

NÍVEL SOCIOECONÔMICO DAS ESCOLAS BRASILEIRAS: METODOLOGIA PARA AMPLIAR SUA COBERTURA E POTENCIAIS USOS

Maria Teresa Gonzaga Alves^I

Clarissa Guimarães Rodrigues^{II}

Adriano Souza Senkevics^{III}

<https://doi.org/10.24109/9786558011026.ceppe.v9.5971>

RESUMO

Este estudo apresenta a metodologia e os resultados de um indicador de nível socioeconômico (NSE) para o total de 164.840 escolas brasileiras com matrículas de escolarização, em funcionamento no ano de 2021. Foram utilizados indicadores primários, calculados com dados de questionários respondidos pelos estudantes nas avaliações e exames nacionais, e indicadores secundários - que caracterizam o contexto social das escolas, como os beneficiários do Programa Bolsa Família (PBF) -, além do NSE de estudantes transferidos entre estabelecimentos de ensino. Esses indicadores foram agregados em uma escala de NSE por meio de um modelo de respostas graduais da Teoria de Resposta ao Item. 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

^I Diretora de Estudos Educacionais do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep); professora da Universidade Federal de Minas Gerais (UFMG); doutora em Educação pela UFMG

^{II} Pesquisadora-Tecnologista do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep); doutora em Demografia pela Universidade Federal de Minas Gerais (UFMG).

^{III} Pesquisador-Tecnologista do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep); mestre e doutor em Educação pela Universidade de São Paulo (USP).

participam das avaliações e exames educacionais. A relação entre resultados educacionais e o NSE representa as desigualdades do País. É possível concluir que 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; educação básica; desigualdades sociais; Teoria de Resposta ao Item.

INTRODUÇÃO

O nível socioeconômico (NSE) é um indicador de grande importância para a compreensão da desigualdade na educação, tendo em vista a conhecida associação entre resultados educacionais e a origem social dos estudantes (Coleman *et al.*, 1966). No nível das escolas, essa associação é, em geral, mais forte (Sirin, 2005), de modo que a composição social dos estudantes é um fator muito importante na investigação sobre a eficácia escolar e na orientação das políticas públicas educacionais (Willms, 2010; Xavier; Alves, 2015; Sciffer; Perry; McConney, 2022).

A metodologia de cálculo do NSE de estudantes e escolas é conhecida e já foi aplicada inúmeras vezes no Brasil (Soares, 2005; Soares; Andrade, 2006; Alves; Soares, 2009; Alves; Gouvêa; Viana, 2014; Alves; Soares; Xavier, 2014; Barros *et al.*, 2019). Em geral, esse cálculo utiliza dados produzidos por meio de questionários contextuais respondidos pelos estudantes que participam do Sistema de Avaliação da Educação Básica (Saeb), do Exame Nacional do Ensino Médio (Enem) e de outras avaliações ou exames educacionais. Esses questionários incluem itens que podem ser utilizados para estimar o NSE dos estudantes, e a média dessa medida é usualmente tomada como o NSE das escolas.

Há uma vasta literatura que demonstra a validade do indicador de NSE, obtido por meio dessa metodologia, para contextualizar e melhorar a compreensão dos resultados dos estudantes nas avaliações educacionais, como o Saeb, em nível nacional, o Programa Internacional de Avaliação de Estudantes (Pisa) e o Third International Mathematics and Science Study (TIMSS), em nível internacional (Soares, 2005; Rutkowski; Rutkowski, 2013; Broer; Bai; Fonseca, 2019; Barros *et al.*, 2019; Avvisati, 2020; Sciffer; Perry; McConney, 2022). Entretanto, essa metodologia restringe a possibilidade de cálculo do NSE às escolas que participam dessas avaliações ou exames, o que, geralmente, não alcança a metade das escolas registradas no Censo Escolar da Educação Básica. Por conseguinte, 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

de pequeno porte, localizados em cidades pouco populosas, 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.

Ampliar a abrangência da medida de NSE para quase a totalidade dos estabelecimentos de ensino do Brasil é de fundamental importância, especialmente após a regulamentação do Novo Fundeb, Lei nº 14.113, de 25 de dezembro de 2020 (Brasil, 2020). A lei inclui em seus dispositivos legais o NSE dos educandos como variável-chave para a distribuição de recursos do Valor Anual por Aluno Fundeb (Vaaf) e Valor Anual Total por Aluno (Vaat), bem como para o monitoramento da redução das desigualdades de resultados educacionais vinculadas à distribuição do Valor Aluno Ano por Resultados (Vaar). Ainda que mencione “educandos”, com base na literatura sobre o efeito de composição social na escola – também conhecido como efeito de pares (Van Ewijk; Slegers, 2010; Firpo; Jales; Pinto, 2015) –, argumentamos que é adequado operacionalizar essa norma legal por meio de uma medida de NSE da escola. Ao lado disso, a disponibilidade de um indicador abrangente e preciso de NSE é essencial para o monitoramento das metas que serão pactuadas para o novo Plano Nacional de Educação (PNE) para o decênio 2024-2034. Essa abrangência ampliada e metodologia robusta possibilitam diagnósticos e prognósticos mais precisos, fundamentais para direcionar decisões de políticas públicas com um impacto significativo na educação nacional. À guisa de exemplo, é possível que o novo PNE demande recortes socioeconômicos para aferição dos cumprimentos de suas metas com vistas a reduzir as desigualdades.

O objetivo deste trabalho é apresentar uma metodologia para se calcular uma medida de NSE de quase a totalidade das escolas ativas de educação básica no Brasil no ano de 2021. Para isso, utilizamos, além da média escolar do NSE dos estudantes participantes de avaliações e exame (como tradicionalmente já se fazia para contextualizar os resultados das avaliações educacionais), estimativas de NSE a partir de estudantes transferidos entre distintos estabelecimentos de ensino e dados públicos que caracterizam o contexto social das escolas. Especificamente, trazemos à análise dados sobre o percentual de alunos beneficiários do Programa Bolsa Família (PBF) em cada escola, além de informações sobre a localização do estabelecimento de ensino.

A principal novidade dessa metodologia em relação aos trabalhos prévios é a utilização das informações obtidas pela análise das trajetórias escolares dos estudantes. Na educação básica brasileira, a maioria dos estudantes é transferida de escola pelo menos uma vez durante a sua trajetória escolar (Rigotti; Hadad, 2018; Malaguth, 2022). Isso ocorre tanto porque poucas escolas ofertam todas as etapas de ensino – sobretudo no setor público, que concentra a maioria das matrículas – quanto porque os familiares se mudam de cidade, de bairro ou decidem transferir os filhos de escola. Cada estudante, durante a sua trajetória escolar, terá grande chance de se matricular em uma ou mais escolas que tiveram o seu NSE calculado devido à participação de seus estudantes em avaliações e exames educacionais.

Com base na literatura que demonstra a associação entre origem social, destinos e escolhas escolares das famílias, bem como os padrões da segmentação entre escolas públicas e privadas no País, assumimos que os estudantes transferidos “levam” a sua condição social e econômica de uma escola para outra (Soares; Collares, 2007; Alves, F., 2010; Alves, M. T. G., 2010; Costa; Koslinski, 2011; Resende; Nogueira; Nogueira, 2011; Alves *et al.*, 2013). As transferências dos estudantes entre escolas permitem, portanto, construir uma medida do NSE das escolas que não participam das avaliações e exames para expandir o número de escolas com NSE estimado.

Na próxima seção, explicamos os detalhes dessa metodologia, que possibilitou o cálculo do NSE de quase a totalidade das escolas de educação básica registradas no Censo Escolar de 2021. Informações complementares estão no Apêndice. Posteriormente, a validade da medida obtida é analisada na seção de resultados. Por último, nas considerações finais, discutimos a importância da medida para as políticas públicas educacionais.

1 METODOLOGIA

O cálculo do NSE escolar foi adaptado da metodologia proposta por Soares e Alves (2023), que utilizou três categorias de indicadores. A primeira, que chamaremos de *NSE primário* da escola, é a média do NSE dos estudantes calculada a partir das respostas aos itens dos questionários contextuais do Saeb e/ou Enem, entre 2011 e 2021. Utilizamos as medidas de NSE primários, já computadas pelo Inep, referentes aos anos de 2011, 2013, 2015, 2019 e 2021. A essas, acrescentamos as medidas calculadas pelos autores de modo a se ter indicadores calculados com os dados do Enem para todo o período, com base em Barros *et al.* (2019), e com os dados do Saeb para a edição de 2017, com metodologia similar à utilizada oficialmente pelo Inep. Considerando as edições do Saeb e do Enem, foram utilizados 12 NSE primários; cada escola pode ter estimativas para vários desses indicadores, dependendo da participação de seus estudantes no Saeb e Enem.

Em seguida, partindo-se de três diferentes coortes de nascimento – 2000/2001, 2002/2003, 2008/2009 –, os estudantes foram acompanhados ao longo de sua trajetória escolar de modo que se mapeassem as suas transferências entre escolas. Assim, geraram-se medidas de *NSE secundário* para escolas que não participaram do Enem e/ou Saeb no período considerado. Dessa maneira, foram calculadas até três medidas de NSE secundários por escola, uma para cada coorte.

Adicionamos à base de dados as seguintes variáveis contextuais: uma variável indicadora de localização urbana/rural, o componente renda do Índice de Desenvolvimento Humano Municipal (IDHM-Renda), referente ao município de localização da escola, e o percentual de alunos cujas famílias participam do

Programa Bolsa Família (PBF) – aqui entendido como uma *proxy* de estudantes próximos à linha de pobreza. Esse último, por sua vez, é um importante indicador de NSE escolar, frequentemente utilizado como substituto de medidas como a que propomos neste texto. Porém, por ser discriminativo do segmento mais empobrecido da população brasileira, pode não ser suficiente para distinguir estudantes pertencentes a classificações de renda acima da linha de pobreza extrema, média e alta; daí porque o utilizamos em conjunto com outras medidas.

