

36º Encontro Anual da ANPOCS

GT06 – Desigualdade e Estratificação Social

**Medida de nível socioeconômico de alunos e
escolas com as informações das avaliações
educacionais em larga escala**

Maria Teresa Gonzaga Alves

José Francisco Soares

Universidade Federal de Minas Gerais

Grupo de Avaliação e Medidas Educacionais (GAME)

Águas de Lindoia, SP, 21 a 25 de outubro de 2012

1. Introdução¹

Este trabalho apresenta um índice de nível socioeconômico (NSE) para descrever as escolas de educação básica do Brasil. Os dados utilizados provêm de avaliações da educação básica feitas pelo Instituto Nacional de Estudo e Pesquisas Educacionais Anísio Teixeira (Inep), autarquia vinculada ao Ministério da Educação (MEC). Essas avaliações são compostas por testes cognitivos e questionários contextuais, que coletam informações sobre as características demográficas, sociais e familiares dos alunos das escolas selecionadas. Esta é uma pesquisa básica, cujos resultados poderão ser utilizados em outros estudos que investiguem temas relacionados à educação e também em estratificação social.

O desenvolvimento do índice de NSE seguiu várias etapas, que serão descritas a seguir. Na próxima seção deste texto, será revista, brevemente, a literatura que justifica as dimensões incluídas no cálculo do índice de NSE. Em seguida, na metodologia, serão apresentadas as bases de dados utilizadas, a preparação das mesmas, os critérios para a seleção e críticas dos itens dos questionários contextuais, a montagem do banco de dados usado neste trabalho e os procedimentos estatísticos utilizados para a estimação do NSE. Na seção dos resultados serão apresentadas as análises descritivas do índice de NSE e a validação do mesmo. Nas considerações finais, serão discutidos as aplicações e os limites do índice de NSE calculado por escola.

2. Fundamentação teórica

A definição de nível socioeconômico, na literatura, não é unívoca. Não há um consenso absoluto sobre quais as dimensões devem ser consideradas para medir esse construto nas pesquisas sociais quando se pretende distinguir a posição dos indivíduos ou famílias na estrutura social. As decisões dos

¹Os autores gostariam de agradecer ao CNPq e ao Instituto Unibanco pelos apoios concedidos para a execução do presente trabalho.

pesquisadores dependem de justificativas teóricas, mas também da disponibilidade de dados empíricos que expressem as dimensões importantes do construto. Neste trabalho, assume-se que o nível socioeconômico é um construto teórico que sintetiza as características dos indivíduos em relação à sua renda, ocupação e escolaridade, permitindo a criação de estratos ou classes de indivíduos semelhantes em relação a estas características.

O estudo das diferenças sociais ocupa posição central na literatura sociológica contemporânea. Embora o interesse pelo tema tenha raízes na sociologia clássica, foi nos Estados Unidos, a partir dos anos 1950, no âmbito das pesquisas em estratificação social e mobilidade social, que as medidas empíricas de nível socioeconômico passaram a ser operacionalizadas (DUNCAN, 1993; BLAU; DUNCAN, 1967). Tais pesquisas foram impulsionadas pelos avanços metodológicos propiciados pela aplicação da estatística com o auxílio computacional aos dados populacionais – censitários ou por amostragem – o que também ocorreu no Brasil algumas décadas depois (PASTORE, 1979; PASTORE; SILVA 2000; SCALON, 1998; SANTOS, 2005, por exemplo).

Entre os sociólogos, os indicadores utilizados nas pesquisas são quase sempre relacionados à posição ocupacional. Nos estudos empíricos, parte-se do registro de todas as ocupações de uma sociedade, por exemplo, as que aparecem nos censos demográficos. Em seguida, estas ocupações são codificadas em grupos sociologicamente relevantes. De acordo com as preferências e questões substantivas do pesquisador atribui-se aos diferentes grupos um valor considerando, por exemplo, suas características quanto às credenciais exigidas (ou nível de instrução), o grau de especialização e responsabilidades envolvidas, bem como o retorno financeiro. Há estudos que argumentam sobre a importância de outras informações como a idade do indivíduo, o tempo de experiência na ocupação e o gênero (AGUIAR; FERNANDES; NEVES, 2007). Dependendo da tradição teórica e da finalidade do estudo, os autores classificam os indivíduos em categorias distintas ou então em estratos contínuos, que são utilizados de forma semelhante em muitos estudos internacionais. Por exemplo, os estratos de classe proposto por Erikson e

Goldthorpe (1992) ou a escala de NSE desenvolvida por Gazenboom, de Graaf e Traiman (1992) para uso em pesquisas comparativas internacionais.

Como se percebe, a operacionalização de indicadores relacionados à posição ocupacional dos indivíduos é bastante complexa. A identificação precisa da posição ocupacional depende no mínimo do conhecimento do título da ocupação (por exemplo, professora) e das atividades executadas na ocupação (por exemplo, ministra aulas no ensino fundamental). Adicionalmente, informações como o nível na hierarquia (direção, supervisão, por exemplo), o tempo de exercício e o setor econômico (público ou privado) são também requeridas. Hauser e Warren (1997) estimam que sejam necessários de dois a três minutos do tempo de aplicação de um questionário para se coletar as informações importantes para a descrição da ocupação e, depois, até quinze minutos para a correta classificação segundo algum critério previamente definido. O tempo e recursos exigidos para a análise das informações e para a codificação das ocupações restringem esse modelo a poucos estudos.

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

O Critério de Classificação Econômica Brasil (ou simplesmente Critério Brasil), que é adotado pela Associação Brasileira de Empresas de Pesquisa (Abep) com

o objetivo de estimar o poder de compra das famílias urbanas, tem princípios bastante semelhantes, mas operacionalmente é bem mais exequível². O cálculo do Critério Brasil depende de informações diretas, que a maioria das pessoas consegue responder, tais como os bens e serviços disponíveis no domicílio e a escolaridade do chefe da família. A sua fórmula de cálculo é simples, pois envolve apenas uma soma de pontos para situar os indivíduos nas classes (A1, A2, B1, B2, C1, C2, D e E). Os resultados são considerados bastante fidedignos para inferir sobre o poder de compra das famílias urbanas com rendimento de até cerca de trinta salários mínimos, o que compreende a maioria absoluta da população. Por estes motivos, o Critério Brasil tem ampla aplicação nas pesquisas de mercado e é adotado também por outros estudos que necessitam segmentar os indivíduos para estudar temas variados, como o comportamento eleitoral, avaliação e acesso aos serviços públicos ou privados etc.

Nas pesquisas educacionais o debate sobre classes e estratos sociais é um tema central. Não só por que a educação é uma das dimensões fundamentais nos estudos de estratificação e mobilidade social, mas, também, devido a forte correlação entre resultados escolares e o nível socioeconômico e cultural das famílias, comprovada por ampla evidência empírica em vários países do mundo desde meados dos anos 1960 (COLEMAN et al, 1966; JENCKS et al, 1972) e também no Brasil (ALVES; FRANCO, 2008; FERRÃO; FERNANDES, 2001; SOARES, 2005; SOARES; ALVES, 2003; SOARES, CÉSAR; MAMBRINI, 2001).

