

**INDICADOR DE NÍVEL  
SOCIOECONÔMICO DOS  
INSCRITOS NO ENEM:  
CONCEPÇÃO, METODOLOGIA  
E RESULTADOS**

Gabriela Thamara de Freitas Barros  
Adriano Souza Senkevics  
Adolfo Samuel de Oliveira  
Adriano Ferreti Borgatto

2019  
SÉRIE DOCUMENTAL  
TEXTOS PARA  
DISCUSSÃO

**47**



REPÚBLICA FEDERATIVA DO BRASIL

MINISTÉRIO DA EDUCAÇÃO | **MEC**

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS  
EDUCACIONAIS ANÍSIO TEIXEIRA | **INEP**


DIRETORIA DE ESTUDOS EDUCACIONAIS | **DIRED**

# **TEXTO PARA DISCUSSÃO 47**

## **INDICADOR DE NÍVEL SOCIOECONÔMICO DOS INSCRITOS NO ENEM: CONCEPÇÃO, METODOLOGIA E RESULTADOS**

Gabriela Thamara de Freitas Barros  
Adriano Souza Senkevics  
Adolfo Samuel de Oliveira  
Adriano Ferreti Borgatto

Brasília-DF  
Inep/MEC  
2019

 Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep)  
É permitida a reprodução total ou parcial desta publicação, desde que citada a fonte.

#### DIRETORIA DE ESTUDOS EDUCACIONAIS (DIRED)

---

##### COORDENAÇÃO DE EDITORAÇÃO E PUBLICAÇÕES

Carla D' Lourdes do Nascimento – carla.nascimento@inep.gov.br

Valéria Maria Borges – valeria.borges@inep.gov.br

##### REVISÃO

*Português:*

Elenita Rodrigues

*Inglês:*

Walkíria de Moraes Teixeira da Silva

##### REVISÃO GRÁFICA

Elenita Rodrigues

##### NORMALIZAÇÃO E CATALOGAÇÃO NA FONTE

Clarice Rodrigues da Costa

##### PROJETO GRÁFICO

Raphael Freitas

##### CAPA

Marcos Hartwich

##### DIAGRAMAÇÃO E ARTE-FINAL

José Miguel dos Santos

#### EDITORIA | DISTRIBUIÇÃO

---

INEP/MEC – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA

Setor de Indústrias Gráficas - Quadra 04 - Lote 327, Térreo, Ala B – CEP 70.610-908 – Brasília-DF – Brasil

Fones: (61) 2022-3077, 2022-3078 – dired.publicacoes@inep.gov.br

TIRAGEM: 1.000 exemplares

Publicada online em dezembro de 2019

A exatidão das informações e os conceitos e opiniões emitidos  
são de exclusiva responsabilidade dos autores.

ESTA PUBLICAÇÃO NÃO PODE SER VENDIDA. DISTRIBUIÇÃO GRATUITA.

#### Dados Internacionais de Catalogação na Publicação (CIP)

---

Indicador de nível socioeconômico dos inscritos no Enem : concepção, metodologia e resultados /  
Gabriela Thamara de Freitas Barros ... [et al.]. – Brasília : Instituto Nacional de Estudos e Pesquisas  
Educacionais Anísio Teixeira, 2019.

68 p. : il. – (Série Documental. Textos para Discussão, ISSN 1414-0640 ; 47)

DOI: 10.24109/1414-0640.TD.2019.47

1. Indicadores socioeconômicos. 2. Exame Nacional do Ensino Médio. 3. Alunos: condições  
socioeconômicas. I. Barros, Gabriela Thamara de Freitas. II. Série..

CDU 371.212:316.3

---



# SUMÁRIO

INTRODUÇÃO.....	7
<b>1</b> REFERENCIAL TEÓRICO .....	<b>8</b>
1.1 O que é o nível socioeconômico familiar? .....	9
1.2 Nível socioeconômico e Teoria de Resposta ao Item.....	10
<b>2</b> PROCEDIMENTOS METODOLÓGICOS.....	<b>16</b>
2.1 Preparação das Bases de Dados .....	16
2.2 Extração da Amostra .....	20
2.3 Matriz de correlação policórica.....	21
2.4 Verificação do pressuposto da unidimensionalidade .....	25
2.5 Validação do ordenamento das categorias.....	29
2.6 Verificação do pressuposto da invariância.....	29
2.7 Verificação do pressuposto da independência local.....	33
2.8 Calibração dos itens: estimação dos parâmetros e do erro-padrão .....	34
2.9 Estimação do nível socioeconômico dos indivíduos .....	38
2.10 Segmentação e interpretação da escala.....	39

<b>3</b>	<b>RESULTADOS</b> .....	<b>41</b>
	CONSIDERAÇÕES FINAIS .....	<b>50</b>
	REFERÊNCIAS .....	<b>52</b>
	APÊNDICES .....	<b>59</b>
	Apêndice A – Curvas características dos itens, por questão – Enem 2011-2017 .....	<b>61</b>
	Apêndice B – Classificação dos bens de consumo e contratação de serviços .....	<b>64</b>
	Apêndice C – Distribuição das Alternativas de cada item do NSE pelos níveis segmentados da escala .....	<b>65</b>