Finalmente, o banco de dados utilizado para cálculo do NSE final é constituído dos indicadores gerados pela discretização (isto é, a transformação de uma escala intervalar para uma categórica ordinal) das variáveis descritas acima. Todas as variáveis foram discretizadas usando classes definidas pelo desvio-padrão, com exceção da variável localização, originalmente dicotômica (urbana/rural), e do percentual de alunos beneficiários do PBF, dicotomizado para uma variável indicador de maioria/minoria de beneficiários entre o corpo discente da escola. Com relação às variáveis discretizadas, a primeira classe agrega os valores abaixo de -1,5 desvios-padrão. Os outros intervalos têm tamanho igual de meio desvio-padrão. A última classe engloba as escolas com desvio-padrão acima de 1,5 ou 2, dependendo da amplitude da variável. O número de classes varia entre 7 e 9.

Com base nesses dados, executamos a modelagem da Teoria de Resposta ao Item (TRI) por meio do modelo de respostas graduais, discutido adiante. Esse método permite gerar uma medida agregada de todos os itens utilizados na composição do NSE e, assim, permitir uma estimativa bastante sólida do que as medidas têm em comum, isto é, da magnitude do NSE familiar de cada estudante. Em um primeiro momento, essa estimativa é produzida em uma escala arbitrária com média 0 e desvio-padrão 1; assim, realizamos uma transformação de escala para reposicionar os valores estimados em média 50 e desvio-padrão 10, de modo que o indicador se torne mais facilmente compreensível.

1.1 NSE PRIMÁRIO

O NSE primário das escolas corresponde à média do NSE de seus estudantes calculada a partir de suas respostas aos questionários contextuais nas avaliações educacionais. Os itens dos questionários empregados no cálculo medem, direta ou indiretamente, a escolaridade e a renda familiar. A utilização dessas informações para estimar o NSE dos indivíduos e das escolas – uma medida de composição social – encontra respaldo na literatura especializada (Buchmann, 2002; Sirin, 2005; Soares, 2005; Rutkowski; Rutkowski, 2013; Alves; Soares; Xavier, 2014; Broer; Bai; Fonseca, 2019; Barros *et al.*, 2019; Avvisati, 2020; Sciffer; Perry; McConney, 2022).

Foram utilizados os NSE primários estimados pelo Inep com dados do Saeb e Enem de 2011 a 2013; Saeb e Enem de 2015; e Saeb de 2019 a 2021 (Barros *et al.* 2019;

Brasil. Inep, 2014, 2021, 2023a). Além disso, calculamos o NSE primário das escolas com os dados do Saeb 2017, Enem 2017, 2018, 2019, 2020 e 2021 especialmente para uso neste trabalho. No total, utilizamos 12 NSE primários. O Quadro 1 resume as bases de dados para o cálculo do NSE primário e a referência do indicador.

QUADRO 1
COMPILADO DAS MEDIDAS DE NSE PRIMÁRIO UTILIZADAS NO CÁLCULO FINAL DO INDICADOR DE NSE

Ano	Fonte dos dados	
	Saeb	Enem
2011-2013	Brasil, 2014	
2012	-	Barros <i>et al.</i> , 2019
2014	-	Barros <i>et al.</i> , 2019
2015	Brasil, 2016	
2016	-	Barros <i>et al.</i> , 2019
2017	Elaboração própria	
2018	-	Elaboração própria
2019	Brasil, 2021	Elaboração própria
2020	-	Elaboração própria
2021	Brasil, 2023	Elaboração própria

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Censo Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021¹.

Para calcular o NSE dos estudantes, utilizamos um modelo da TRI. A TRI engloba um conjunto 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. Nas avaliações educacionais, esse é um desafio comum, visto que, por razões diversas, os estudantes deixam sem respostas itens dos questionários. A TRI consegue calibrar os itens com dados parcialmente ausentes junto com as respostas observadas nesses itens para estimar o traço latente.

O modelo de resposta gradual da TRI, adequado a respostas politômicas (Samejima, 1969), foi empregado para estimar o NSE dos estudantes. Essa classe de modelos tem como pressuposto a unidimensionalidade, isto é, a existência de um único construto latente dominante no conjunto de dados, o qual deve ser verificado antes do ajuste do modelo.

¹ O acesso aos microdados completos para reproduzir os resultados desta pesquisa pode ser solicitado, exclusivamente, por meio do Serviço de Acesso a Dados Protegidos (Sedap). Disponível em Brasil. Inep (2023b).

A heurística para a estimação dessa medida envolveu três fases. Inicialmente, testamos os pressupostos desse modelo. 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 com auxílio do pacote *mirt* do *software* R (Chalmers, 2012)².

Escore individuais dos estudantes foram agregados por código dos estabelecimentos de ensino, segundo médias aritméticas simples, para constituir um escore médio por escola. O NSE médio das escolas calculado nessa etapa descreve fidedignamente as desigualdades regionais, entre redes de ensino e escolas, tal como resultados anteriores (Alves; Soares; Xavier, 2014; Barros *et al.*, 2019).

Cabe ressaltar que a escala do NSE, estimada por meio de itens de questionários respondidos por estudantes na escola, tem uma correlação muito alta com medidas similares calculadas com informações fornecidas por responsáveis pelos alunos. Por exemplo, a Secretaria de Educação do Estado do Rio Grande do Sul coletou, durante a matrícula em 2009, informações sobre escolaridade, ocupação e renda dos pais dos estudantes, a fim de produzir uma medida do NSE das escolas da rede. O índice de NSE produzido com esses dados administrativos tem uma correlação de 0,95 (sendo o máximo equivalente a 1,0) com uma medida de NSE calculada com dados obtidos por meio de questionários respondidos pelos alunos que participaram do Saeb e Enem, entre 2007 e 2011, considerando 4.015 escolas em comum nos dois estudos (Alves; Soares; Xavier, 2014). A mesma análise foi feita com a escala de NSE da pesquisa longitudinal Geres, que coletou dados diretamente junto aos pais dos estudantes e observou uma correlação de 0,93, considerando 229 escolas em comum (Alves; Soares; Xavier, 2014). Ou seja, no nível macro, a descrição da realidade socioeconômica vista tanto por um indicador obtido por meio das respostas de alunos a um questionário contextual quanto por um indicador obtido por meio das respostas dos pais é essencialmente a mesma.

O uso dessa medida do NSE para contextualizar os resultados educacionais e para a formulação de políticas públicas que visam a reduzir o impacto das desigualdades socioeconômicas nesses resultados encontra, portanto, amplo respaldo teórico e empírico. Contudo, há alguns municípios brasileiros cujas escolas não participam das avaliações educacionais e há outros com poucas escolas com dados disponíveis. Por essa razão, propomos produzir uma medida mais robusta e abrangente de NSE.

Para isso, vamos tomar os valores estimados nessa etapa e a análise das trajetórias dos estudantes para inferir um valor de NSE secundário para as escolas sem informação, como explicaremos na sequência. A expansão do NSE das escolas garantirá que todos os municípios do País tenham o NSE calculado.

² Os procedimentos são, em linhas gerais, os mesmos descritos nestas publicações: Brasil. Inep (2014); Barros *et al.* (2019); Brasil. Inep (2021); Brasil. Inep (2023a).

1.2 NSE SECUNDÁRIO

Esta nova formulação do indicador de NSE proposta pelo Inep, inspirada no estudo de Soares e Alves (2023), traz um indicador adicional, denominado NSE secundário, que comporá a cesta de variáveis usadas para estimar o NSE final. O NSE secundário é obtido a partir do acompanhamento de uma coorte de nascimento ao longo de sua trajetória escolar na educação básica. O uso de uma coorte de nascimento se justifica porque nela está contida toda a população de determinada idade x , independentemente do indivíduo ter ou não registro de matrícula em determinada escola em um ou mais anos letivos específicos. Desde 2007, a criança, ao ser matriculada no sistema de ensino, tem o seu registro no Censo Escolar com um código de identificação único. Assim como outros pesquisadores (Osório, 2021; Cedeplar, 2021), verificamos que o tamanho populacional das coortes de indivíduos nascidos a partir de 2000 registrados no Censo Escolar aproxima-se do tamanho da população nas projeções demográficas por idade simples do Instituto Brasileiro de Geografia e Estatística (IBGE, 2018). Dessa forma, pode-se dizer que a cobertura populacional do Censo Escolar é similar à cobertura populacional das projeções do Censo Demográfico.

Vale dizer que a inclusão do indicador de NSE secundário, que contribui para lançar luz sobre o NSE de escolas antes invisibilizadas, é uma grande novidade metodológica trazida por Soares e Alves (2023), e seu cálculo só é possível pela existência de um rico e fidedigno conjunto de dados disponíveis no painel longitudinal do Censo Escolar (Brasil. Inep, 2017), algo relativamente novo e bastante específico do contexto brasileiro.

Ao acompanhar os estudantes de determinada coorte de nascimento ao longo dos diversos anos disponíveis do Censo Escolar, é possível identificar as escolas pelas quais esses estudantes transitaram ao longo do período. A transferência do estudante entre escolas distintas traz a oportunidade de se estimar o NSE para aquelas escolas que não possuem o NSE primário. Para tanto, parte-se do pressuposto de que o NSE médio dos estudantes de uma determinada escola (NSE primário) é uma boa *proxy* do NSE desses estudantes. Assim, o estudante, ao ser transferido de uma *escola B*, com estimativa do NSE primário, para uma *escola C*, sem estimativa do NSE primário, carrega consigo o NSE primário médio da *escola B* e, assim, fornece uma estimativa do NSE para a *escola C* de destino. A mesma lógica é dada para uma escola de partida ou saída. Por exemplo, o estudante que se mudou de uma *escola A*, exclusiva de educação infantil – e, portanto, sem o NSE primário –, para a *escola B*, fornece uma informação para o NSE da *escola A*. A informação “carregada” por um único estudante é uma informação pequena e pode ser imperfeita, mas, como são milhões de estudantes analisados, as informações trazidas por milhões compõem um sólido indicador.

Para a construção do indicador de NSE secundário, o primeiro passo é definir as coortes de nascimento que serão usadas como base para o acompanhamento dos estudantes e suas transferências entre escolas ao longo do percurso escolar.