Assim, nas avaliações educacionais a elaboração de questionários contextuais, que contenham as informações relevantes e que sejam fidedignos, constitui uma etapa crucial do processo avaliativo (FRANCO, et al, 2003). Um modelo que inspira avaliações educacionais em vários países do mundo foi proposto pelo Programa Internacional de Avaliação de Estudantes, cuja sigla é PISA. Para contextualizar os resultados comparativos entre países, o PISA desenvolveu o índice de status socioeconômico, conhecido por ISEI (*International Socio-Economic Index of Occupational Status*), que sintetiza as dimensões teóricas

² O Critério Brasil é desenvolvido a partir das informações coletadas pelo Levantamento Socioeconômico realizado anualmente pelo IBOPE, por meio de uma pesquisa probabilística de domicílios nas dez principais regiões metropolitanas do país. Informações em: <http://www.abep.org/novo/Content.aspx?SectionID=84>, consultado em 2 de abril de 2012.

destacadas nos estudos em estratificação social. O ISEI é derivado da informação sobre a ocupação dos pais, a qual é atribuída um escore escalonado de acordo com a educação e a renda da família medida de forma indireta pela posse de bens.

Um ponto crítico na tarefa de coletar esses indicadores nas pesquisas educacionais decorre do fato de que os questionários são, geralmente, respondidos pelos próprios alunos na sala de aula. Isto não ocorre nas pesquisas populacionais que contam com entrevistadores treinados para aplicar e preencher os questionários domiciliares. Por isto, a definição dos indicadores importantes para a contextualização dos resultados, a formulação dos itens com clareza e precisão, a extensão dos questionários, a adequação à faixa etária e a possibilidade de itens total ou parcialmente não respondidos são elementos críticos a serem considerados. Contudo, a experiência acumulada pelas agências de avaliação educacional tem produzido resultados satisfatórios.

Tendo em vista esses referenciais, os dados educacionais disponíveis e a experiência anterior (SOARES; ANDRADE, 2006; ALVES; SOARES, 2009), assume-se neste trabalho que o nível socioeconômico é um construto latente, ou seja, não diretamente observável, e que sintetiza unidimensionalmente a relação entre a escolaridade, a ocupação e a renda das famílias. Entretanto, estas variáveis não estão diretamente contempladas em todos os questionários contextuais das avaliações educacionais do Inep. Assim, o desenvolvimento desse índice exigiu decisões metodológicas para operacionalizar o construto, descritas a seguir.

3. Metodologia

O índice de nível socioeconômico construído nesta pesquisa foi estimado com base em uma ampla evidência empírica, com cobertura de quase uma década de avaliações educacionais. Utilizaram-se dados produzidos Inep, nas avaliações educacionais, que estão disponíveis para acesso público. Embora essas avaliações não se constituam em uma amostra aleatória das famílias

brasileiras, elas têm representatividade bastante satisfatória das escolas de educação básica, pois compreendem um amplo espectro de perfis de alunos.

A partir destas diferentes fontes, um banco de dados único e representativo dos “perfis de famílias de alunos que participaram das avaliações” foi construído. Devido aos diversos formatos com que as bases de dados originais são apresentadas, a construção do banco de dados manejável exigiu inúmeras tarefas tais como: a obtenção dos dados e preparação dos mesmos em bancos de dados manipuláveis; a análise dos questionários contextuais e seleção dos itens relacionados às dimensões do índice de NSE; a compatibilização dos itens por meio de recodificações; a ponderação dos dados; a análise estatística dos itens; a estimação do NSE; a preparação do banco de dados final com o índice de NSE; e a identificação das escolas. Depois de descrever cada uma destas etapas, serão apresentados os resultados obtidos.

3.1. Os dados

Os dados analisados neste trabalho provêm das avaliações da educação básica. Foram consideradas as seguintes bases de dados: o Sistema de Avaliação da Educação Básica (Saeb), nas edições de 2001, 2003 e 2005; a Prova Brasil de 2005, 2007 e 2009; e Exame Nacional do Ensino Médio (Enem) realizados nos anos de 2007, 2008 e 2009. Todas estas avaliações são acompanhadas por questionários contextuais que contêm indicadores relacionados ao construto de interesse desta pesquisa. Cada uma das modalidades de avaliação contribui com informações específicas que, no conjunto, compõem um quadro bem rico dos tipos de escolas espalhadas pelo país.

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

A inclusão do Saeb na base construída neste estudo é importante por que as suas respectivas amostras constituem evidências estatisticamente representativas de todos os tipos de escolas. O seu uso possibilita a ligação entre as escolas públicas e privadas, já que a amostra do Saeb inclui alunos de ambos os tipos de redes. Assim, os alunos da amostra Saeb têm um papel importante no processo de estimação dos parâmetros dos itens que compõem o índice de NSE. Isto permite um melhor ajuste do modelo de análise estatístico adotado no trabalho em comparação ao que seria obtido com uma base de dados composta só por um tipo de escola. No entanto as escolas do Saeb não são identificadas e, portanto, o seu respectivo NSE não é utilizado nos outros aspectos dessa pesquisa.

A Prova Brasil avalia, a cada dois anos, as escolas públicas urbanas que oferecem o ensino fundamental, o que compreende em torno de 85% das matrículas neste nível de ensino. Participam desta avaliação alunos das séries finais dos segmentos do ensino fundamental (5º e 9º ano). Devido ao uso dos resultados da Prova Brasil em vários programas de apoio educacional no nível federal, a adesão à Prova Brasil tem sido crescente desde sua implantação em 2005. Nas bases de Prova Brasil há os códigos das escolas, cujos nomes e localização puderam ser consultados no sistema do Inep na etapa final deste trabalho.

O Enem é um exame anual e voluntário, ao qual se submetem alunos concluintes do ensino médio. Embora não seja obrigatório, o escore obtido neste exame possui finalidades bastante motivadoras para a participação – uso nos vestibulares de para universidades públicas, bolsas para universidades privadas e outros – o que têm estimulado a adesão crescente, tornando-o bastante representativo das escolas de ensino médio das redes pública e privada em todo o país.

Neste trabalho, foram considerados os dados somente dos alunos que terminaram o ensino médio no ano de aplicação do Enem. Optou-se por não utilizar os dados anteriores a 2007 porque eles não possuem o código de identificação de escola (no lugar há uma máscara, ou número fantasia que

impedem a identificação da escola) e a finalidade de calibrar a base geral para estimar o NSE das escolas já estava garantida pelo Saeb que possui uma amostra criteriosamente selecionada. Também não se considerou a base de dados do Enem 2010, embora disponível, porque houve uma mudança profunda no questionário contextual aplicado nesta edição, o que o tornou completamente atípico em comparação a todas as outras bases de dados.

Em avaliações educacionais, é comum que uma parte dos alunos faça o teste, mas deixe o questionário totalmente em branco. Por exemplo, de 4.148.721 casos registrados na base de dados do Enem 2009, 1.339.445 eram alunos concluintes naquele ano, mas destes apenas 848.063 responderam o questionário contextual, mesmo que parcialmente³. Quando isto ocorreu, estes casos foram excluídos das respectivas bases de dados. Posteriormente, outra exclusão de casos foi feita em função da “não resposta” à maioria dos itens considerados para a estimação do NSE, conforme será explicado a seguir.

No total, foram selecionadas dezoito bases de dados, visto que em cada edição do Saeb e da Prova Brasil há informações sobre três e duas séries, respectivamente. Os números de alunos e escolas, em cada uma das bases de dados, após a primeira seleção de casos estão sintetizados no Quadro 1.

³ Embora o número de inscritos no Enem seja crescente, o percentual de alunos que não responderam ao questionário contextual aumentou continuamente no período considerado neste trabalho. Em 2007, entre os concluintes daquele ano, 22% deixaram o questionário Enem totalmente em branco. Em 2009, o percentual de alunos concluintes que não respondeu chegou a 36,7%.