# INDICADOR DE NÍVEL SOCIOECONÔMICO DOS INSCRITOS NO ENEM: CONCEPÇÃO, METODOLOGIA E RESULTADOS\*

Gabriela Thamara de Freitas Barros<sup>I</sup>

Adriano Souza Senkevics<sup>II</sup>

Adolfo Samuel de Oliveira<sup>III</sup>

Adriano Ferreti Borgatto<sup>IV</sup>

## RESUMO

---

Medidas de nível socioeconômico familiar são indicadores contextuais de suma importância para a investigação das desigualdades de acesso, trajetória e aprendizagem dos estudantes. Com o intuito de subsidiar esse debate, o objetivo deste texto é apresentar a concepção, metodologia e resultados de um indicador de nível socioeconômico dos inscritos do Exame Nacional do Ensino Médio (Enem), procurando detalhar o referencial teórico, o processo criativo e os procedimentos técnico-metodológicos para a geração de uma medida sintética da posição social dos indivíduos. Para tanto, são utilizadas as bases de dados dos inscritos

---

\* Este trabalho contempla alguns tópicos, revisados e adaptados, da dissertação de mestrado defendida no Programa de Pós-graduação em Métodos e Gestão em Avaliação da Universidade Federal de Santa Catarina, intitulada "Procedimentos para a construção de indicadores por meio da teoria de resposta ao item: a criação de uma medida de nível socioeconômico familiar" (BARROS, 2016).

<sup>I</sup> Pesquisadora da Diretoria de Estudos Educacionais (Dired/Inep), bacharel em Estatística pela Universidade de Brasília (UnB) e mestre em Métodos e Gestão da Avaliação pela Universidade Federal de Santa Catarina (UFSC). E-mail: gabriela.barros@inep.gov.br

<sup>II</sup> Pesquisador da Diretoria de Estudos Educacionais (Dired/Inep), especialista em Gestão de Políticas Públicas pela Universidade de Brasília (UnB), mestre e doutorando em Educação pela Universidade de São Paulo (USP). E-mail: adriano.senkevics@inep.gov.br

<sup>III</sup> Pesquisador da Diretoria de Estudos Educacionais (Dired/Inep), bacharel e licenciado em Ciências Sociais, mestre e doutor em Educação pela Universidade de São Paulo (USP). E-mail: adolfo.oliveira@inep.gov.br

<sup>IV</sup> Professor associado do Departamento de Informática e Estatística da Universidade Federal de Santa Catarina (INE/UFSC). E-mail: adriano.borgatto@ufsc.br

do Enem entre 2011 e 2017 para a construção de um indicador que sintetiza informações de renda familiar e escolaridade parental, por meio do modelo de respostas graduais da Teoria de Resposta ao Item (TRI). Em seguida, o texto explora um conjunto de validações do indicador, além de exemplificar potenciais aplicações e limitações. Sabendo que a construção de medidas numéricas em ciências sociais requer a explicitação dos esquemas teóricos e conceituais que informam a análise, o texto estará especialmente interessado na articulação entre um plano teórico-abstrato e outro operacional-empírico para o estudo da estratificação social e da desigualdade educacional.

Palavras-chave: desigualdade entre classes; estratificação social; indicadores socioeconômicos; Pierre Bourdieu; Teoria de Resposta ao Item. ■

## **ABSTRACT**

### *SOCIOECONOMIC STATUS INDEX OF ENEM APPLICANTS: CONCEPTION, METHODOLOGY AND RESULTS*

Measures of family socioeconomic level are exceedingly important contextual indicators for the investigation of the inequalities of access, trajectory and learning of students. To support this debate, this text presents the conception, methodology and results of one socioeconomic status index of applicants at the National High School Exam (acronymed in Portuguese as Enem), providing details on the theoretical framework, the creative process, and the technical procedures for generating a synthetic measure of individuals' social position. To do so, we used Enem's databases from 2011 to 2017 to construct an indicator that synthesizes information about family income and parental schooling, through the graded-response model from Item Response Theory (IRT). Subsequently, the text explores a set of validations of the indicator, as well as exemplifies potential applications and limitations. Knowing that the construction of numerical measures in social sciences requires the explanation of the theoretical and conceptual schemata that inform the analysis, the text will primarily entertain the articulation between a theoretical-abstract framework and an operational-empirical one for the study of social stratification and educational inequality.

*Keywords: Item Response Theory; Pierre Bourdieu; social inequality; social stratification; socioeconomic indicators.* ■



## INTRODUÇÃO

---

Nem a mobilidade social nem as fronteiras indistintas mudam o fato de que a pirâmide é uma dura, sólida realidade e de que as diferenças entre a classe situada nos níveis superiores da pirâmide e as classes situadas nos níveis inferiores são de fato muito grandes em termos de riqueza, renda, poder, responsabilidade, estilo e qualidade de vida e tudo o mais que compõe a textura da existência. Isso pode ser deplorado, ou louvado, ou declarado lamentável mas inevitável, ou visto de outro modo qualquer. O que não se pode ou não se deve fazer é ignorar a existência de tais divisões e a importância crucial para a vida da sociedade em que ocorrem. (Miliband, 1999, p. 483).