Optamos por trabalhar com duas coortes de nascimento mais velhas, para dar a oportunidade de seus estudantes percorrermos toda a educação básica e “carregarem” consigo informações do NSE primário. Foram consideradas a coorte de nascidos entre 1º de julho de 2000 e 30 de junho de 2001 (N = 3.466.539) e a coorte de nascidos entre 1º de julho de 2002 e 30 de junho de 2003 (N = 3.236.151), cuja trajetória escolar hipotética está representada na Figura 1. São estudantes com idade prevista de entrada no 1º ano do ensino fundamental em 2007 e 2009, respectivamente, quando o Censo Escolar já estava coletando os dados individualizados por estudante, o que torna possível acompanhá-los longitudinalmente. Também são estudantes que teriam a chance de atravessar os diversos ciclos de aplicação do Saeb e, portanto, terem seus dados dos questionários contextuais usados na construção do NSE primário. Por fim, são estudantes que teriam um tempo para potencialmente alcançar e concluir o ensino médio, o que aumenta a chance de terem participado também do Enem.

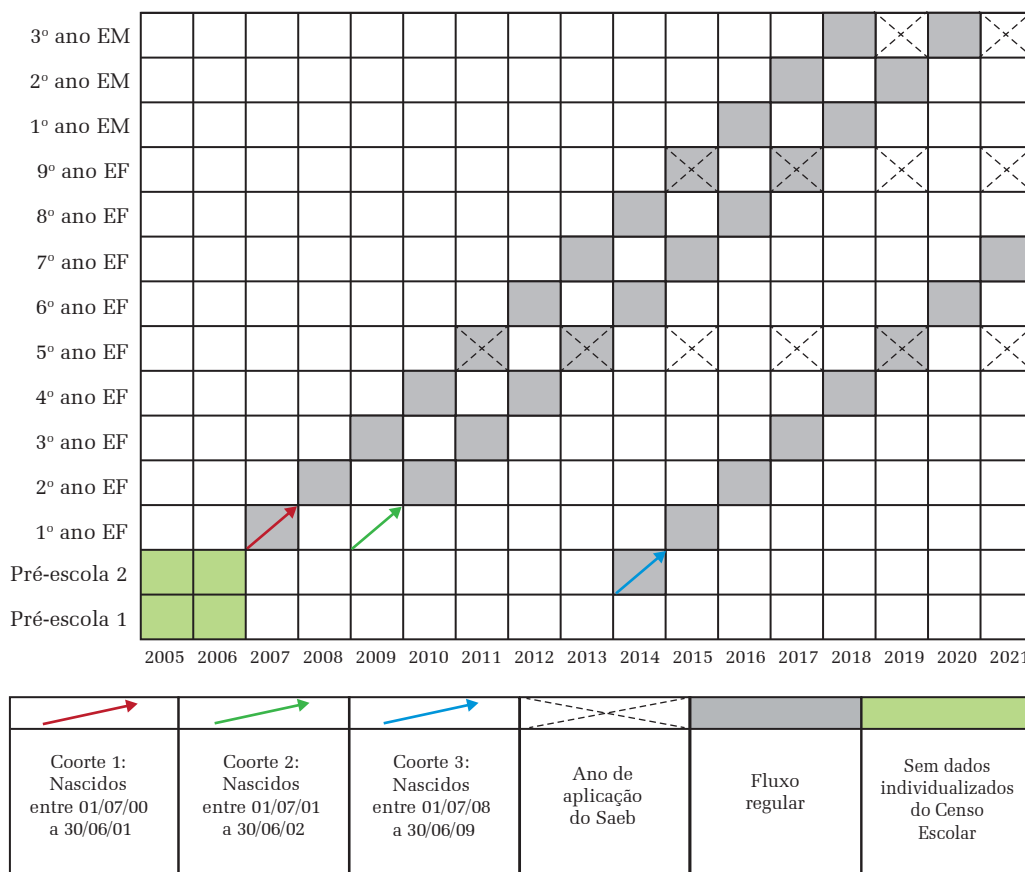


FIGURA 1
REPRESENTAÇÃO GRÁFICA DA TRAJETÓRIA ESCOLAR TEÓRICA DAS TRÊS COORTES DE NASCIMENTO SELECIONADAS PARA O ESTUDO POR MEIO DO DIAGRAMA DE LEXIS

Fonte: Elaboração própria

Além dessas duas coortes, utilizamos também uma coorte de nascimento mais jovem com representatividade na educação infantil, já que as escolas exclusivas de educação infantil formam um grupo relevante das escolas sem indicador de NSE primário, uma vez que não participam do Saeb. Foram considerados os estudantes nascidos entre 1º de julho de 2008 e 30 de junho de 2009 (N = 3.012.868). Conforme mostra a Figura 1, os estudantes com trajetória regular dessa coorte fizeram parte do público-alvo do Saeb 2019, 5º ano, e os estudantes irregulares podem ter participado do Saeb 2021, 5º ano.

Uma vez definidas as coortes de nascimento que serão acompanhadas, o segundo passo metodológico é construir a base de dados para se estimar o NSE secundário. Inicialmente, são geradas três bases de dados, uma para cada coorte. Cada base de dados tem em suas linhas as observações, definidas pela combinação do código do estudante, código da escola em que esteve matriculado e ano escolar (a Tabela 1 informa a quantidade de linhas de cada base de dados na coluna “Observação”). Para cada uma das três bases de dados, faz-se o pareamento com o arquivo contendo os NSE primários. Nas linhas estão as 178.370 escolas ativas da educação básica, segundo o Censo Escolar de 2021, e nas colunas estão as 12 variáveis que representam os NSE primários calculados com base no Saeb e Enem nos diversos ciclos já apresentados.

TABELA 1
TAMANHO DAS COORTES DE NASCIMENTO UTILIZADAS PARA O CÁLCULO DO NSE SECUNDÁRIO

Coortes de Nascimento	Definição	Tamanho da População	Observação: combinação do código do estudante, ano e código da escola em que esteve matriculado
Coorte 1	Nascidos entre 1º de julho de 2000 e 30 de junho de 2001	3.466.539	38.891.286
Coorte 2	Nascidos entre 1º de julho de 2002 e 30 de junho de 2003	3.236.151	40.509.169
Coorte 3	Nascidos entre 1º de julho de 2008 e 30 de junho de 2009	3.012.868	27.313.674

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Censo Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

Construída a base de dados que atribui os NSE primários das escolas a cada combinação de estudante e sua(s) escola(s) frequentada(s), parte-se para o terceiro passo metodológico: gerar o NSE secundário das escolas. Para cada estudante, agrega-se o NSE primário das escolas em que ele esteve matriculado. Essa medida é “carregada” para a(s) escolas(s) que ele frequentou e que não possuía(m) o NSE primário. Em seguida, para cada escola, agrega-se o NSE primário dos estudantes que tiveram registro de matrícula na escola. Como são até 12 medidas de NSE primário, o cálculo final do NSE secundário é dado pela média de até 12 medidas do NSE primário

médio de seus estudantes. Uma explicação mais detalhada sobre esse procedimento, com ilustrações, está presente no Apêndice.

O resultado dessa construção metodológica do NSE secundário é um banco de dados no qual as linhas são representadas pela totalidade de escolas com matrículas de escolarização em funcionamento em 2021, com quatro variáveis: código da escola, NSE secundário da Coorte 1, NSE secundário da Coorte 2 e NSE secundário da Coorte 3.

Os histogramas do NSE secundário para cada uma das três coortes de nascimento estão representados na Figura 2.

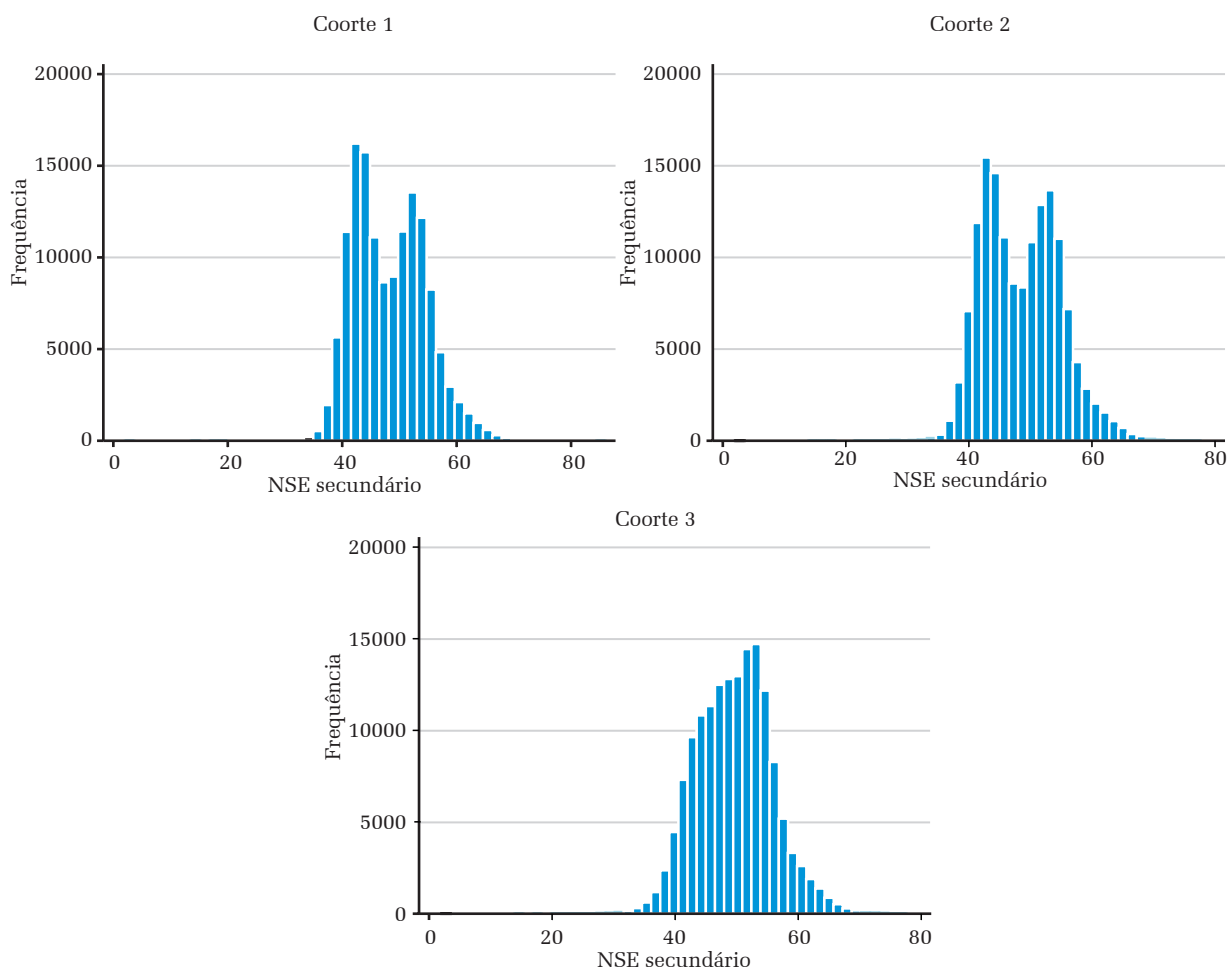


FIGURA 2
HISTOGRAMA DO NSE SECUNDÁRIO DAS ESCOLAS POR COORTE

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

A Tabela 2 apresenta os resultados das correlações lineares do NSE secundário com cada um dos NSE primários e o NSE produzido por Soares e Alves (2023). Observa-se correlações elevadas – a maioria por volta de 0,90 – para todos os indicadores apresentados.