Quadro 1: fontes dos dados, ano da avaliação, número de alunos e escolas (não cumulativo*)

No	Fonte	Ano	Série/ Ano	Alunos	Escolas
1	Saeb	2001	4 ^a	114.512	4.049
2	Saeb	2001	8 ^a	100.792	2.818
3	Saeb	2001	11 ^a	72.415	1.834
4	Saeb	2003	4 ^a	98.188	3.573
5	Saeb	2003	8 ^a	84.600	2.575
6	Saeb	2003	11 ^a	52.484	1.319
7	Saeb	2005	4 ^a /5 ^o	83.929	2.811
8	Saeb	2005	8 ^a /9 ^o	66.353	1.358
9	Saeb	2005	11 ^a /12 ^o	66.825	682
10	Prova Brasil	2005	4 ^a /5 ^o	1.927.280	29.250
11	Prova Brasil	2005	8 ^a /9 ^o	1.348.541	20.847
12	Prova Brasil	2007	4 ^a /5 ^o	2.3103.02	37.483
13	Prova Brasil	2007	8 ^a /9 ^o	1.798.963	27.381
14	Prova Brasil	2009	5 ^o	2.539.911	43.582
15	Prova Brasil	2009	9 ^o	1.995.390	31.963
16	Enem	2007	12 ^o (**)	979.585	22.985
17	Enem	2008	12 ^o (**)	1.029.902	24.379
18	Enem	2009	12 ^o (**)	848.036	28.446

(*) Os números de alunos e escolas não são cumulativos, por que um mesmo aluno ou escola pode estar presente em diferentes avaliações.

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

3.2. Seleção e preparação das variáveis

A identificação dos itens dos questionários contextuais para a estimação do índice de NSE se deu a partir da leitura dos questionários ou livros de códigos das avaliações. De uma maneira geral, todos os questionários contêm itens relacionados ao nível socioeconômico das famílias. Eles incluem, também, itens

relacionados ao acesso a serviços públicos (luz, água, rua pavimentada), condições construtivas da moradia e o consumo cultural das famílias, que, embora relacionados, extrapolam a definição teórica adotada neste trabalho.

Na leitura dos questionários, todos os itens que guardassem relação, direta ou indireta, com a posição ocupacional e nível de instrução dos pais dos alunos e a renda familiar foram considerados. Foram destacados 22 itens nas diferentes bases de dados. Nesta etapa, constatou-se uma grande variação na formulação dos itens entre as avaliações e também entre as edições de uma mesma avaliação. Isto, porém, não inviabilizou o uso das bases de dados de forma conjunta, mas aumentou muito o tempo para a preparação dos dados e algumas perdas no nível de qualidade da informação de alguns deles.

Em seguida, os itens selecionados foram recodificados, de forma que valores mais altos foram atribuídos às categorias ordinais que indicam uma posição mais favorável em relação ao construto e de forma padronizada em todas as avaliações. Devido às diferentes formulações nas questões, os itens com menos categorias ordinais se tornaram obrigatoriamente padrões na recodificação. Isto prejudicou especialmente alguns itens relacionados ao nível de renda da família (“banheiro”, “TV” e “Rádio”), que perderam a capacidade de discriminação pela redução do número de categorias. Contudo, a inclusão de outros indicadores garantiu a qualidade do modelo de análise estatístico.

Nesta etapa, considerou-se importante analisar as ocorrências de “ausência de informação” nas variáveis destacadas para o cálculo do índice de NSE em cada um dos bancos de dados. O percentual de não resposta aos itens varia muito entre as bases de dados, mas a maioria dos alunos responde dez itens ou mais. Por isto, decidiu-se que seriam considerados apenas os alunos que responderam a partir desse limite.

Com todos os bancos de dados prontos isto é, com todas as variáveis padronizadas, calculou-se a correlação entre as variáveis indicadoras do construto a partir de uma amostra dos dados. Como se tratam de variáveis ordinais utilizou-se a correlação policórica, conforme implementada no software

R. A correlação indicou que as variáveis “trabalho do aluno” e “geladeira duplex” (ambas indicadoras do nível de renda da família) não fornecem informações consistentes com o construto latente, definido pelas outras variáveis. A evidência é que o “trabalho do aluno” não reflete necessariamente o seu nível socioeconômico porque há alunos que trabalham e têm NSE mais alto e alunos que não trabalham e têm NSE mais baixo⁴. O item “geladeira duplex” não está em todos os questionários contextuais e, quando incluído, tem formulação confusa para o tipo de respondente. Talvez por estas razões esse item não contribua para a medida do construto como esperado.

Decidiu-se ainda excluir as variáveis “alfabetização” da mãe e do pai (indicadoras do nível de escolaridade) e manter apenas a variável “grau de escolaridade”. As variáveis de “alfabetização”, embora com correlação positiva com o construto, estão presentes em poucas bases de dados e, de certa forma, elas já estão contempladas no “grau de escolaridade”. Isto reduziu a incidência de perfis raros de tipos de alunos na matriz que reúne todas as combinações possíveis entre as variáveis utilizadas na etapa final do trabalho. Para a estimação do índice de NSE, ficaram, ao final, dezoito variáveis, que estão sintetizadas no Quadro 2 (no final desta seção), que contém os indicadores, itens selecionados, os rótulos das variáveis e as categorias de respostas de cada um deles após o trabalho de recodificação, bem como as bases de dados nas quais eles foram selecionados.

Finalmente, em cada um dos bancos de dados, foram incluídas variáveis correspondentes aos itens não contemplados nos respectivos questionários, com o código de “dado ausente” para todos os casos. Por exemplo, o item “renda familiar” está presente somente nos questionários das três edições do Enem. Para criar um banco de dados único com dados de todos os alunos, criou-se esta variável nos outros bancos de dados (Saeb e Prova Brasil) a qual foi atribuída o código de “dado ausente” para todos os casos. A ausência de

⁴ Alves Brito (2011) pesquisou os padrões de trabalhos entre crianças e adolescentes no Brasil, no período de 1988 a 2005 e constatou que este fenômeno não tem uma lógica única e linear. Há uma enorme variação quanto aos padrões de alocação do tempo com o trabalho nessa faixa etária, mas também mudanças no período analisado.

informação em parte dos casos não inviabiliza o processo de estimação do NSE devido à especificidade do modelo de análise, conforme explicado a seguir.