Mensurar as disparidades socioeconômicas é uma prerrogativa de sociedades desiguais. Uma vez que se reconhece que as desigualdades não se restringem à distribuição de bens e recursos econômicos entre os indivíduos, extrapolando para outras dimensões da vida, a exemplo do processo de escolarização, discussões sobre a trajetória, aprendizagem e experiências de estudantes nas escolas brasileiras demandam a utilização de medidas socioeconômicas para se melhor apreender as nuances da realidade. Não foi com outro intuito que tem sido gestado, desde o século passado, indicadores de nível socioeconômico aplicados à área educacional. Aqui, voltaremos nossa atenção a um indicador de nível socioeconômico familiar dos inscritos no Exame Nacional do Ensino Médio (Enem), índice de especial valor para a compreensão do perfil de quem conclui a escolarização básica no Brasil, bem como de quem se candidata à continuidade dos estudos no nível superior.

Em razão de seu caráter como um texto para discussão, apresentaremos em detalhes o referencial teórico e os procedimentos metodológicos para cálculo do indicador. Nosso intuito, neste documento, é antes explicar meticulosamente os fundamentos teóricos e metodológicos da produção do Indicador de Nível Socioeconômico (Inse) do que explorar em profundidade suas potenciais aplicações para as pesquisas educacionais. Esperamos registrar o acúmulo de um processo disparado no âmbito do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Brasil. Inep, 2014) e, assim, fazer jus aos esforços empregados para a constituição de um indicador de nível socioeconômico com validade, confiabilidade e relevância para o estudo das desigualdades sociais e educacionais. O que é o indicador de nível socioeconômico familiar? Quais são suas bases teóricas e conceituais? De que maneira os dados do questionário socioeconômico do Enem podem ser empregados para a geração de um índice sintético da posição social dos inscritos? Como se produz uma medida interpretável sociologicamente e comparável ao longo do tempo? Que usos podem ser extraídos desse indicador? Quais cautelas devem ser observadas nesse uso?

Para responder essas e outras questões, este texto está organizado em quatro seções, além desta introdução. Na próxima seção, exploraremos o referencial teórico do Inse, procurando contextualizar alguns embates teóricos em torno da definição

do construto e referenciar estudos anteriores que computaram indicadores de nível socioeconômico, com especial atenção para aqueles baseados na Teoria de Resposta ao Item (TRI). Em seguida, apresentaremos em detalhes os procedimentos metodológicos para o cálculo do Inse ora proposto, explicando as escolhas tomadas ao longo desse processo e suas consequências. Depois, apresentaremos e discutiremos os resultados do indicador a fim de ilustrar algumas de suas formas de representação e aplicações. Por fim, as considerações finais revisarão os principais pontos do texto e concluirão com sugestões de perspectivas para o aprofundamento do estudo.

## 1 REFERENCIAL TEÓRICO

---

Sabendo que a busca por critérios robustos, válidos e confiáveis para estratificar populações ou subpopulações em termos socioeconômicos não é recente, é preciso reconhecer que persistem até a atualidade disputas em torno dos marcos teóricos sobre os quais se deve ancorar uma medida socioeconômica familiar de potencial utilidade para as pesquisas em estratificação e desigualdade social. De acordo com a literatura na área, ainda no início da década passada se notava uma ausência de consenso sobre a mensuração e o significado conceitual de medidas socioeconômicas (Bollen; Glanville; Stecklov, 2001); publicações recentes nos permitem atualizar essa constatação e trazê-la especificamente para o caso brasileiro, a exemplo de Kamakura e Mazzon (2013).

Em contrapartida, é consenso entre a maioria dos sociólogos da atualidade que a classe social é a principal forma de estratificação social das nações modernas em todo o mundo (Giddens; Sutton, 2016). Ainda que o capitalismo – assim como outros modos de produção descritos na história das civilizações – comporte variados critérios de estratificação social, a exemplo das castas e dos estamentos, é preciso reconhecer que as classes sociais tais como as conhecemos se definem com o surgimento do modo de produção capitalista, do Estado moderno, da propriedade privada e da divisão social do trabalho (Hirano, 2002). Aprender dimensões da estratificação por classes sociais tem sido um dos desafios que acompanham as ciências sociais quantitativas desde o seu surgimento.

Embora existam inúmeras estratégias para se obter uma medida de nível socioeconômico familiar, a literatura nacional e internacional tradicionalmente privilegia a natureza tripartite do nível socioeconômico a partir da *educação parental*, da *renda familiar* e da *ocupação parental* como os três principais indicadores da posição das famílias na estratificação social (Buchmann, 2002; Sirin, 2005). No campo educacional, esforços nesse sentido remontam aos primeiros achados de que a escola se apresenta como uma instituição-chave para a reprodução das desigualdades sociais (Coleman et al., 1966; Bourdieu, 2015; Bourdieu; Passeron, 2014; Connell et al., 1982; Patto, 1999). Para entender como esse indicador opera no estudo dessas disparidades, é preciso compreender suas bases teóricas e seu construto.