TABELA 2
CORRELAÇÕES LINEARES DE PEARSON DO NSE SECUNDÁRIO DAS ESCOLAS COM OS NSE PRIMÁRIOS CALCULADOS A PARTIR DE DISTINTAS FONTES

Correlações lineares de Pearson	Coorte 1		Coorte 2		Coorte 3		
	N	Corr.	N	Corr.	N	Corr.	
NSE artigo Soares e Alves (2023)	126.532	0,94	127.494	0,94	118.699	0,90	
NSE's primários	NSE_2011_2013	69.034	0,93	69.043	0,93	62.196	0,90
	NSE_2015	61.325	0,92	61.380	0,92	54.446	0,89
	NSE_2017	57.055	0,90	57.238	0,90	53.457	0,85
	NSE_SAEB_2019	67.744	0,93	68.023	0,93	61.517	0,89
	NSE_SAEB_2021	68.865	0,91	69.229	0,91	63.052	0,88
	NSE_ENEM_2012	26.854	0,92	26.869	0,92	20.288	0,92
	NSE_ENEM_2014	28.090	0,93	28.108	0,93	21.156	0,92
	NSE_ENEM_2016	30.296	0,93	30.345	0,92	23.057	0,91
	NSE_ENEM_2018	29.348	0,93	29.376	0,93	21.789	0,92
	NSE_ENEM_2019	28.303	0,93	28.332	0,93	20.680	0,92
NSE_ENEM_2020	28.947	0,92	29.080	0,92	21.408	0,91	
NSE_ENEM_2021	28.740	0,91	29.020	0,91	21.370	0,90	

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

1.3 VARIÁVEIS DE CONTEXTO

As variáveis de contexto incluídas na modelagem do NSE final são: localização da escola (urbana/rural), percentual de estudantes beneficiários do PBF e o componente renda do Índice de Desenvolvimento Humano Municipal (IDHM-Renda). De modo geral, a localização das escolas em áreas rurais indica o atendimento escolar de uma população economicamente mais desfavorecida, dado que, frequentemente, essas regiões enfrentam maior dificuldade no acesso a serviços básicos e oportunidades educacionais. Esse indicador pode sinalizar a necessidade de políticas públicas direcionadas para mitigar desigualdades educacionais substanciais. Contudo, vale ressaltar que, embora possa ser um indicador útil, existem variações significativas dentro das áreas rurais e, portanto, tal marcador deve ser usado com cautela e em conjunto com outros indicadores para uma análise mais precisa, motivo pelo qual ele compõe uma das variáveis da modelagem final.

Utilizar o percentual de estudantes beneficiários do PBF por escola emerge como uma *proxy* eficaz para avaliar o NSE escolar. Esse indicador permite uma

rápida identificação de instituições que acolhem uma parcela mais vulnerável da população, revelando potenciais desafios associados, como acesso limitado a recursos educacionais e maior risco de evasão escolar. Escolas com uma alta percentagem de beneficiários geralmente estão situadas em regiões de maior vulnerabilidade social, onde a renda familiar é limitada e as condições de vida são mais precárias. Embora seja uma ferramenta valiosa, é crucial reconhecer suas limitações, pois não encapsula completamente a diversidade de contextos presentes em uma comunidade escolar, necessitando ser complementada com outros indicadores para uma análise mais robusta e completa.

Adicionalmente, incluímos na base o IDHM-Renda, que mensura o nível de renda da população de um município e pode ser utilizado como um indicador para conhecer o NSE do contexto no qual as escolas estão inseridas. Ao explorar esse índice, é possível identificar os municípios que apresentam desafios socioeconômicos mais acentuados, oferecendo uma imagem inicial das disparidades que podem ser encontradas no ambiente educacional dessas regiões. Mesmo considerando o fato de o IDHM-Renda ser uma medida mais agregada que os índices escolares calculados, notamos empiricamente que ele contribui para a estimação do NSE das escolas, em especial entre as escolas que dispõem de poucos indicadores de NSE primário.

Além das variáveis mencionadas nesta seção, testamos outras que, após primeiras modelagens empíricas, foram retiradas do modelo por motivos de parcimônia. Especificamente, testamos a variável dependência administrativa (pública/privada) – utilizada na proposta inicial de Soares e Alves (2023) –, mas essa mostrou-se pouco discriminativa do NSE das escolas brasileiras, uma vez consideradas as demais variáveis presentes no modelo.

2 RESULTADOS

A estimativa de um NSE composto de variáveis primárias, secundárias e de contexto social da escola gera uma medida com bastante informação entre -2 e 2 desvios-padrão, conforme se observa na Figura 3. Entre as variáveis que, individualmente, contribuem para a estimativa global do índice, constam as variáveis primárias e secundárias do NSE. Demais informações, tais como a localização da escola, o IDHM-Renda e a participação dos estudantes no PBF contribuíram com relativamente menos informação.

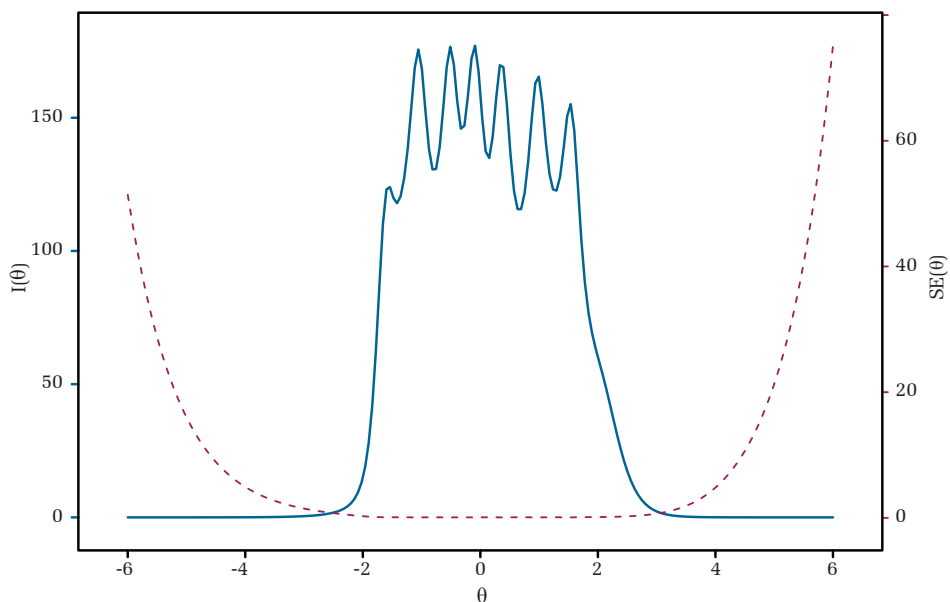


FIGURA 3

CURVA DE INFORMAÇÃO TOTAL E ERRO-PADRÃO DO NSE – BRASIL – 2011-2021

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Censo Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

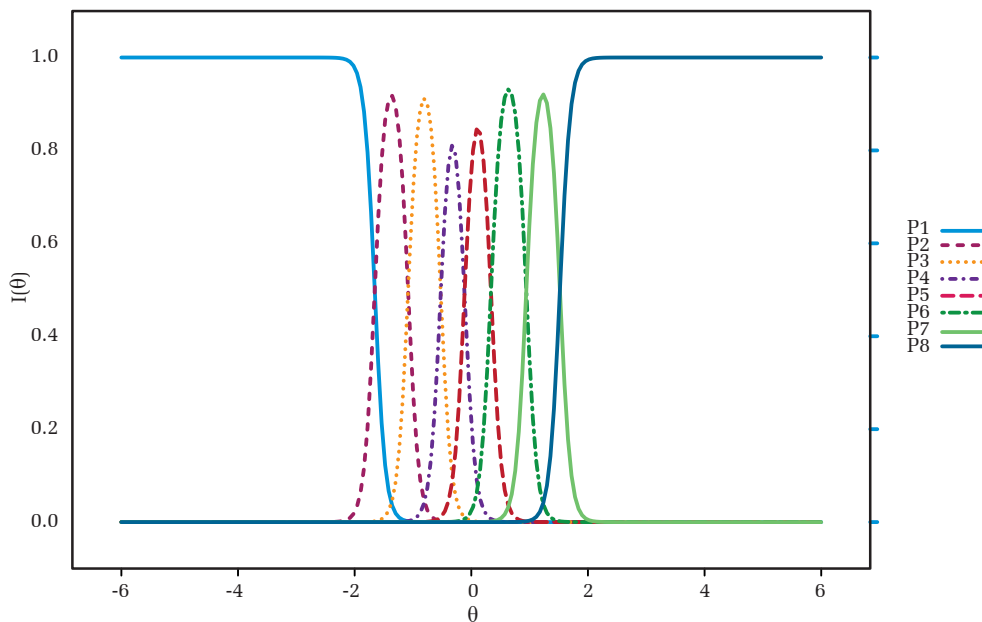


FIGURA 4

CURVA CARACTERÍSTICA DE UM ITEM DO NSE SECUNDÁRIO – BRASIL – 2011-2021

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

Um exemplo da curva característica do item de uma variável secundária do NSE, referente à coorte de nascimento 2000-2001, está ilustrada na Figura 4. Esse dado reforça a qualidade de uma medida secundária de NSE bastante discriminativa ao longo da escala.