Quadro 2: Indicadores, itens, descrição, categorias e fonte de dados

Indicador	Item	Descrição	Categorias	Fonte (bases de dados com item disponível)
Nível educacional dos pais	EscolaMãe	Até que série sua mãe ou a mulher responsável por você estudou?	1. Nunca estudou/não completou a 4ª série. 2. Completou a 4ª série, mas não completou a 8ª série. 3. Completou a 8ª série, mas não completou o Ensino Médio. 4. Completou o Ensino Médio, mas não completou a faculdade. 5. Completou a faculdade. 9. Não sabe/Sem Informação.	Saeb: 2001, 2003, 2005 (4º/8ª/11ª); Prova Brasil: 2005, 2007, 2009 (5º / 9º); Enem: 2007, 2008, 2009
	EscolaPai	Até que série seu pai ou o homem responsável por você estudou?		
Nível de rendimento da família	Banheiro2	Em sua casa tem banheiro?	1. Não tem 2. Tem um 3. Tem mais de um 9. Sem informação	Saeb: 2001, 2003, 2005, (4º/8ª/11ª); Prova Brasil: 2005, 2007, 2009 (5º / 9º)
	TV2	Em sua casa tem televisão em cores?	1. Não tem 2. Tem uma 3. Tem mais de uma 9. Sem informação	Saeb: 2001, 2003, 2005 (4º/8ª/11ª); Prova Brasil: 2005, 2007, 2009 (5º / 9º); Enem: 2007, 2008, 2009
	Rádio2	Em sua casa tem rádio?	1. Não tem 2. Tem um 3. Tem mais de um 9. Sem informação	Saeb: 2001, 2003, 2005 (4º/8ª/11ª); Prova Brasil: 2005, 2007, 2009 (5º / 9º); Enem: 2007, 2008, 2009
	Geladeira	Em sua casa tem geladeira?	1. Não tem 2. Tem uma ou mais 9. Sem informação	Saeb: 2001, 2003, 2005 (4º/8ª/11ª); Prova Brasil: 2005, 2007, 2009 (5º / 9º); Enem: 2007, 2008, 2009
	Freezer	Em sua casa tem freezer?	1. Não tem 2. Sim, tem 9. Sem informação	Saeb: 2001, 2003, 2005 (4º/8ª/11ª); Prova Brasil: 2005, 2007, 2009 (5º / 9º),
	Máquina	Em sua casa tem uma máquina de lavar roupa (não é tanquinho)?	1. Não tem 2. Sim, tem 9. Sem informação	Saeb: 2001, 2003, 2005 (4º/8ª/11ª); Prova Brasil: 2005, 2007, 2009 (5º / 9º); Enem: 2007, 2008, 2009
	Aspirador	Em sua casa tem um aspirador de pó?	1. Não tem 2. Sim, tem 9. Sem informação	Saeb: 2001, 2003, 2005 (4º/8ª/11ª); Prova Brasil: 2005, 2007(5º / 9º).
	Vídeo_DVD	Em sua casa tem videocassete?	1. Não tem 2. Tem 9. Sem informação	Saeb: 2001, 2003, 2005(4º/8ª/11ª); Prova Brasil: 2005, 2007, 2009 (5º / 9º); Enem: 2007, 2008 , 2009

Continua

Indicador	Item	Descrição	Categorias	Fonte (bases de dados com item disponível)
Nível de rendimento da família	Computador	Em sua casa tem um computador?	1. Não 2. Sim, sem acesso à internet 3. Sim, com acesso à internet. 9. Sem informação	Saeb: 2001, 2003, 2005 (4 ^o /8 ^a /11 ^a); Prova Brasil: 2005, 2007, 2009 (5 ^o / 9 ^o); Enem: 2007, 2008, 2009
	Automóvel	Em sua casa tem automóvel?	1. Não tem 2. Sim, tem um. 3. Tem mais de um 9. Sem informação	Saeb: 2001, 2003, 2005 (4 ^o /8 ^a /11 ^a); Prova Brasil: 2005, 2007, 2009 (5 ^o / 9 ^o); Enem: 2007, 2008, 2009
	Empregada	Em sua casa trabalha empregada doméstica?	1. Não tem 2. Sim, uma diarista (faxineira, etc.) uma ou duas vezes por semana. 3. Uma ou mais, todos os dias. 9. Sem informação	Saeb: 2001, 2003, 2005 (4 ^o /8 ^a /11 ^a); Prova Brasil: 2005, 2007, 2009 (5 ^o / 9 ^o)
	RendaF	Renda Familiar (somando a do respondente e a das pessoas que moram com ele)	1. Nenhuma/ Até 1 salário mínimo 2. De 1 a 2 salários mínimos 3. De 2 a 5 salários mínimos 4. De 5 a 10 salários mínimos 5. Mais de 10 salários mínimos 9. Sem informação/Não sabe	Enem:, 2007, 2008, 2009
Ocupação dos pais	ÁreaM	Área que a mãe trabalha ou trabalhou	1. Área A – mais informal/primário [agrupamento dos itens: (A) Na agricultura, no campo, em fazenda ou na pesca. (F) Trabalhador do setor informal (sem carteira assinada). (G) Trabalha em casa em serviços (costura, cozinha, aulas particulares etc.).] 2. Área B – mais formalizado [agrupamento dos itens: (B) Na indústria. (C) No comércio, banco, transporte ou outros serviços. (D) Funcionário público do governo federal, estadual ou municipal. (E) Profissional liberal, professor ou técnico de nível superior.]; 9. No lar/ Não trabalha/ Não sabe/Sem informação	Enem: 2007, 2008, 2009
	AreaP	Área que o pai trabalha ou trabalhou		

Continua

Indicador	Item	Descrição	Categorias	Fonte (bases de dados com item disponível)
Ocupação dos pais	OcupaçãoM	Posição da ocupação da mãe	1. Nível 1 [agrupamento dos itens: (G) Trabalho temporário, informal, sem carteira assinada. (H) Trabalho por conta própria.] 2. Nível 2 [agrupamento dos itens: (D) Empregado no setor privado, com carteira assinada. (E) Funcionário público (federal, estadual ou municipal), sem função de direção. (F) Militar (guarda-civil, polícia estadual ou Forças Armadas), sem posto de comando.] 3. Nível 3 [agrupamento dos itens: (A) Gerente, administrador ou diretor de empresa privada. (B) Funcionário público (federal, estadual, municipal), com funções de direção. (C) Militar (guarda-civil, polícia estadual ou Forças Armadas), com posto de comando.	Enem: 2007, 2008
	OcupaçãoP	Posição da ocupação do pai	9. Desempregado Aposentado/ Outra situação/ /Sem informação	

3.3. Preparação do banco de dados final: perfis e ponderação

Após as análises dos itens e a seleção de casos na etapa anterior, restaram 12.040.106 alunos com informações válidas para a estimação do índice de NSE. Dentre estes milhões de alunos, há vários cujos padrões de resposta aos itens selecionados são absolutamente idênticos. Estes constituem perfis de alunos que podem ser agregados – reduzindo-se o número de casos na base de dados – e as suas ocorrências computadas e analisadas de forma ponderada. Em termos computacionais, é muito mais viável estimar o índice de NSE com base em perfis de alunos ponderados do que para cada um deles individualmente. Neste processo, um perfil de aluno com muitas ocorrências terá um peso maior na estimação do NSE do que um perfil mais raro.

Esta é a opção adotada neste trabalho. Criou-se um banco de dados com os perfis de alunos segundo as possíveis combinações entre as variáveis e registrando-se as ocorrências – i.e., o número de alunos – destes perfis em cada um dos bancos de dados.

Entretanto, algumas considerações tiveram que ser feita antes de se computar os perfis e continuar o processo de criação do banco de dados final. As bases de dados da Prova Brasil e do Enem são constituídas pelas populações-alvo que participam das avaliações por adesão dos sistemas de ensino (no caso da Prova Brasil) ou por inscrição voluntária no teste (no caso do Enem). As escolas destes alunos não são a priori selecionadas por amostragem representativa de suas respectivas populações. Porém, o crescente número de participantes nesses exames e a presença de alunos da maioria das escolas que oferecem ensino médio no país os tornam, em termos gerais, bastante representativos. Assim, com os dados dessas avaliações, cada perfil que se configurou no processo de agregação das variáveis teve os seus respectivos números de alunos somados.

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

Para evitar distorções que ocorreriam com a agregação simples de bancos de dados com números de alunos muito diferentes corrigiu-se a proporção das ocorrências em cada perfil nas respectivas avaliações fixando-se o valor de um milhão de casos para todas as bases. Isto foi feito por regra de três simples e teve como consequência garantir que todas as bases tiveram a mesma influência no processo de cálculo do NSE.