## 1.1 O QUE É O NÍVEL SOCIOECONÔMICO FAMILIAR?

Desde 2014, o Inep tem envidado esforços na construção de indicadores contextuais visando à melhor compreensão do fenômeno educativo e de seus condicionantes. Em vista disso, foi desenvolvido o Inse, a partir de dados do Sistema de Avaliação da Educação Básica (Saeb) e do Exame Nacional do Ensino Médio (Enem), como uma medida sintética de características socioeconômicas familiares dos alunos das escolas brasileiras de ensino fundamental e médio (Brasil. Inep, 2014). Seu construto pode ser enunciado como a agregação de uma medida direta de renda familiar (mensurada em faixas de salário-mínimo), uma medida educacional (mensurada pela escolaridade do pai e da mãe) e medidas de patrimônio familiar, expressas na posse de bens de consumo e na contratação de serviços, adotadas como proxies ou medidas indiretas de renda. Importa ressaltar que o referencial teórico que fundamenta esse indicador se respalda na noção de espaço social do sociólogo francês Pierre Bourdieu (1996).

Segundo uma abordagem bourdieusiana, os dois principais tipos de capital que estruturam o espaço social são o *capital econômico* e o *capital cultural*. Dentro desse esquema interpretativo, a estrutura de classe pode ser compreendida como um espaço constituído por pelo menos dois eixos (Weininger, 2015): (1) o volume total de capital (econômico e cultural) possuído por seus titulares; e (2) a composição do capital possuído, que versa sobre a relativa preponderância de cada um deles. Outro elemento importante desse arcabouço interpretativo é a trajetória dos titulares de capital, que, ao ocupar várias posições desse espaço ao longo da vida, pode levar a variações no volume e na composição de seus capitais, bem como a movimentos de “conversão” e “reconversão” deles.

Em suma, Bourdieu (1996, 2007) se propõe a abordar a distribuição de capitais de forma multidimensional, procurando alocar indivíduos e grupos em posições específicas do espaço social. Ademais, os principais eixos estruturantes do espaço social – o volume e a composição – são vistos, de um ponto de vista metodológico e teórico, como dimensões contínuas, relativas e, em certa medida, hierárquicas. Na prática, isso significa dizer que não existem linhas inerentes de clivagem definindo limites estruturais entre as classes, como se perceberia em uma perspectiva puramente marxista ou weberiana, mas que dentro desse gradiente de posições objetivas a identificação de situações ou frações de classes obedeceria estritamente a uma conveniência heurística (Weininger, 2015). Daí decorre que a distribuição contínua de posições no espaço social, por meio da apreensão do volume e composição de diferentes tipos de capital, poderia ser representada por uma escala contínua, teórica, que sintetiza de maneira abstrata a posição social de um indivíduo ou de um coletivo. Esta é a escala sobre a qual se fundamenta o já mencionado indicador.

Evidentemente, a teoria bourdieusiana encontra, no indicador de nível socioeconômico, uma possível operacionalização de parte de seus componentes, mas que não se encerra em uma medida sintética quantitativa. Para Bourdieu, tão importante quanto os rendimentos familiares e as credenciais escolares são os conflitos de ordem simbólica fruto de disputas

classificatórias pela apropriação e valorização de bens materiais e culturais, inscritas em conceitos sofisticados desenvolvidos pelo sociólogo a exemplo de *capital cultural incorporado* (Bourdieu, 2015). Em razão de seu caráter subjetivo, essas dimensões são raramente transformadas em indicadores quantitativos, a despeito de sua relevância para compreender a “violência simbólica” nas hierarquias de diferenciação e distinção social (Bourdieu, 2014). Também resta subteorizado na sociologia quantitativa outro ponto sensível dos escritos de Bourdieu: a relação entre estrutura de classe e ação de classe, isto é, a teoria de ação dos indivíduos intermediada pelo conceito de *habitus* (Bourdieu, 1996). Com isso, pouco se aprofunda acerca do “papel dos sistemas simbólicos ou culturais na produção e reprodução das estruturas sociais” (Nogueira; Nogueira, 2016, p. 40). Cientes dessas limitações, reforçamos que, embora o indicador ora proposto tenha um pé na teoria bourdieusiana no tocante aos construtos econômico e cultural, também o fundamentamos em abordagens neoweberianas de longa tradição nas Ciências Sociais quantitativas.

Essa ponderação nos leva a concluir que o Inse está preocupado com a alocação dos estudantes e suas famílias em uma posição relativa na estratificação social, e não com a definição de classes propriamente dita. É importante ter em mente que estratos sociais não equivalem a classes sociais. De acordo com o sociólogo mexicano Rodolfo Stavenhagen (1973, p. 150), “não é a ocupação, nem o montante de rendimentos, nem o estilo de vida, que constitui o principal critério para a constituição de uma classe social, embora todos constituam critérios secundários”. Assim, ao trabalharmos com um indicador de nível socioeconômico, o que buscamos é uma medida sintética de uso descritivo para a interpretação da estratificação social, sem compromisso com um uso analítico para a compreensão da dinâmica histórica de formação das classes. É no escopo da caracterização dos estratos sociais que o indicador encontra sua maior utilidade e, por isso, ancora-se em medidas familiares da renda corrente e permanente<sup>1</sup> e da escolaridade parental – dois critérios reconhecidamente relevantes para a definição das oportunidades de vida dos sujeitos.