Evidentemente, a acurácia das estimativas varia em função da quantidade de indicadores disponíveis para cada estabelecimento de ensino. Escolas que tiveram seu NSE calculado através da utilização de indicadores abrangendo as três categorias apresentam uma maior precisão. Por outro lado, essa precisão é reduzida para as escolas cuja estimativa do NSE se restringiu ao uso do NSE secundário, inferido a partir da média dos NSE primários das instituições para as quais os alunos foram transferidos, bem como das características contextuais. De modo geral, a presença de pelo menos um indicador de NSE primário contribui para uma diminuição da variância. A Tabela 3 detalha as estatísticas descritivas relacionadas ao desvio-padrão das estimativas, levando-se em conta a presença ou ausência de dados de NSE primário. É notável que a média dos desvios-padrão para o conjunto de escolas avaliadas com base em algum NSE primário é menor do que aquela observada para escolas cujas avaliações foram conduzidas sem tais dados.

TABELA 3
ESTATÍSTICAS DESCRITIVAS DO DESVIO-PADRÃO DO NSE DAS ESCOLAS POR GRUPO DE DISPONIBILIDADE DE DADOS

Grupos	Média	Nº de escolas	Desvio-padrão	Mínimo	Máximo
1 - Escola possui algum NSE primário	0,110	88.940	0,047	0,046	0,667
2 - Escola possui algum NSE secundário	0,180	75.900	0,074	0,119	0,521
Total	0,143	164.840	0,070	0,046	0,667

Fonte: Elaboração própria.

Por sua vez, a Tabela 4 apresenta a abrangência do novo indicador de NSE proposto neste estudo, assim como a sua comparação com outras metodologias vigentes. Na coluna “Total”, apresenta-se a distribuição de escolas da educação básica brasileiras, com matrículas de escolarização em funcionamento no ano de 2021, considerando categorias específicas que caracterizam a invisibilidade das escolas nas avaliações e exame nacionais: escolas de áreas rurais, exclusivas de educação infantil, escolas pequenas, escolas indígenas, ou localizadas em terras indígenas, e escolas localizadas em comunidade remanescente de quilombos.

Na coluna “Escolas com apenas o NSE primário”, observa-se que o indicador não alcança 3% do total de escolas exclusivas de educação infantil e do total de escolas muito pequenas (aquelas com 10 estudantes matriculados ou menos). Na

coluna “Escolas com apenas o NSE do Saeb 2021”, constata-se a inexistência do indicador para escolas exclusivas de educação infantil e escolas muito pequenas, além de um percentual muito baixo de alcance de escolas rurais (31,1%), indígenas (9,3%) e localizadas em comunidade remanescente de quilombo (33,1%). Por fim, a última coluna apresenta a medida do NSE proposta neste texto, que abrange 92% do total de escolas de educação básica ativas no Brasil em 2021. Cumpre notar que essa medida alcança quase 100% das escolas rurais e localizadas em comunidade remanescente de quilombos, cerca de 90% das escolas pequenas e escolas indígenas e cerca de 80% das escolas exclusivas de educação infantil. Esse resultado evidencia a ampla abrangência dessa nova medida de NSE, o que demonstra utilidade desse indicador para a focalização das políticas públicas de combate às desigualdades educacionais.

TABELA 4
DISTRIBUIÇÃO DAS ESCOLAS DE EDUCAÇÃO BÁSICA NO BRASIL EM ATIVIDADE NO ANO DE 2021, POR TIPO E LOCALIZAÇÃO DIFERENCIADA

Tipo e localização das escolas	Total de escolas com matrícula de escolarização ativas em 2021	Escolas apenas com o NSE primário		Escolas apenas com o NSE do SAEB 2021		Escolas com a nova metodologia do NSE	
	N	N	%	N	%	N	%
Rurais	53.557	21.255	39,7	16.677	31,1	52.430	97,9
Exclusivas de Educação Infantil	41.885	1.123	2,7	0	0,0	32.931	78,6
Pequenas							
20 estudantes ou menos matriculados	14.535	319	2,2	19	0,1	12.866	88,5
21 a 50 estudantes matriculados	23.225	1.939	8,3	492	2,1	20.566	88,6
Indígenas	3.464	674	19,5	322	9,3	3.061	88,4
Localizadas em comunidade remanescente de quilombos	2.548	1.046	41,1	843	33,1	2.507	98,4

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Censo Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

Do total de 178.370 escolas ativas da educação básica em 2021 com matrículas de escolarização, apenas 13.530 (7,5%) ficaram sem a medida do NSE. Dessas, 66% (8.954) são escolas exclusivas de educação infantil e 3.240 (24%) são escolas privadas que não participaram da amostra do Saeb nem receberam estudantes com NSE secundário.

Considerando que um dos objetivos do NSE proposto é utilizá-lo na focalização de políticas públicas – entre as quais o Fundeb –, agregamos o indicador para uma base municipal por meio da média do NSE escolar ponderado pela quantidade de matrículas de escolarização no Censo Escolar 2021. Uma vez que o indicador foi calculado para um quantitativo bastante expressivo de escolas (quase 165 mil), conseguimos obter uma medida de NSE de todos os 5.570 municípios brasileiros. A Figura 5 ilustra o histograma do NSE em nível municipal.

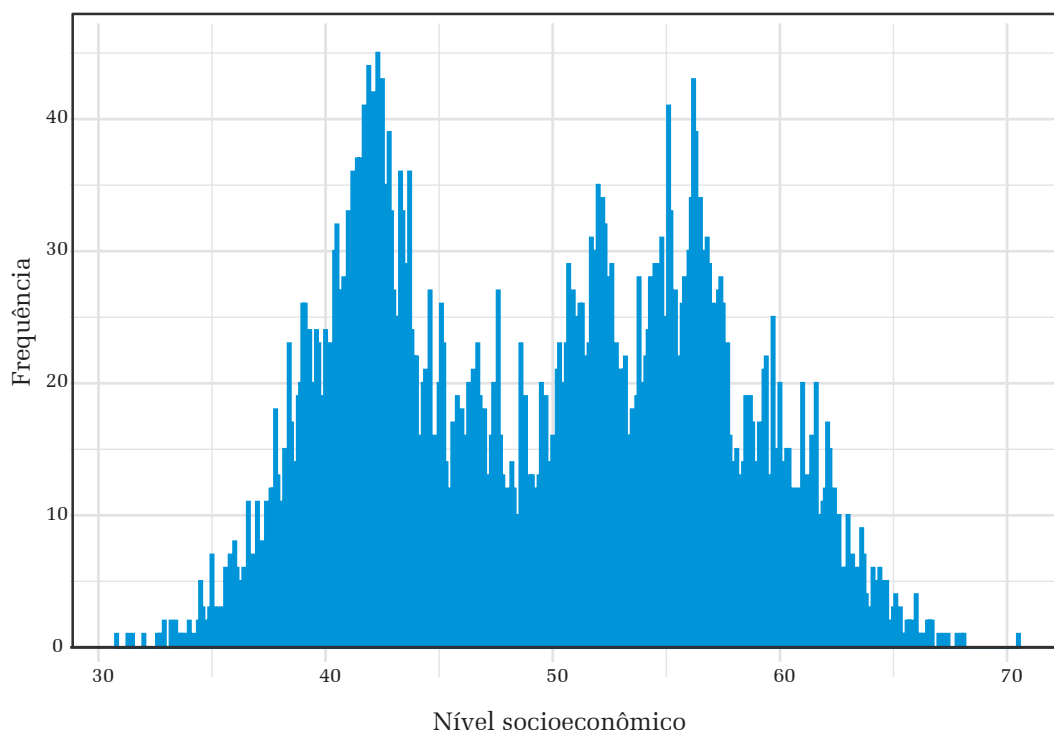


FIGURA 5

**HISTOGRAMA DO NSE MÉDIO MUNICIPAL PONDERADO PELO NÚMERO DE MATRÍCULAS
- BRASIL**

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Censo Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

A distribuição do resultado do NSE é muito desigual no território brasileiro, com um padrão de desigualdade muito semelhante a outros indicadores de contextos sociais e econômicos. Observa-se na Figura 6 um cartograma da distribuição do NSE médio por

município. Conforme a escala de cores da legenda, as cores esverdeadas representam uma concentração maior de alunos oriundos de famílias com NSE mais elevado; já as cores avermelhadas representam NSEs mais baixos. O cartograma ilustra que os municípios das regiões Sudeste, Sul e Centro-Oeste tendem a concentrar estudantes de maior NSE familiar, contrariamente ao que se observa nas regiões Norte e Nordeste.