A agregação das dezoito variáveis para a estimação do índice de NSE produziu um banco de dados com 1.622.627 perfis. Diante de cada um, registrou-se o número de ocorrências nas respectivas bases de dados pela soma simples ou ponderada do número de alunos que se enquadravam em cada um deles. Ao final, a soma das ocorrências na nova grandeza corresponde aos pesos utilizados por perfil no processo de estimação do índice de NSE.

3.4. Modelo de análise

Para agregar os itens em uma única medida do índice de NSE empregou-se um modelo da Teoria da Resposta ao Item (TRI). A TRI engloba uma série de modelos matemáticos cujo objetivo principal é a obtenção de medidas de construtos latentes baseadas na probabilidade de um indivíduo escolher uma determinada resposta a um item dicotômico e/ou ordinal (HAMBLETON, 1973). Uma vantagem da TRI em relação aos métodos convencionais é a possibilidade de estimar a medida do NSE mesmo com dados incompletos para algum dos indicadores, capacidade crucial para este estudo, cujos dados provêm de questionários com itens parcialmente em comum.

O modelo da TRI escolhido para a criação do índice de NSE é o especificado para respostas graduadas, desenvolvido por Samejima (1969). Trata-se de um modelo utilizado para estimação de um construto latente através de itens cujas opções de respostas possuam uma ordenação. Devido ao método utilizado, são considerados apenas os itens para os quais os indivíduos forneceram resposta, desconsiderando aqueles em branco. Ou seja, considerou-se não apenas a resposta ao item, mas também quantos alunos (soma ponderada dos perfis) responderam daquela forma. Com a correção dos perfis para a ordem de um milhão, realizada na etapa final de preparação da base de dados, alguns perfis muito raros ficaram reduzidos a valores ínfimos, e, portanto, com peso equivalente a zero. Estes foram mantidos na base de dados, embora não tenham contribuído para o processo de estimação dos parâmetros do modelo. No entanto, tiveram o seu NSE calculado da mesma forma que os outros padrões.

4. Resultados

4.1. Análise dos itens

A qualidade de cada item como componente da medida característica latente – o índice de NSE – é avaliada analisando-se a curva característica do item e a sua curva de informação.

Por exemplo, o gráfico 1 apresenta a curva da variável “automóvel” (um item do nível de renda familiar), que possui três categorias ordinais. No eixo das abscissas está colocado o índice de NSE dos alunos e, no eixo das ordenadas, a probabilidade de um indivíduo com dado NSE está em cada uma das categorias. Pode-se notar que as alternativas 1 e 3 (respectivamente, as opções “não possui” e “tem mais de um automóvel”) correspondem aos extremos da escala. Além disso, à medida que aumenta o NSE, diminui a probabilidade da opção 1 e, mais a frente na escala, da opção 2 (“tem um automóvel”).

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

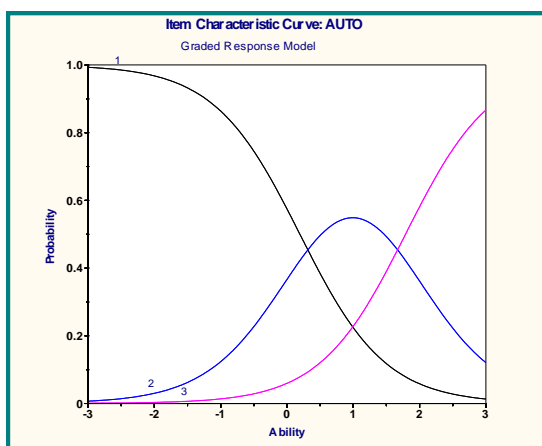


Gráfico 1: curva característica do item automóvel

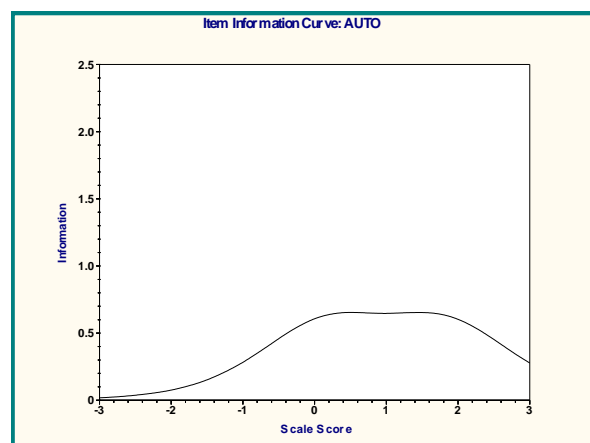


Gráfico 2: curva de informação do item automóvel

O item “geladeira” cumpre este papel. Embora um bem de consumo dos mais comuns, há ainda famílias que, no momento da resposta do questionário pelo

aluno, não o possuíam. Isto se nota no extremo inferior da escala do índice de NSE, no gráfico 3. O índice é dicotômico (“tem geladeira” ou “não tem”), logo a curva característica indica o ponto de inflexão, a partir do qual praticamente todas as famílias passam a responder que “tem geladeira”. Este ponto está situado abaixo do nível -1, na escala de NSE. A curva de informação do item mostra que ele é importante para distinguir famílias com NSE situado no extremo inferior da escala.

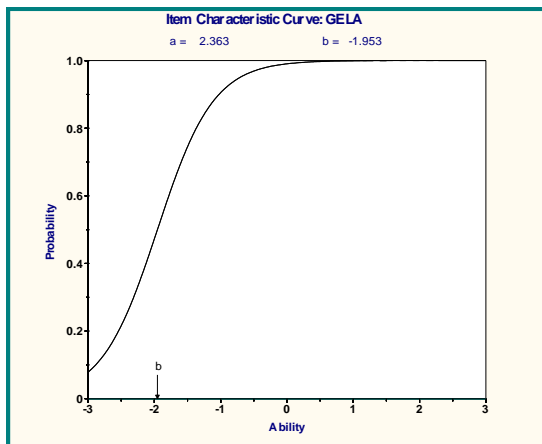


Gráfico 3: curva característica do item geladeira

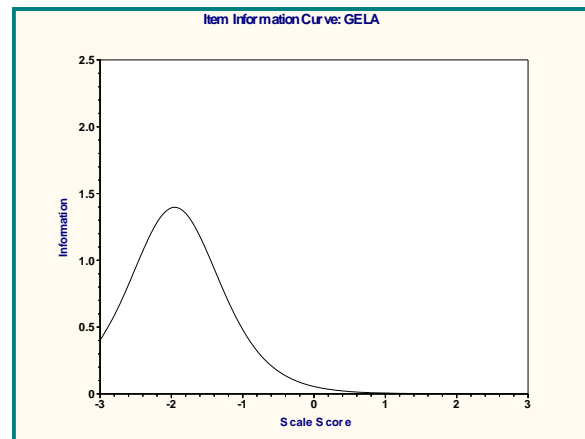


Gráfico 4: curva de informação do item geladeira

Outro exemplo interessante para ilustrar a análise dos itens é o referente à “posição ocupacional dos pais do aluno”. As categorias originais do item foram agrupadas em três níveis, sendo o mais baixo o referente às ocupações informais e por conta própria; o nível intermediário, referente aos empregos formais; e o nível mais alto, agrupando os pais com com posição de chefia ou direção em empregos formais. Os gráficos 5 e 6 mostram as curvas do item referente a ocupação da mãe do aluno. Nota-se, no gráfico da esquerda, que a probabilidade de se encontrar entre os perfis mães em cargos de direção só cresce nos extremo superior da escala de NSE. Este item contribui para a distinção de famílias com NSE mais alto, conforme o gráfico da direita, porque a partir do ponto médio da escala a curva de informação segue constante. É muito difícil distinguir a posição de famílias situadas nos extremos de qualquer escala de NSE. Mais ainda entre as famílias de alta posição. Não obstante, conseguiu-

se, com este item, atingir um bom nível de informação sobre perfis de famílias mais privilegiadas.