## 1.2 NÍVEL SOCIOECONÔMICO E TEORIA DE RESPOSTA AO ITEM

A Teoria de Resposta ao Item (TRI) é uma modelagem desenvolvida para a mensuração de traços latentes (logo, variáveis não diretamente observadas) e envolve um conjunto de técnicas matematicamente complexas que procuram representar a relação entre a probabilidade de um indivíduo apresentar determinada resposta a uma questão em função do seu traço latente e das características intrínsecas dessa questão (Andrade; Tavares; Valle, 2000). A partir da década de 1980, a TRI vem se tornando a técnica predominante no campo dos testes cognitivos (Pasquali, 2013). Mais recentemente, a utilização da TRI para cálculo de indicadores contextuais tem ganhado proeminência na pesquisa nacional, não se restringindo a uma técnica para cômputo

<sup>1</sup> Para uma diferenciação de renda corrente e permanente, ver Kamakura e Mazzon (2013).

da proficiência em testes padronizados e tampouco à área educacional (Soares, 2005); em comum, os estudos que lançam mão dessa modelagem estatística buscam mensurar variáveis latentes, tais como a satisfação no trabalho, o grau de ansiedade e depressão, a situação socioeconômica e, evidentemente, as já mencionadas habilidades cognitivas. Daí por que o emprego da TRI tem extrapolado os tradicionais usos psicométricos e tem compreendido, igualmente, aplicações sociométricas como as de interesse neste texto. Antes de avançar sobre esse tópico, é preciso dar um passo atrás e compreender quais vantagens da TRI justificariam o seu emprego. Em suma, iniciaremos tentando responder por que não adotamos métodos já consolidados para a geração de uma medida de nível socioeconômico familiar.

No Brasil, uma das medidas mais utilizadas para representar a condição socioeconômica é proposta pela Associação Brasileira de Empresas de Pesquisa (Abep): o Critério de Classificação Econômica Brasil (ou simplesmente Critério Brasil). Esse indicador objetiva estratificar, com base em um critério robusto e confiável, a população brasileira em termos socioeconômicos – aqui, referimo-nos em particular ao novo Critério Brasil, atualização proposta por Kamakura e Mazzon (2013). Para tanto, a medida é construída a partir de informações objetivas, precisas e de fácil coleta e operacionalização, de modo a prover um sistema de pontuação padronizada que sirva como um estimador do poder de consumo dos domicílios brasileiros. Seu construto é composto por 35 itens que interrogam sobre a quantidade possuída de 21 categorias de bens e serviços, os quais incluem um conjunto de itens sobre bens de conforto, contratação de empregados mensalistas, existência de rede geral de distribuição de água e de pavimento no trecho da rua do domicílio, e grau mais elevado de educação alcançado pelo chefe de família. É capaz de discriminar grandes grupos de acordo com sua capacidade de consumo de produtos e serviços acessíveis a uma parte significativa da população, e situa os domicílios em sete estratos (classes A, B1, B2, C1, C2, D ou E)<sup>2</sup>. Por esses motivos, o Critério Brasil tem ampla aplicação nas pesquisas de mercado e é adotado também por estudos que segmentam os indivíduos para investigar temas variados, tais como o comportamento eleitoral, avaliação e acesso a serviços públicos ou privados etc. Em razão da metodologia consolidada e da utilização difundida, o Critério Brasil inspira inúmeros indicadores Brasil afora; na área de Educação, não é diferente.

Quando o Inep se propôs à construção de um indicador de nível socioeconômico familiar, percebeu-se que não seria possível replicar a metodologia do Critério Brasil sem adaptações. Faziam-se necessárias propriedades ausentes no indicador da Abep, sendo as principais delas a possibilidade de construir uma escala interpretável sociologicamente, a facilidade de lidar com dados faltantes (*missing*) em um questionário autopreenchido, a equalização de uma única escala para populações distintas e a comparabilidade das medidas ao longo do tempo. Isso porque o fim precípuo do indicador contextual no Inep era gerar uma média

<sup>2</sup> Para segmentar esses estratos, os autores constroem um modelo de classes latentes a partir dos 35 itens socioeconômicos selecionados e apresentam seus resultados em “vintis” da distribuição. Em seguida, por meio de uma análise das diferenças e similaridades entre os vinte grupos, estes são agregados para apenas sete estratos de modo a facilitar sua interpretação (Kamakura; Mazzon, 2013).

escolar do nível socioeconômico do corpo discente, de modo a empregar essa medida para contextualizar o desempenho dos alunos nas avaliações em larga escala (Brasil. Inep, 2014). Naquele momento, foram utilizadas as respostas dos estudantes dos 5º e 9º anos do ensino fundamental e da 3ª série do ensino médio ao questionário do Saeb<sup>3</sup> 2011 e 2013, bem como os concluintes do ensino médio quando da inscrição no Enem dos mesmos anos. Trata-se, pois, de populações, questionários e aplicações distintas. Daí porque optou-se pela adoção da TRI como estratégia metodológica para viabilizar a construção desse indicador e permitir uma comparabilidade longitudinal, além da interpretação sociológica do indicador.