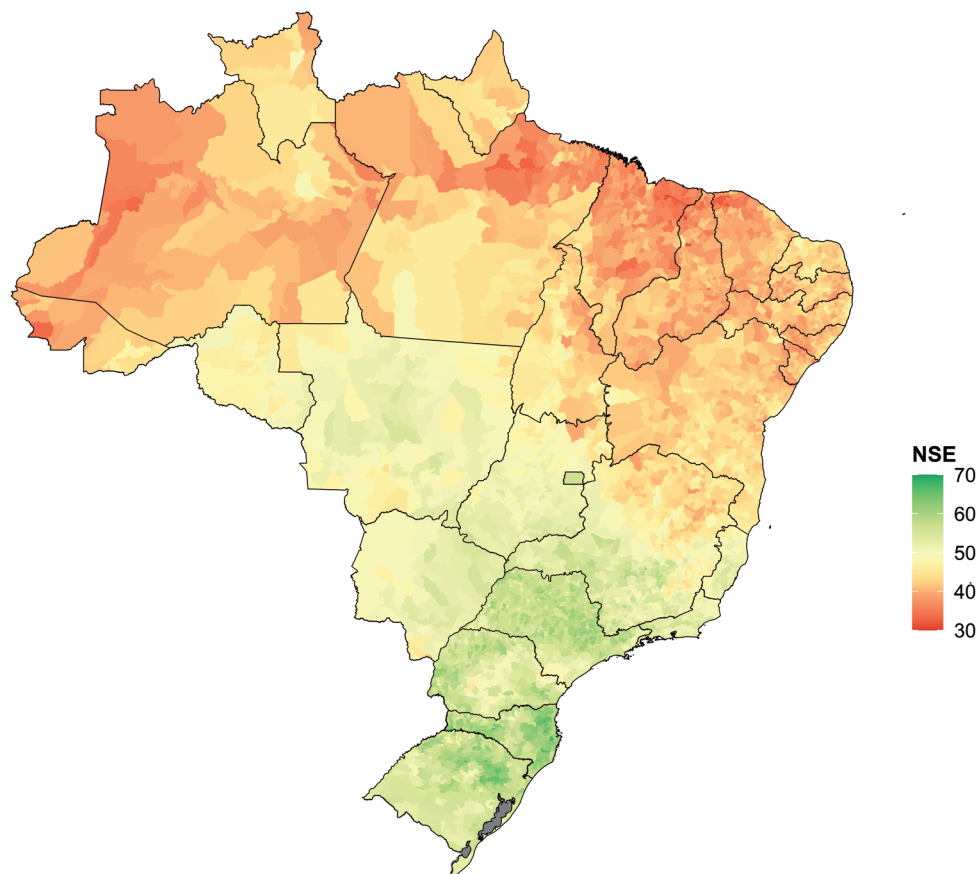


FIGURA 6
CARTOGRAMA DA DISTRIBUIÇÃO DO NSE MÉDIO MUNICIPAL – BRASIL

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Censo Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

A seguir, são apresentadas as validações internas e externas do indicador. Analisamos um conjunto de variáveis que identificam a relação do NSE com contextos mais ou menos vulneráveis, com o intuito de checar a robustez da medida sintética aqui proposta. Observa-se, pela Figura 7, o que já seria esperado no contexto brasileiro

em relação às desigualdades regionais e por tipo de escola em relação à composição socioeconômica de seu alunado. Em média, escolas municipais, rurais, localizadas no Norte e Nordeste, em terras indígenas e quilombolas são aquelas que apresentam a menor média de NSE de seus educandos.

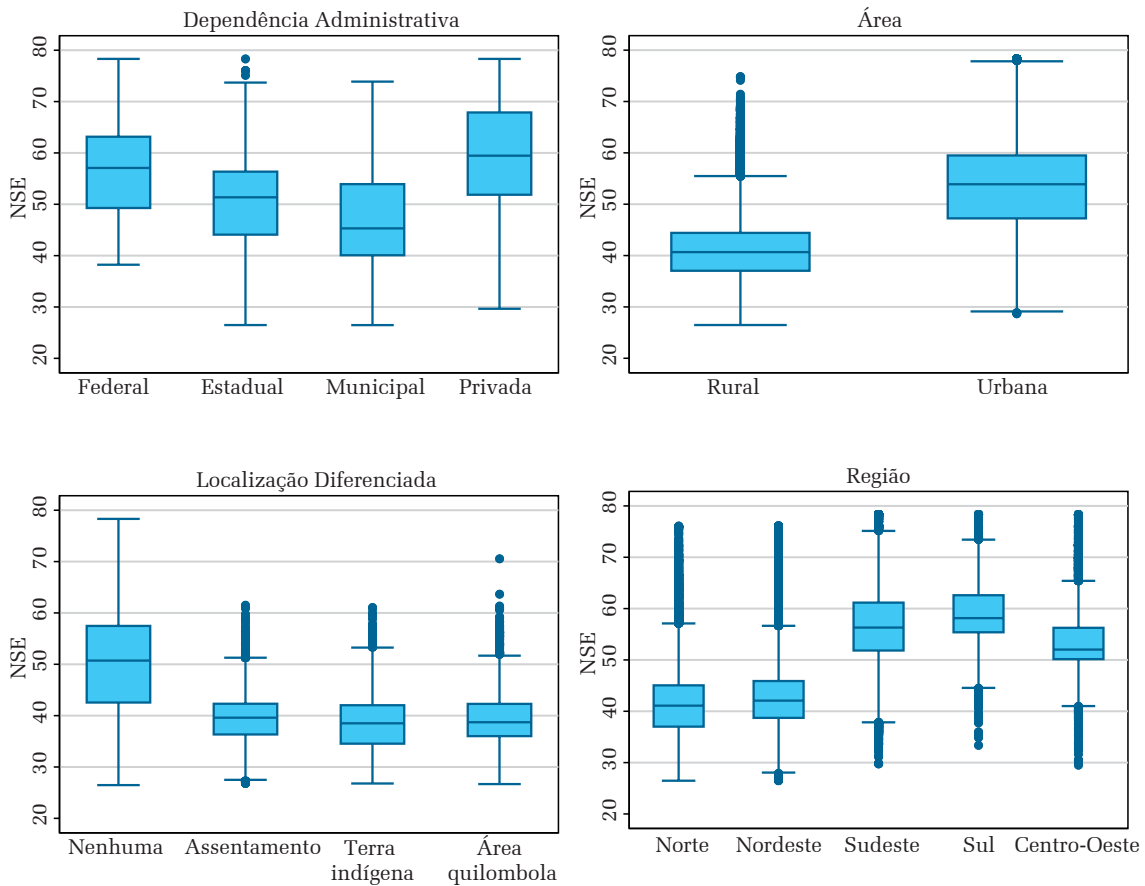


FIGURA 7

VARIAÇÃO DO INDICADOR DE NSE POR DEPENDÊNCIA ADMINISTRATIVA, ÁREA, REGIÃO E LOCALIZAÇÃO DIFERENCIADA – BRASIL

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Censo Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

Para analisar a consistência externa do indicador, a Tabela 5 apresenta as correlações lineares do indicador proposto neste estudo e um conjunto de outras medidas sintéticas que buscam captar o NSE municipal e escolar.

TABELA 5
CORRELAÇÕES LINEARES DE PEARSON ENTRE O NSE PROPOSTO E UM CONJUNTO DE INDICADORES SOCIOECONÔMICOS DE ÂMBITO MUNICIPAL E ESCOLAR

Indicadores	Descrição	Correlação
Indicadores por município		
T_ANALF25M	Taxa de analfabetismo da população com 25 anos ou mais por município	-0,88
T_FUND25M	Taxa de conclusão do ensino fundamental da população com 25 anos ou mais	0,68
T_MED25M	Taxa de conclusão do ensino médio da população com 25 anos ou mais	0,61
T_SUPER25M	Taxa de conclusão do ensino superior da população com 25 anos ou mais	0,64
PMPOB	Proporção de pobres (indivíduos com renda domiciliar <i>per capita</i> ≤ R\$140,00 mensais)	-0,92
PPOB	Proporção de vulneráveis à pobreza (indivíduos com renda domiciliar <i>per capita</i> ≤ R\$255,00 mensais)	-0,96
RDPC	Renda domiciliar <i>per capita</i>	0,90
RENOCUP	Rendimento médio dos ocupados com 18 anos ou mais	0,82
IDHM	IDHM - Índice de Desenvolvimento Humano Municipal	0,91
IDHM_EDUC	IDHM - Educação	0,80
IDHM_LONG	IDHM - Longevidade	0,83
IDHM_REND	IDHM - Renda - Renda municipal <i>per capita</i>	0,93
Indicadores por escola		
IOEB	Índice de Oportunidades da Educação Brasileira	0,68
ESCS - PISA 2018	NSE médio das escolas participantes do Pisa 2018. Indicador calculado pela OCDE	0,77
SES - PIRLS 2021	NSE médio das escolas participantes do Pirls 2021. Indicador calculado pela IEA	0,79

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Censo Escolar 2007-2021, Saeb 2011-2021, Enem 2011-2021, Pnud; Ipea; FJP, 2013 e Centro de Liderança Pública (CLP) 2017.

Nota: O IOEB utiliza as seguintes fontes: índice de desenvolvimento da educação básica, escolaridade do professor e número de horas-aula/dia, experiência do diretor, taxa de atendimento da educação infantil (Inep) e taxa líquida de matrícula do Ensino Médio (IBGE).

A Figura 8 ilustra a correlação entre o NSE médio municipal e a proporção municipal de vulneráveis à pobreza (isto é, pessoas que vivem em domicílios com renda *per capita* inferior a meio salário mínimo: R\$ 255,00 em 2010). Em resumo, essa ilustração representa graficamente a informação oferecida na tabela anterior

de que a correlação entre os dois indicadores é bastante elevada, da ordem de $-0,96$. Isso indica uma correlação quase perfeita entre as informações fornecidas pelos dois indicadores, reforçando a robustez do indicador ora proposta na mensuração da vulnerabilidade municipal.

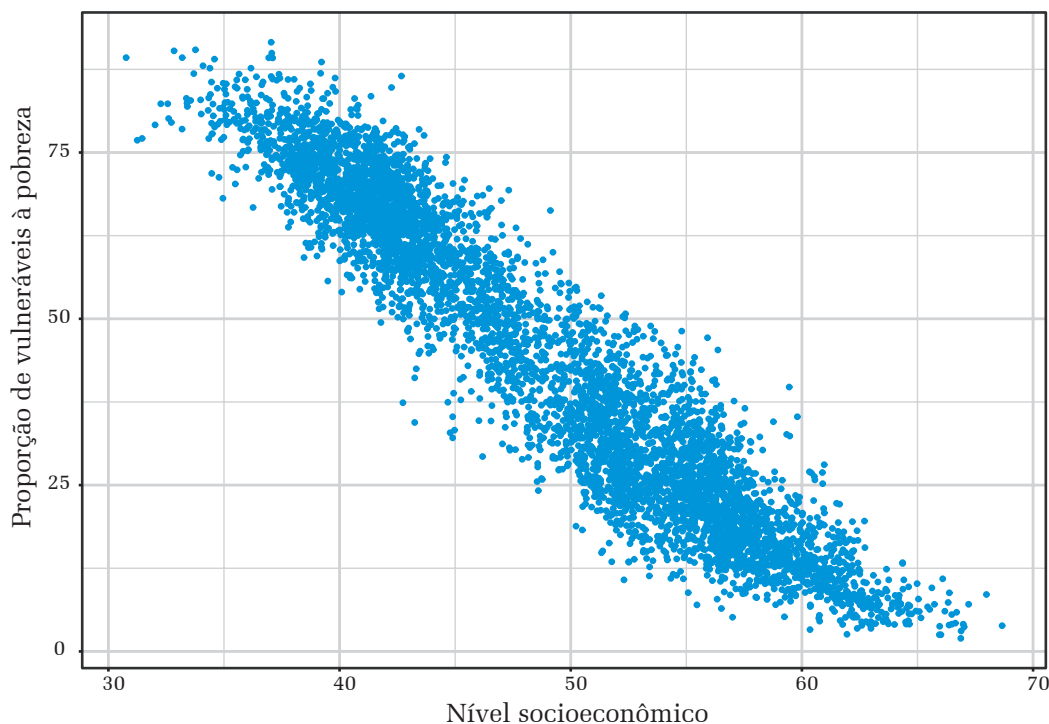


FIGURA 8

CORRELAÇÃO ENTRE A PROPORÇÃO DE VULNERÁVEIS À POBREZA NOS MUNICÍPIOS E O NSE MÉDIO MUNICIPAL – BRASIL

Fonte: Elaboração própria a partir de dados do Painel Longitudinal do Censo Escolar 2007-2021, Saeb 2011-2021 e Enem 2011-2021.

