keywords: Schools' socioeconomic status; primary and secondary education; educational inequalities; social scales and measures; Item Response Theory

Resumen

Una medida del nivel socioeconómico de las escuelas brasileñas a partir de indicadores primarios y secundarios

El artículo presenta una medida del nivel socioeconómico (NSE) para casi todas las escuelas brasileñas. Indicadores primarios – calculados con los datos de los cuestionarios respondidos por los alumnos en las evaluaciones educativas – e indicadores secundarios – que caracterizan el contexto social de los centros escolares – se agregaron en una escala de NSE mediante un modelo de la Teoría de Respuesta al Ítem. El NSE se calculó para más de 180.000 centros escolares. La metodología dio visibilidad a escuelas pequeñas o de zonas remotas que, hasta entonces, eran invisibles para las políticas públicas, ya que no participaban en las evaluaciones educativas. La relación entre los resultados educativos y el NSE representa las desigualdades del país. El NSE de las escuelas es útil para la investigación académica y tiene el potencial de contribuir a la mejora de las políticas públicas destinadas a reducir las desigualdades educativas.

Palabras clave: nivel socioeconómico de las escuelas; educación primaria y secundaria; desigualdades educativas; escalas y medidas sociales; Teoría de Respuesta al Ítem

Résumé

Une mesure du niveau socio-économique des écoles brésiliennes à l'aide d'indicateurs primaires et secondaires

L'article présente une mesure du niveau socio-économique (NSE) pour presque toutes les écoles brésiliennes. Des indicateurs primaires – calculés à partir de données provenant de questionnaires auxquels les élèves ont répondu dans le cadre d'évaluations éducatives – et des indicateurs secondaires – qui caractérisent le contexte social des écoles – ont été agrégés sur une échelle de NSE au moyen d'un modèle de théorie de la réponse aux items. Le NSE a été calculé pour plus de 180.000 écoles. La méthodologie a donné une visibilité aux petites écoles ou à celles situées dans des zones reculées qui, jusqu'alors, étaient invisibles pour les politiques publiques, puisqu'elles ne participent pas aux évaluations éducatives. La relation entre les résultats scolaires et les NSE représente les

inégalités du pays. Les NSE des écoles est utile pour la recherche académique et peut contribuer à l'amélioration des politiques visant à réduire les inégalités éducatives.

Mots-clés : niveau socio-économique des écoles ; enseignement primaire et secondaire ; inégalités éducatives ; échelles et mesures sociales ; Item Response Theory

Artigo submetido à publicação em 15 de fevereiro de 2023.

Artigo aprovado para publicação em 23 de agosto de 2023.

Opinião Pública adota a licença Creative Commons CC-BY.