Ao cotejar a medida socioeconômica do Critério Brasil com um indicador calculado por meio da TRI, Soares, Souza e Pereira (2004) ressaltam que a TRI permite uma maior discriminação dos indivíduos por extrair mais informações do conjunto de dados, possibilitando, ainda, conhecimento mais aprofundado sobre cada item que compõe o critério. Por destrinchar as características sociométricas dos itens, a TRI também permite que instrumentos de coleta parcialmente comuns sejam conjugados na geração de uma única medida. Do ponto de vista da produção de estatísticas oficiais, esse passo representa um grande avanço metodológico, na medida em que cada escola teria apenas um indicador de nível socioeconômico que expressasse o perfil familiar de seu alunado, e não um indicador para cada série/ano escolar ou para cada avaliação ou exame de que ela participasse. Além do mais, esse indicador poderia ser comparado ao longo do tempo, levando à criação de uma série temporal que contemplasse variações na situação socioeconômica do corpo discente de cada estabelecimento de ensino. Outro ponto destacado pelos autores é a possibilidade de lidar com dados faltantes. Diferentemente do Critério Brasil, os questionários do Saeb e do Enem aplicados pelo Inep são autopreenchidos, isto é, preenchidos pelos próprios respondentes. No caso do Enem a partir de 2010, o preenchimento total do instrumento é obrigatório para efetivar a inscrição do candidato; no caso do Saeb, porém, a aplicação é feita em papel e, com frequência, envolve taxas de não resposta relativamente altas, geralmente sobre itens de construtos mais complexos para os respondentes (e.g. escolaridade dos pais ou responsáveis)<sup>4</sup>. Ao adotar uma modelagem por TRI, os dados ausentes passam a contabilizar como itens não apresentados ao respondente e não inviabilizam o cômputo do indicador<sup>5</sup>. Em síntese, existem diversas vantagens de se utilizar a TRI para calcular o indicador de nível socioeconômico, entre as quais destacamos: é uma técnica adequada para medir o que

<sup>3</sup> O Sistema de Avaliação da Educação Básica (Saeb) abrange um conjunto de avaliações em larga escala realizadas a cada dois anos a estudantes dos 5º e 9º anos do ensino fundamental e da 3ª série do ensino médio no Brasil. Os testes cognitivos são aplicados em conjunto com um questionário respondido pelo próprio estudante. Pelo menos desde 1995 esse questionário apresenta itens socioeconômicos, embora tenha havido variações no construto, enunciado e alternativa das questões ao longo do tempo.

<sup>4</sup> Para uma discussão metodológica dos instrumentos de pesquisa do Inep, com ênfase na coleta do item cor/raça, ver Senkevics, Machado e Oliveira (2016).

<sup>5</sup> Vale uma ressalva: quanto menor o número de itens respondidos, maior será o erro-padrão daquele escore produzido; a magnitude desse erro é estimada no modelo para cada indivíduo e fornecida na geração dos resultados. De todo modo, seguindo Barros (2016), desconsideraremos para o cálculo final do indicador os indivíduos que responderam menos de cinco itens do construto.

não é observado diretamente; permite um tratamento natural dos dados faltantes; permite estimar uma medida de nível socioeconômico mesmo com itens parcialmente comuns entre os instrumentos; trabalha inicialmente com variáveis fortemente correlacionadas; permite conhecer características dos itens utilizados e escolher apenas aqueles que contribuíram de forma significativa para a criação da medida; gera uma escala comparável ao longo do tempo; discrimina os indivíduos e suas famílias ao longo da escala com base no padrão de resposta dos itens apresentados; entre outras serventias fartamente documentadas na literatura (Hambleton; Swaminathan; Rogers, 1991; Soares; Souza; Pereira, 2004; Soares, 2005; Sartes; Souza-Formigoni, 2013).

No campo educacional brasileiro, os primeiros estudos que geraram medidas socioeconômicas por meio da TRI datam de 2003. No ano em questão, Soares (2003) calculou um indicador de nível socioeconômico como variável de controle para modelar a influência do professor e do ambiente escolar sobre a proficiência dos alunos no Sistema Mineiro de Avaliação Educacional (Simave), referente à aplicação de 2002; ainda nesse ano, José Francisco Soares e Juliana Mambrini (2003) compararam o efeito de um escore produzido pelo Critério Brasil e outro por um modelo específico da TRI sobre a nota dos candidatos ao vestibular da UFMG. Trata-se de dois estudos que geraram indicadores socioeconômicos pela TRI com aplicações bastante práticas para a pesquisa empírica. Desde então, outros trabalhos se debruçaram sobre aspectos teóricos e metodológicos de indicadores como estes, e têm convergido sobre a utilização do modelo de resposta gradual, unidimensional, de dois parâmetros (Soares, 2005; Soares; Andrade, 2006; Alves; Soares, 2009; Costa, 2010; Soares; Alves, 2013; Alves; Soares; Xavier, 2014; Brasil. Inep, 2014; Machado, 2014; Barros, 2016). Explicaremos os termos desse modelo abaixo.