CONSIDERAÇÕES FINAIS

O Novo Fundeb traz em seus dispositivos legais a necessidade de se mensurar o NSE dos educandos, tendo em vista que essa medida é um insumo essencial para a operacionalização dessa importante política pública. Atribui-se ao Inep a definição da metodologia e o cálculo desse indicador – art. 18 da Lei nº 14.113/2020 (Brasil, 2020) – em periodicidade bienal. Como demonstrado neste trabalho, o NSE da escola calculado com indicadores primários e secundários resulta em uma medida robusta, válida, com grande capacidade de captar diferenças entre grupos de estudantes. Dessa forma, argumentamos que ele atende de forma adequada à lei e deve ser adotado como

parâmetro desse dispositivo legal, além de orientar o processo de monitoramento do cumprimento das metas do novo PNE.

Em síntese, o indicador de NSE tradicionalmente produzido pelo Inep é calculado com o objetivo de contextualizar, para melhor compreensão, os resultados dos estudantes nas avaliações educacionais. Dessa forma, essa medida está restrita às escolas que participam do Saeb ou Enem, o que contempla no máximo 40% do total de escolas de educação básica brasileiras. Considerando o desenho amostral do Saeb e o perfil dos estudantes que se inscrevem no Enem (estudantes concluintes do ensino médio), a grande maioria das escolas que possuem um corpo discente mais vulnerável – portanto, escolas que deveriam ser o foco das políticas públicas atuais de redução das desigualdades – não tem o NSE calculado.

Para ampliar a abrangência desse importante indicador, a principal inovação metodológica proposta neste estudo foi incluir, como variável adicional na cesta de variáveis usadas para estimar o NSE, a estimativa de NSE fornecida pelos estudantes transferidos entre escolas. Ao se transferirem entre escolas, os estudantes carregam consigo o seu NSE e fornecem informação para a estimativa do NSE da escola de chegada ou de saída. Cada estudante fornece uma informação pequena e imprecisa, mas, como foram usados dados de mais de 9 milhões de estudantes de três coortes de nascimento distintas, foi possível criar um indicador que adiciona informação útil, principalmente por não estar presente nos outros indicadores.

Conforme visto, usamos a TRI como metodologia para a construção do NSE final. Como é amplamente conhecido, esse modelo sintetiza em um único fator – o traço latente – a informação presente nos diferentes indicadores utilizados, cada um deles informativo e limitado à sua maneira. Essa metodologia usa dados reais, por meio de uma construção sólida e auditável, que produz a medida do NSE das escolas e sua agregação para os municípios sem a necessidade de imputação de dados.

Desse modo, foi possível expandir o número de escolas com NSE calculado – sobretudo entre as escolas rurais, pequenas, indígenas, localizadas em território quilombola, exclusivas de educação infantil. A expansão da cobertura de escolas com NSE calculado permitiu estimar um indicador de NSE municipal mais robusto e para todos os 5.570 municípios do Brasil. 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.

DISPONIBILIZAÇÃO DOS DADOS

Para aqueles interessados em analisar mais profundamente o perfil socioeconômico de cada escola, é possível acessar os microdados do NSE por escola diretamente na página deste texto. Esses dados, disponibilizados detalhadamente,

facilitam a realização de análises mais acuradas e podem ser um recurso valioso para pesquisadores, educadores e gestores educacionais na elaboração de estratégias e políticas focadas.

REFERÊNCIAS

ALVES, F. Escolhas familiares, estratificação educacional e desempenho escolar: quais as relações. *Dados*, Rio de Janeiro, v. 53, n. 2, p. 447-468, 2010.

ALVES, M. T. G. Dimensões do efeito das escolas: explorando as interações entre famílias e estabelecimentos de ensino. *Estudos em Avaliação Educacional*, São Paulo, v. 21, n. 46, p. 271-296, 2010.

ALVES, M. T. G. et al. Fatores familiares e desempenho escolar: uma abordagem multidimensional. *Dados*, Rio de Janeiro, v. 56, n. 3, p. 571-603, 2013.

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, v. 15, n. 1, p. 1-30, jun. 2009.

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

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, v. 57, n. 1, p. 129-166, 2014.

AVVISATI, F. The measure of socio-economic status in Pisa: a review and some suggested improvements. *Large-Scale Assessments in Education*, [S. l.], v. 8, n. 8, Jun. 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, 47).

BRASIL. Lei nº 14.113, de 25 de dezembro de 2020. Regulamenta o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb), de que trata o art. 212-A da Constituição Federal; revoga dispositivos da Lei nº 11.494, de 20 de junho de 2007; e dá outras providências. *Diário Oficial da União*, Brasília, DF, 25 dez. 2020. Seção 1, p. 1. Edição Extra.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Indicador de nível socioeconômico das escolas de educação básica (Inse)*: nota técnica. Brasília, DF: Inep, 2014.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Estimativas de fluxo escolar a partir do acompanhamento longitudinal dos registros de aluno do Censo Escolar do período 2007-2016*: nota técnica. Brasília, DF: Inep, 2017.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Indicador de nível socioeconômico do Saeb 2019*: nota técnica. Brasília, DF: Inep, 2021.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Indicador de nível socioeconômico do Saeb 2021*: nota técnica. Brasília, DF: Inep, 2023a.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Serviço de Acesso a Dados Protegidos (Sedap)*. Brasília, DF, 2023b. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/gestao-do-conhecimento-e-estudos-educacionais/cgdi/servico-de-acesso-a-dados-protegidos-sedap>>. Acesso em: 22 nov. 2023.

BROER, M.; BAI, Y.; FONSECA, F. *Socioeconomic inequality and educational outcomes: evidence from twenty years of TIMSS*. Washington, DC: IEA; Springer, 2019. v. 5.

BUCHMANN, C. Measuring family background in international studies of education: conceptual issues and methodological challenges. In: PORTER, A. C.; GAMORAN, A. (Ed.). *Methodological advances in cross-national surveys of educational achievement*. Washington DC: National Academy Press, 2002. p. 150-197.

CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL (CEDEPLAR). *Estimativas populacionais municipais por idade simples e cálculo das taxas de escolarização líquida*. Belo Horizonte: Cedeplar, 2021. (material disponível sob solicitação).

CHALMERS, R. P. Mirt: a multidimensional item response theory package for the R environment. *Journal of Statistical Software*, [S. l.], v. 48, n. 6, p. 1-29, May 2012.

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

COSTA, M.; KOSLINSKI, M. C. Quase-mercado oculto: disputa por escolas “comuns” no Rio de Janeiro. *Cadernos de Pesquisa*, [S. l.], v. 41, n. 142, p. 246-266, jan./abr. 2011.

FIRPO, S.; JALES, H.; PINTO, C. Measuring peer effects in the Brazilian school system. *Applied Economics*, [S. l.], v. 47, n. 32, p. 3414-3438, Feb. 2015.

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.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Projeções da População: Brasil e unidades da Federação: revisão 2018*. 2. ed. Rio de Janeiro: IBGE, 2018.

MALAGUTH, T. Z. *Migrações e fluxo escolar da coorte de estudantes de 2008 a 2019, em Minas Gerais*. 2022. 129 f. Dissertação (Mestrado em Demografia) - Faculdade de Ciências Econômicas, Universidade Federal de Minas Gerais, Belo Horizonte, 2022.

OSÓRIO, R. *Um guia para o uso dos painéis da Pesquisa Nacional por Amostra de Domicílio (Pnad) Continua*. Brasília, DF, 2021. Webinar. Disponível em: <<https://www.youtube.com/watch?v=pJXmDiEOiAE>>. Acesso em: 13 nov. 2023.

RESENDE, T. F.; NOGUEIRA, C. M. M.; NOGUEIRA, M. A. Escolha do estabelecimento de ensino e perfis familiares: uma faceta a mais das desigualdades escolares. *Educação & Sociedade*, Campinas, v. 32, n. 117, p. 953-970, out./dez. 2011.

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*. [S. l.], 2018. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000266085/PDF/266085eng.pdf.multi.page=1&zoom=auto,-16,842>>. Acesso em: 13 nov. 2023.

RUTKOWSKI, D.; RUTKOWSKI, L. Measuring socioeconomic background in Pisa: one size might not fit all. *Research in Comparative and International Education*, [S. l.], v. 8, n. 3, p. 259-278, 2013.

SAMEJIMA, F. Estimation of latent ability using a response pattern of graded responses. *Psychometrika*, [S. l.], v. 34, p. 1-97, 1969.