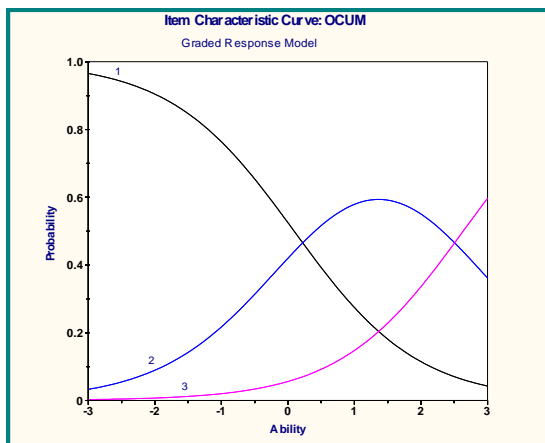


Gráfico 5: curva característica do item posição da ocupação da mãe

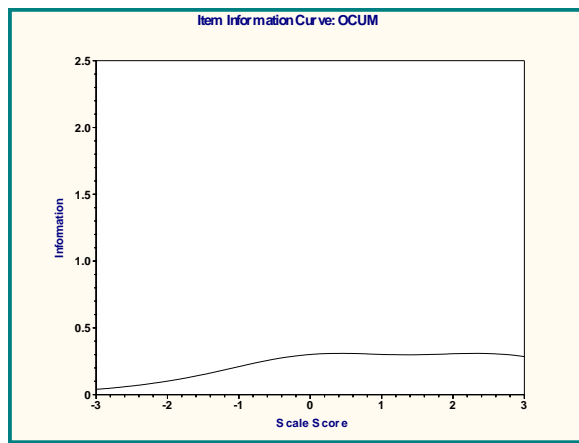


Gráfico 7: curva característica do item posição da ocupação da mãe

A mesma análise foi realizada com todos os itens. Os resultados revelaram-se bem satisfatórios quanto a relação dos itens com o construto latente, bem como o grau de informação que eles trazem sobre o perfis das famílias ao longo da escala de NSE.

Outra evidência importante sobre a qualidade dos itens se faz pela comparação entre as frequências esperadas com as frequências observadas das categorias de cada um deles. As proporções obtidas, de uma forma geral, são muito próximos, o que denota um alto grau de precisão nas estimativas. A tabela com esses resultados não está incluída no trabalho por motivo de espaço.

4.2. Medida de NSE

O índice de NSE estimado para os perfis dos alunos incluídos produz uma medida padronizada com média em torno de zero. A tabela 1 apresenta as estatísticas descritivas do índice de NSE para todos os perfis.

Tabela 1: Estatísticas descritivas do índice de NSE por perfil de alunos (valores sem ponderação)

	N	Mínimo	Máximo	Média	Desvio-Padrão
NSEPerfil	1.622.627	-2,31	2,31	0,06	0,57

Fonte: elaboração própria

O gráfico 7 apresenta o histograma do NSE para os perfis ponderados pelo número de alunos incluídos em cada perfil. Notas-se que há perfis de alunos ao longo de praticamente toda a distribuição.

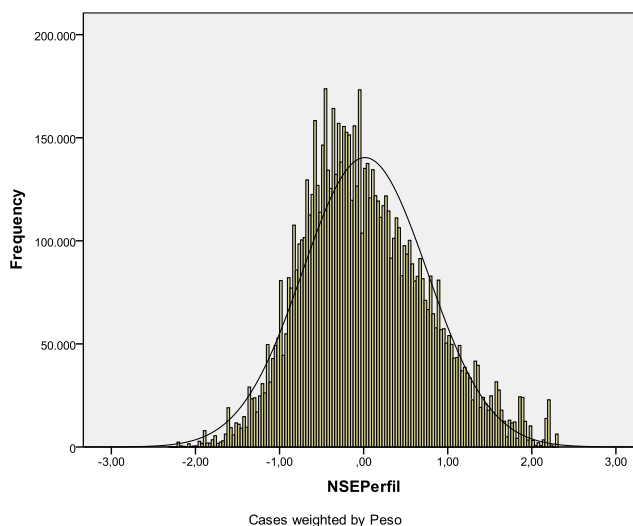


Gráfico 7: histograma do NSE – perfil de alunos ponderados

Para estimar o NSE por escola, os escores TRI calculados para cada perfil (o índice de NSE) foram incorporados aos bancos de dados nos quais a variável de identificação da escola (código do Inep) está presente. Isto ocorre nas três edições da Prova Brasil, que possui dois bancos de dados em cada edição (um por série – 5º e 9º ano) e nas três edições do Enem. Para isto, fez-se a compatibilização dos padrões de respostas aos itens nos perfis com as respostas de cada um dos alunos incluídos nesses bancos de dados. Em seguida, estes nove bancos de dados foram empilhados – isto é, transformados em um único banco de dados. O NSE de cada escola corresponde à média dos escores estimados pela TRI – ou o NSE – dos alunos por escola. Computou-se

ainda os desvios-padrão dessas médias – considerando-se a imprecisão das medidas dos alunos⁵ – e o número de alunos por escola.

Ao final deste processo, estimou-se o NSE para 76.006 escolas. Para evitar distorções nas médias, considerou-se o índice de NSE apenas das escolas que possuem informação de pelo menos 15 alunos. Assim, o banco de dados finais contém informações sobre 69.906 escolas, 85% delas em área urbana⁶.

A tabela 2 apresenta as estatísticas descritivas do índice de NSE das escolas e o gráfico 6, apresenta o histograma do NSE das escolas. Nota-se, que o NSE das escolas se concentra em valores menores do que o NSE dos perfis. Isto é esperado, pois somente se existisse uma escola em que todos os alunos tivessem o NSE igual ao valor máximo ou então ao valor mínimo haveria possibilidade de ocorrer coincidência entre os valores extremos das tabelas. Tal situação é praticamente improvável de acontecer.

Tabela 2: Estatísticas descritivas do índice de NSE das escolas

	N	Mínimo	Máximo	Média	Desvio-Padrão
NSE da escola	69906	-1,83	2,01	-,0,07	0,47

Fonte: elaboração própria

No gráfico 8, nota-se que há maior concentração de escolas com valores abaixo de zero. Isto também é esperado, haja vista que 91,5% das escolas representadas na base de dados são públicas – estaduais ou municipais – nas quais parte dos alunos vêm de famílias com baixo NSE⁷. Tal constatação

⁵ O desvio-padrão calculado para o NSE das escolas levou em consideração a imprecisão das médias dos alunos que foram sintetizadas para a produção desse índice. Isto porque o NSE estimado para cada aluno incluído na análise tem um grau de incerteza que varia em função do número de itens que ele respondeu.

⁶ A sinopse estatística da educação básica com base no Censo Escolar 2010 registra a existência, no Brasil, de 75.201 estabelecimentos de ensino de nível fundamental e 24.196 de ensino médio. Estes segmentos educacionais, registrados como números separados, podem, muitas vezes, pertencer ao mesmo estabelecimento de ensino.