Proposto há cinquenta anos por Fumiko Samejima (1969), o modelo de resposta gradual é apropriado para itens politômicos cujas respostas estão dispostas em uma escala ordinal; assim, esse modelo permite extrair informações sobre cada categoria – correspondente às alternativas de resposta ordenadas das questões do instrumento – e não apenas sobre respostas dicotomizadas em corretas/incorretas. Para tanto, o modelo é executado em duas etapas. Na primeira, de acordo com Andrade, Tavares e Valle (2000), supondo que os escores das categorias de um item  $i$  são arranjados em ordem do menor para o maior e denotados por  $k = 0, 1, \dots, m_i$  onde  $(m_i + 1)$  é o número de categorias do  $i$ -ésimo item, a probabilidade de um indivíduo  $j$  escolher uma determinada categoria ou outra superior do item  $i$  é dada pela seguinte extensão do modelo logístico de dois parâmetros:

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_{i,k})}}$$

com  $i = 1, 2, \dots, l, j = 1, 2, \dots, n$  e  $k = 0, 1, \dots, m_i$ , em que:

$\theta_j$  é o traço latente do  $j$ -ésimo indivíduo,

$P_{i,k}^+(\theta_j)$  é a probabilidade de um indivíduo  $j$  com traço latente  $\theta_j$  escolher a categoria  $k$  ou outra superior do item  $i$ ,

$a_i$  é o parâmetro de discriminação do item  $i$ ,

$b_{i,k}$  é o parâmetro de dificuldade da  $k$ -ésima categoria do item  $i$ .

E, para a segunda etapa, a probabilidade de um indivíduo  $j$  escolher uma categoria  $k$  do item  $i$  é:

$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j)$$

Ou seja, trata-se simplesmente da diferença entre a probabilidade de escolher uma categoria  $k$  ou outra superior do item  $i$  e a probabilidade de escolher uma categoria  $k + 1$  ou outra superior do mesmo item  $i$ . Isso permite modelar as probabilidades de cada categoria isoladamente e, com isso, adaptar o modelo logístico dicotômico de dois parâmetros para um modelo de respostas graduais adequado para itens politômicos, costumeiramente utilizado em questionários contextuais e socioeconômicos.

Diz-se que o modelo tem dois parâmetros porque, para cada item do construto, afere-se o comportamento do *parâmetro de discriminação* ( $a_i$ ) e, para cada categoria de resposta dos itens, do *parâmetro de dificuldade* ( $b_i$ ). O segundo parâmetro indica a magnitude do traço latente necessária para que seja provável a opção por determinada categoria de resposta – no caso, o quão elevado deve ser o nível socioeconômico familiar para que haja mais de 50% de chance de o indivíduo possuir um carro em sua residência, por exemplo. Já o primeiro parâmetro representa o grau de discriminação de um determinado item, isto é, sua capacidade de distinguir indivíduos de nível socioeconômico baixo daqueles de nível socioeconômico alto. Como se pode imaginar, a conjugação desses dois parâmetros para um amplo leque de itens do construto permite mapear larga extensão da escala de nível socioeconômico familiar e, assim, possibilita a estimação do nível socioeconômico de cada indivíduo e sua alocação em determinada posição da escala.

Por *unidimensional*, entende-se um modelo de TRI que assume a predominância de um único traço latente a explicar a variabilidade dos itens considerados para o cômputo do indicador (Andrade; Tavares; Valle, 2000); em nosso caso, esse traço latente, representado por  $\theta$  (*teta*), é, precisamente, o nível socioeconômico familiar. Mais à frente, quando da validação dessa hipótese, traremos informações adicionais sobre a unidimensionalidade do nosso modelo. Por ora, importa ressaltar que a unidimensionalidade é condição *sine qua non* para adoção do referido modelo, e o atendimento a essa hipótese está no cerne da validação do próprio construto.

## 2 PROCEDIMENTOS METODOLÓGICOS

---

Neste trabalho, optamos por adaptar a metodologia para cálculo do Inse adotada pelo Inep (2014). Ademais, partimos do estudo de Barros (2016) para delinear as principais escolhas metodológicas desta pesquisa e, assim, decidimos por desenvolver um indicador de nível socioeconômico familiar segundo o modelo unidimensional de resposta gradual



(Samejima, 1969) da TRI, tomando por base as respostas dadas pelos inscritos no questionário socioeconômico do Enem de 2011 a 2017.

Para viabilizar as análises almejadas, realizamos as seguintes etapas, nesta ordem: (1) compatibilização dos itens entre as distintas edições do exame; (2) preparação das bases de dados; (3) extração de uma amostra; (4) obtenção de correlações entre os itens; (5) verificação do pressuposto da unidimensionalidade; (6) validação do ordenamento das categorias; (7) verificação do pressuposto da invariância (função diferencial do item – DIF); (8) verificação do pressuposto da independência local; (9) estimação dos parâmetros (calibração) dos itens e de seus respectivos erros-padrão; (10) estimação do nível socioeconômico dos indivíduos; (11) segmentação e interpretação da escala em nível individual; e, finalmente, (12) análises dos resultados do indicador. Explicaremos detalhadamente cada um dos procedimentos listados acima ao longo desta e da próxima seção.