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

SCIFFER, M. G.; PERRY, L. B.; MCCONNEY, A. The substantiveness of socioeconomic school compositional effects in Australia: measurement error and the

- relationship with academic composition. *Large-scale Assessments in Education*, [S. l.], v. 10, n. 21, p.1-22, Nov. 2022.
- SIRIN, S. R. Socioeconomic status and student achievement: a meta-analyse review of research. *Review of Educational Research*, [S. l.], v. 75, n. 3, p. 417-453, 2005.
- SOARES, J. F.; ALVES, M. T. G. Uma medida do nível socioeconômico das escolas brasileiras utilizando indicadores primários e secundários. *Opinião Pública*, Campinas, v. 29, n. 3, p. 575-605, set./dez. 2023.
- 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, jan./mar. 2006.
- SOARES, J. F.; COLLARES, A. C. M. Family resources and cognitive performance by primary school students in Brazil. *Dados*, Rio de Janeiro, v. 3, p. 1-37, 2007.
- SOARES, T. M. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. *Pesquisa Operacional*, [S. l.], v. 25, n. 1, p. 83-112, jan./abr. 2005.
- VAN EWIJK, R.; SLEEGERS, P. The effect of peer socioeconomic status on student achievement: a meta-analysis. *Educational Research Review*, [S. l.], v. 5, n. 2, p. 134-150, 2010.
- 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, v. 26, n. 61, p. 216-243, jan./abr. 2015.
- WILLMS, J. D. School composition and contextual effects on student outcomes. *Teachers College Record*, [S. l.], v. 112, n. 4, p. 1008-1037, Apr. 2010.

APÊNDICE

DETALHAMENTO DO CÁLCULO DO NSE SECUNDÁRIO

Essa etapa da metodologia utiliza o arquivo com a trajetória dos estudantes ao longo dos anos de 2007 a 2021. Para cada coorte, o banco de dados contém todas as combinações de estudante x ano x escola, refletindo a mudança de escola em seu percurso. Além dessas três variáveis, há também as 12 variáveis que correspondem ao NSE primário.

Em primeiro lugar, calcula-se o NSE secundário de cada estudante, que corresponde à média, na coluna, dos NSE primários das escolas em que o estudante esteve matriculado em sua trajetória, sempre que houver dado do NSE primário, conforme apresentado na Tabela A1. Observa-se que o estudante 1 possui 15 registros de matrícula, de 2007 a 2021, em sete escolas distintas (A a G). A escola A é exclusiva de educação infantil e não possuía nenhum NSE primário calculado. Ao final, a escola A recebe o NSE primário médio das escolas em que o estudante esteve matriculado (chamado aqui de NSE secundário do estudante) e que possuíam esse indicador. O resultado do cálculo do NSE secundário de cada estudante é um arquivo com a média de cada um dos NSE primários, refletindo as escolas em que ele esteve matriculado, como ilustrado na Tabela A2.

TABELA A1

ILUSTRAÇÃO DO CÁLCULO DO NSE SECUNDÁRIO DO ESTUDANTE E SUA ATRIBUIÇÃO ÀS ESCOLAS EM QUE O ESTUDANTE ESTEVE MATRICULADO DURANTE SUA TRAJETÓRIA ESCOLAR – BRASIL – 2007-2021

Ano	Estudante	Escola	A			B			C			D						
			NSE primário			Soma do NSE primário de todas as escolas que o estudante esteve matriculado durante sua trajetória escolar			Quantidade de vezes que o NSE primário aparece na trajetória escolar do estudante			NSE secundário do estudante						
			nse_p1	nse_p2	nse_p12	med_p1	med_p2	med_p12	mat_p1	mat_p2	mat_p12	nse_s_p1	nse_s_p2	nse_s_p12				
2007	1	A	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2008	1	A	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2009	1	A	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2010	1	B	49,9	50,8	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2011	1	B	49,9	50,8	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2012	1	B	49,9	50,8	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2013	1	B	49,9	50,8	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2014	1	C	50,7	51,3	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2015	1	D	51,1	50,9	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2016	1	D	51,1	50,9	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2017	1	E	49,9	52,0	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2018	1	E	49,9	52,0	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2019	1	F	46,2	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2020	1	G	51,5	49,8	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4
2021	1	G	51,5	49,8	...	601,6	560,0	...	236,9	...	12	11	...	5	50,1	50,9	...	47,4

Fonte: Elaboração própria.

TABELA A2
RESULTADO DO CÁLCULO DO NSE SECUNDÁRIO DO ESTUDANTE

Estudante	Escola	D			
		NSE secundário do estudante			
		nse_s_p1	nse_s_p2	...	nse_s_p12
1	A	50,1	50,9	...	47,4
1	B	50,1	50,9	...	47,4
1	C	50,1	50,9	...	47,4
1	D	50,1	50,9	...	47,4
1	E	50,1	50,9	...	47,4
1	F	50,1	50,9	...	47,4
1	G	50,1	50,9	...	47,4

Fonte: Elaboração própria.

Em segundo lugar, calcula-se o NSE secundário de cada escola, que corresponde à média, na coluna, de cada um dos NSE secundários dos estudantes que estiveram matriculados em cada escola, sempre que houver dado da média do NSE secundário dos estudantes. Uma ilustração desta etapa está presente nas Tabelas A3 e A4.

TABELA A3
ILUSTRAÇÃO DO CÁLCULO DO NSE SECUNDÁRIO DA ESCOLA

Escola	Estudante	A						B						C						D					
		NSE secundário do estudante						Somatório do NSE secundário do estudante por escola						Total de estudantes com informação de NSE secundário na escola						NSE secundário da escola					
		nse_sec_1	nse_sec_2	...	nse_sec_12	...	nse_sec_2	soma_sec_1	soma_sec_2	...	soma_sec_12	...	soma_sec_2	mat_sec_1	mat_sec_2	...	mat_sec_12	...	mat_sec_2	nse_sec_1	nse_sec_2	...	nse_sec_12	...	nse_sec_2
A	1	50,1	50,9	...	47,4	...	23886,6	23820,1	...	23292,1	...	470	470	...	470	470	...	470	50,8	50,7	...	49,6	...	50,7	49,6
A	23886,6	23820,1	...	23292,1	...	470	470	...	470	470	...	470	50,8	50,7	...	49,6	...	50,7	49,6
A	470	52,4	51,4	...	51,7	...	23886,6	23820,1	...	23292,1	...	470	470	...	470	470	...	470	50,8	50,7	...	49,6	...	50,7	49,6
B	1	50,1	50,9	...	47,4	...	12297,0	12381,3	...	11931,7	...	251	251	...	251	251	...	251	49,0	49,3	...	48,7	...	49,3	48,7
B	12297,0	12381,3	...	11931,7	...	251	251	...	251	251	...	251	49,0	49,3	...	48,7	...	49,3	48,7
B	251	49,1	49,2	...	48,6	...	12297,0	12381,3	...	11931,7	...	251	251	...	251	251	...	251	49,0	49,3	...	48,7	...	49,3	48,7
C	1	50,1	50,9	...	47,4	...	21120,5	21410,9	...	20081,1	...	429	429	...	429	429	...	429	49,2	49,9	...	49,5	...	49,9	49,5
C	21120,5	21410,9	...	20081,1	...	429	429	...	429	429	...	429	49,2	49,9	...	49,5	...	49,9	49,5
C	429	49,2	49,6	...	51,7	...	21120,5	21410,9	...	20081,1	...	429	429	...	429	429	...	429	49,2	49,9	...	49,5	...	49,9	49,5
D	1	50,1	50,9	...	47,4	...	15465,7	15826,7	...	15107,9	...	322	322	...	322	322	...	322	48,0	49,2	...	46,9	...	49,2	46,9
D	15465,7	15826,7	...	15107,9	...	322	322	...	322	322	...	322	48,0	49,2	...	46,9	...	49,2	46,9
D	322	47,2	46,2	...	46,3	...	15465,7	15826,7	...	15107,9	...	322	322	...	322	322	...	322	48,0	49,2	...	46,9	...	49,2	46,9
E	1	50,1	50,9	...	47,4	...	12328,4	12736,7	...	12466,7	...	256	256	...	256	256	...	256	48,2	49,8	...	48,7	...	49,8	48,7
E	12328,4	12736,7	...	12466,7	...	256	256	...	256	256	...	256	48,2	49,8	...	48,7	...	49,8	48,7
E	256	47,7	47,8	...	43,0	...	12328,4	12736,7	...	12466,7	...	256	256	...	256	256	...	256	48,2	49,8	...	48,7	...	49,8	48,7
F	1	50,1	50,9	...	47,4	...	10892,3	11120,5	...	10414,8	...	222	222	...	222	222	...	222	49,1	50,1	...	49,1	...	50,1	49,1
F	10892,3	11120,5	...	10414,8	...	222	222	...	222	222	...	222	49,1	50,1	...	49,1	...	50,1	49,1
F	222	49,2	49,6	...	51,7	...	10892,3	11120,5	...	10414,8	...	222	222	...	222	222	...	222	49,1	50,1	...	49,1	...	50,1	49,1
G	1	50,1	50,9	...	47,4	...	18746,1	18974,4	...	18283,8	...	381	381	...	381	381	...	381	49,2	49,9	...	49,7	...	49,9	49,7
G	18746,1	18974,4	...	18283,8	...	381	381	...	381	381	...	381	49,2	49,9	...	49,7	...	49,9	49,7
G	381	47,4	49,5	...	51,1	...	18746,1	18974,4	...	18283,8	...	381	381	...	381	381	...	381	49,2	49,9	...	49,7	...	49,9	49,7

Fonte: Elaboração própria.

TABELA A4
RESULTADO DO CÁLCULO DO NSE SECUNDÁRIO DA ESCOLA

Escola	D			
	NSE secundário da escola			
	nse_sesc_1	nse_sesc_2	...	nse_sesc_12
A	50,8	50,7	...	49,6
B	49,0	49,3	...	48,7
C	49,2	49,9	...	49,5
D	48,0	49,2	...	46,9
E	48,2	49,8	...	48,7
F	49,1	50,1	...	49,1
G	49,2	49,9	...	49,7

Fonte: Elaboração própria.

Por fim, o cálculo do NSE secundário para cada escola corresponde à média, na linha, dos 12 NSE secundários da escola. O resultado dessa construção metodológica é um banco de dados no qual as linhas são representadas pela totalidade de escolas com matrículas de escolarização em funcionamento em 2021, com quatro variáveis: código da escola, NSE secundário da Coorte 1, NSE secundário da Coorte 2 e NSE secundário da Coorte 3.