⁷ A sinopse estatística da educação básica com base no Censo Escolar 2010 registra a existência, no Brasil, de 81,5% de escolas públicas no nível de educação básica, percentual que inclui do infantil ao ensino médio e todas as modalidades de ensino (regular, EJA, especial, profissional).

baseia-se no gráfico 9 que relaciona o nível socioeconômico das escolas associado às suas dependências administrativas. Observa-se que escolas municipais e estaduais têm NSE semelhantes e em níveis mais baixos do que escolas federais e privadas. As escolas privadas, como esperado agregam os alunos de maior nível socioeconômico.

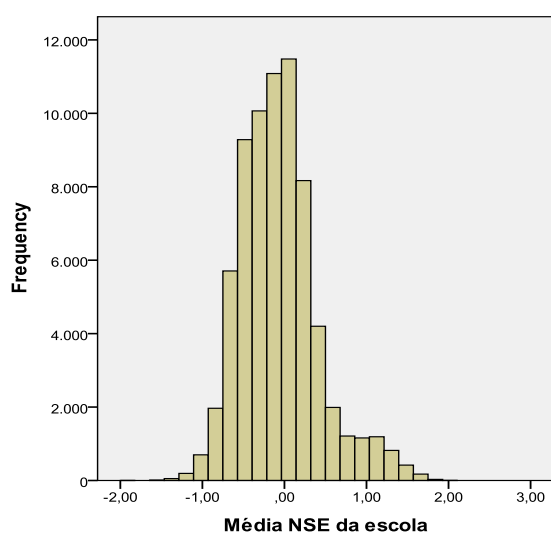


Gráfico 8: histograma do NSE escola

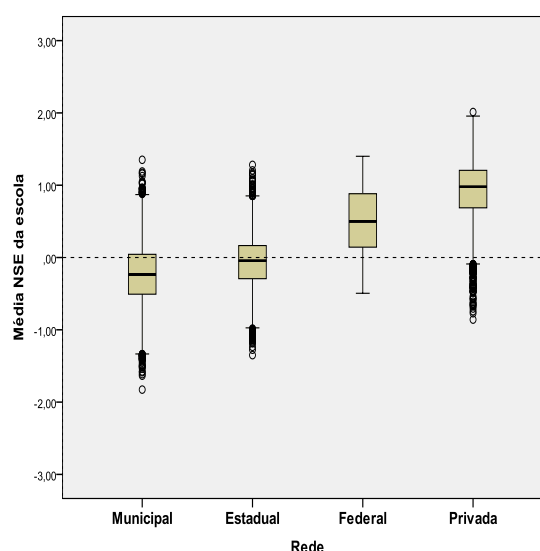


Gráfico 9: Box plot NSE por rede de ensino

Finalmente, para gerar o banco de dados para consulta do NSE de cada escola, fez-se a identificação dos dados das escolas (nome, município e unidade da federação). Os bancos de dados das avaliações possuem o código Inep das escolas, mas não possuem uma variável com nome da escola, com exceção da base de dados da Prova Brasil de 2005 e 2007 que informam os nomes de escolas públicas de ensino fundamental que participaram dos exames naqueles anos. Para complementar os nomes não disponíveis – isto é, nome de escolas que aparecem somente na Prova Brasil 2009 e nas edições do Enem – utilizou-se uma base de dados construída ao longo dos anos no âmbito de vários estudos educacionais conduzidos no Game/UFMG. Feitas as compatibilizações, restaram ainda quase 700 códigos sem identificação. Estes foram consultados no site do Inep.

4.3. Validação dos resultados

Para validar o índice de NSE obtido, calcularam-se correlações entre o valor obtido e outros indicadores. Primeiramente, testou-se a correlação com o NSE da pesquisa Geres: Geração escolar 2005⁸ – que foi calculado com base em informações mais precisas sobre o nível de renda da família e a ocupação e a escolaridade dos pais de alunos matriculados em uma amostra de escolas públicas e privadas. Há 227 escolas em comum nas duas bases de dados. A correlação do NSE estimado nas duas amostras é bem alta, com o valor de 0,96.

Outra medida de NSE disponível provém do Sistema de Avaliação do Rio Grande do Sul (SAERS), realizado pela Secretaria de Estado da Educação daquele estado. Nesse sistema, durante a matrícula para o ano letivo de 2009 os pais dos alunos preencheram um questionário informando sobre a sua ocupação, escolaridade e renda. Há 1.929 escolas em comum nas duas bases de dados (somente escolas públicas estaduais). A correlação entre os índices de NSE obtidos é também alta, atingindo o valor de 0,80.

Testou-se também, com dados obtidos em avaliações da rede estadual realizadas no estado de São Paulo, nos anos de 2008, 2009 e 2010. Neste caso são os pais dos alunos responderam ao questionário na época da avaliação estadual. Neste caso, há 5.139 escolas em comum. A correlação novamente atingiu o valor de muito alto, de 0,88.

Testou-se ainda a correlação entre o NSE dos alunos da base final, agregado por município com o rendimento mensal per capita por município obtido com base no Censo 2010. A base de dados do índice de NSE contém informações sobre todos os municípios brasileiros, ou seja, 5.565 municípios. O resultado da correlação entre os dois indicadores é muito alto, atingindo 0,91. Este resultado é particularmente importante, pois mostra que o NSE capta a realidade econômica do município da mesma forma que a medida da renda per capita

⁸ Para maiores informações sobre a Pesquisa Geres: Geração Escolar 2005, consultar Franco, Brooke, Alves (2008).

obtida por meio do Censo Demográfico. A diferença é que para isso baseia-se em dados coletados com os alunos das escolas.

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

5. Considerações finais

Este trabalho reuniu dados de avaliações educacionais realizadas durante os anos de 2001 a 2009 para produzir um índice de nível socioeconômico (NSE) das escolas de ensino fundamental e médio. O maior desafio do trabalho foi produzir um indicador consistente, haja vista que os dados coletados, por meio de questionários contextuais respondidos pelos alunos nessas avaliações, não foram planejados para esta finalidade. Inúmeras decisões metodológicas tiveram que ser tomadas devido a diferenças entre os questionários. O trabalho envolveu etapas bastante complexas desde a análise e seleção dos itens, a preparação dos bancos de dados, a estimação do NSE e a validação dos resultados. O banco de dados que acompanha este relatório com o resultado final tem um enorme potencial para orientar estudos e pesquisas educacionais. Entretanto, algumas ressalvas para a leitura dos resultados devem ser feitas.

Estão ausentes estabelecimentos de ensino que não se enquadram no perfil das escolas representadas nas bases de dados cujas informações foram agregadas para se produzir o NSE médio das escolas. Por exemplo, não há identificação das escolas privadas que atendem apenas ao ensino fundamental. Este segmento escolar contribuiu para a estimação do NSE porque ele está representado nas bases de dados do Saeb. Mas, como explicado na metodologia deste trabalho, não é possível identificar escolas no Saeb. O setor privado está identificado quando presente nas bases de dados do Enem, que contém informação sobre o estabelecimento onde o aluno concluiu o ensino médio. Eventualmente, estes podem atender também ao ensino fundamental. As escolas localizadas em áreas rurais não estão devidamente representadas.

As que estão presentes na base de dados tiveram seu NSE calculado com base em informações da última edição da Prova Brasil – quando elas passaram a fazer parte do exame, desde que tivessem no mínimo 20 alunos na série avaliada – ou então de escolas rurais de ensino médio (possivelmente técnico), que são muito raras. Estão ausentes, também, escolas cujo número de alunos que participaram das avaliações realizadas no período considerado, é inferior a 15 alunos, para não gerar distorções excessivas na estimativa média.

O NSE de cada escola não deve, de forma alguma, ser absolutizados. Isto porque há uma variação em torno das médias. Qualquer inclusão de novos dados ou de itens pode alterar posições entre escolas próximas na ordenação. Por esta razão, o banco de dados com as informações das escolas contém também a medida do erro padrão da média. Este indica, de forma clara, que as médias possuem uma variação e as hierarquias derivadas não são absolutas. Isso limita a utilidade dos *rankings* de escolas.

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

Um último aspecto a se destacar é que os valores do NSE médio por escola correspondem aos valores da escala original obtida com o ajuste do modelo da Teoria da Resposta ao Item (TRI) – os valores de *Theta*. Estes correspondem aos escores do NSE – tipicamente variam entre -3 a 3. Na apresentação dos resultados por escola, optou-se por não fazer nenhuma transformação na escala original. Isto poderá ser feito posteriormente transformando a escala original em em uma que contenha somente números positivos. Entretanto, não é óbvio que intervalo de valores deve estar contemplado na escala transformada, porque a escolha deve fazer sentido substantivo. Por exemplo, a mera transformação dos valores mínimos e máximos observados para um intervalo de zero a 100 poderia levar a uma interpretação de que há uma escola cujo valor do NSE seria zero,

algo sem muito sentido na realidade empírica. Por esta razão, esta decisão deverá ser tomada depois de uma análise e discussão dos resultados.

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

A inspeção qualitativa dos estabelecimentos de ensino situados nos extremos superior e inferior coincidiu com as percepções sobre a realidade educacional de várias cidades brasileiras. Conseguiu-se, assim, um bom resultado mesmo junto a públicos muito difíceis de se acessar de forma direta ou indireta, como é o caso das famílias de alto poder aquisitivo, cujos filhos estudam em escolas privadas mais elitizadas. Com todos os esforços feitos, pode-se afirmar que o índice de NSE produzido reflete uma ordenação que descreve de forma fidedigna a realidade socioeconômica das escolas incluídas na base de dados.

Em termos práticos, nas pesquisas e estudos educacionais, o índice de NSE pode ser usado como uma forma de criação de grupos homogêneos, pois os agrupamentos de escolas tendem a ser mais estáveis do que a posição específica de uma escola na hierarquia obtida. Dentre dos grupos homogêneos as comparações seriam bem mais legítimas. Hoje a ausência de um indicador como este induz a comparação entre escolas cujos alunados são muito diferentes, como se vê nos *rankings* de escolas cujos alunos participam do Enem ou das escolas públicas pelo Índice de Desenvolvimento da Educação Básica (Ideb).

Referências

AGUIAR, Neuma, FERNANDES, Danielle; NEVES, Jorge Alexandre. Mobilidade Social Feminina. In: AGUIAR, Neuma (Ed.). Desigualdades Sociais, redes de sociabilidade e participação política. Belo Horizonte: Editora UFMG, 2007. p.165-180.

ALVES, Maria Teresa Gonzaga; FRANCO, Creso. A pesquisa em eficácia escolar no Brasil: evidências sobre o efeito das escolas e fatores associados à eficácia escolar. In: N. BROOKE; J. F. SOARES (org.) Pesquisa em eficácia escolar: origem e trajetórias. Belo Horizonte: Editora UFMG, 2008a, p. 482-500.

ALVES, Maria Teresa Gonzaga; SOARES, José Francisco. Medidas de nível socioeconômico em pesquisas sociais: uma aplicação aos dados de uma pesquisa educacional. Opinião Pública, vol.15, nº. 1, pp.1-30, 2009.

ALVES DE BRITO, Murilo. M. Entre o Estudo e o Trabalho: Padrões e Tendências da Alocação de Tempo de Crianças e Adolescentes no Brasil de 1988 a 2005. Pensamento & Realidade, v. 26, p. 45-68, 2011.

BLAU, Peter M.; DUNCAN, Otis Dudley. The American Occupational Structure. New York: Wiley. 1967.

BOURDIEU, Pierre. Escritos de educação. Petrópolis. Vozes, 1998.

BROOKE, Nigel; SOARES, José Francisco. Comentários da seção 1. In:_____. Pesquisa em eficácia escolar: origem e trajetórias: Belo Horizonte: Editora UFMG, 2008, p. 14-22.a

DUNCAN, Otis Dudley. A Socioeconomic Index for All Occupations in REISS, Albert. J. (ed). Occupations and Social Status. New York: Free Press. 1963, p. 109-138.

ERICKSON, Robert; GOLDTHORPE, John H. The CASMIN Project and the American Dream. European Sociological Review, vol. 8, pp. 283-306. 1992.

FERRÃO-BARBOSA, Maria Eugénia; FERNANDES, Cristiano. A escola brasileira faz diferença? uma investigação dos efeitos da escola na proficiência em matemática dos alunos da 4a série. In: FRANCO, Creso (Ed.). Promoção, ciclos e avaliação educacional. Porto Alegre: ArtMed, 2001, p.155-172

FRANCO, Creso; FERNANDES, Cristiano; SOARES, José Francisco; BARBOSA, Maria Eugénia; ALVES, Maria Teresa G. O Referencial teórico na construção dos questionários contextuais do SAEB 2001. Estudos em Avaliação Educacional, v.28, jul-dez, p.39-74. 2003.

GANZEBOOM, Harry. B. G., DE GRAAF, Paul, TREIMAN, Donald. A Standard International Socio-Economic Index of Occupational Status. Social Science Research, v.21, p.1-56. 1992.

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

HOUSER, Robert M.; WARREN, John Robert. Socioeconomic Index of Occupational Status: A Review, Update, and Critique. In: RAFTERY, Adrian (ed.). Sociological Methodology, Cambridge: Blackwell. 1997, p. 177-298.

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

NOGUEIRA, Maria Alice. Capital Cultural (verbete). In.: ZANTEN, Ignès Van (coord.) Dicionário de Educação. Petrópolis: Ed. Vozes, 2011, p. 80-82.

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

PASTORE, José; SILVA, Nelson do Valle. Mobilidade Social no Brasil. São Paulo: Markron. 2000.

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

SANTOS, José Alcides F. Uma classificação socioeconômica para o Brasil. Revista Brasileira de Ciências Sociais, vol.20, no.58, p.27-45, 2005.

SCALON, Maria Celi. Mapeando Estratos: Critérios para Escolha de uma Classificação. Dados - Revista de Ciências Sociais, v.41, n.2, p. p.337-375, 1998.

SOARES, José Francisco. O efeito da escola no desempenho cognitivo de seus alunos. In: A. M. Souza (Ed.). Dimensões da avaliação educacional. Petrópolis: Editora Vozes, 2005, p.174-204

SOARES, José Francisco; ALVES, Maria Teresa Gonzaga. Desigualdades raciais no sistema brasileiro de educação básica. Educação e Pesquisa, v.29, n.1, jan./jun., p.147-165. 2003.

SOARES, José Francisco; ANDRADE, Renato Júdice. Nível socioeconômico, qualidade e equidade das escolas de Belo Horizonte. Ensaio: Avaliação e Políticas Públicas em Educação, v.14, n.50, p.107-126. 2006.

SOARES, José Francisco; CÉSAR, Cibele Comini; MAMBRINI, Juliana. Determinantes de desempenho dos alunos do ensino básico brasileiro: evidências do SAEB de 1997. In: Franco, Creso (Ed.). Promoção, ciclos e avaliação educacional. Porto Alegre: ArtMed, 2001b, p.121-153.