Para desenvolver este estudo, empregamos o pacote estatístico *SAS Enterprise Guide*, versão 4.2, visando compatibilizar os itens das edições do Enem e preparar as bases de dados (etapas 1 e 2, supracitadas). Para a maior parte dos procedimentos – da extração da amostra à interpretação da escala (etapas 3 a 11) –, utilizamos o programa *R*, versão 3.5.0 (R Core Team, 2018), em adição aos pacotes *sampling* 2.8 para a extração da amostra (Tillé; Matei, 2016), *plyr* 1.8.4 para a recodificação de variáveis (Wickham, 2011), *psych* 1.8.4 para correlações e análises fatoriais (Revelle, 2018), *mirt* 1.28 para os modelos de TRI (Chalmers, 2012) e, para a geração de relatórios, *rmarkdown* 1.9 (Allaire et al., 2018) e *knitr* 1.2 (Xie et al., 2018). Por fim, as análises finais (etapa 12) foram desenvolvidas tanto no *R*, versão 3.5.0, quanto no *PASW Statistics*, versão 18.

## 2.1 PREPARAÇÃO DAS BASES DE DADOS

Um dos primeiros desafios enfrentados foi compatibilizar todos os itens relativos ao construto nível socioeconômico entre as sete edições do Enem, entre 2011 e 2017<sup>6</sup>. Embora os questionários do Enem já tivessem sido concebidos para o cálculo de indicadores socioeconômicos, dificuldades como esta decorrem de alterações no enunciando e nas alternativas de respostas das questões, além de inclusões ou exclusões de perguntas, entre instrumentos de diferentes edições, os quais culminaram, no período considerado, em três configurações distintas de itens, as quais foram por nós enunciadas como A (2011), B (2012-2014) e C (2015-2016). O Quadro 1 sintetiza que as configurações A e B se assemelham quanto à forma de coletar a posse de bens de consumo, a contratação de serviços de

<sup>6</sup> Optamos por trabalhar com os dados do Enem a partir de 2011 porque, em 2010, o questionário apresentou itens atípicos que inviabilizam o cálculo do Inse pela metodologia aqui discutida (Alves; Soares; Xavier, 2014), e, em 2009, houve um percentual excepcionalmente elevado de não respostas em um questionário ainda aplicado em papel (Senkevics; Machado; Oliveira, 2016).

empregada doméstica<sup>7</sup> e a escolaridade dos pais, mas diferem quanto à agregação das faixas de renda familiar (em salários mínimos). Já as configurações B e C se assemelham quanto à renda familiar, mas diferem no tocante à posse de bens, à contratação de serviços e à escolaridade dos pais. A configuração C, por sua vez, é a única que apresenta o item de ocupação dos pais.

**QUADRO 1**  
**CONFIGURAÇÃO DOS BLOCOS DE ITENS ORIGINAIS DO INSE NO ENEM 2011-2017**

Ano	Blocos de Itens do NSE				
	Renda Familiar	Bens de Consumo e Serviços	Escolaridade dos Pais	Ocupação dos Pais	
A	2011	A	AB	AB	-
	2012	BC	AB	AB	-
B	2013	BC	AB	AB	-
	2014	BC	AB	AB	-
C	2015	BC	C	C	C
	2016	BC	C	C	C
	2017	BC	C	C	C

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Em um primeiro momento, agregamos o mínimo de categorias necessárias para assegurar a comparabilidade dos itens entre as edições do Enem. Isso foi necessário porque a maioria dos itens relativos à posse de bens de consumo dispunham, entre 2011 e 2014, de quatro categorias de resposta que variavam da “Não tenho” a “3 ou mais”, ao passo que, de 2015 em diante, apresentavam cinco categorias que cobriam de “Não” a “Sim, quatro ou mais”. Como veremos adiante, outras agregações de categorias foram necessárias por força das análises estatísticas posteriormente realizadas. No entanto, para garantir a compatibilização dos itens nesta primeira etapa, assegurando a maior fidedignidade possível aos instrumentos originais e evitando perdas de informação, agregamos apenas as categorias necessárias para permitir comparabilidade entre as edições do Enem.

Há duas ressalvas a serem feitas. Primeiramente, dada a importância documentada na literatura sobre os itens de renda familiar e escolaridade dos pais, não efetuamos nenhuma alteração nos itens originais; para estes, mantemos versões dos itens em seus distintos formatos, tendo em vista que eles foram apresentados a conjuntos diferentes de inscritos. Em segundo lugar, com relação aos itens de posse de bens, compatibilizamos as diferentes versões agregando as categorias necessárias, exceto quando os itens apresentavam alterações substanciais no enunciado (e.g. carro/automóvel, computador/microcomputador), as

<sup>7</sup> Referir-nos-emos a essas profissionais pelo gênero feminino porque, em 2015, cerca de 92% dessas trabalhadoras pertenciam a esse sexo. Vide o *Retrato das desigualdades de gênero e raça*, disponível em: <<http://www.ipea.gov.br/retrato/>>. Acesso em: 2 out. 2018.

quais poderiam ensejar vieses na resposta dos inscritos – nesses casos, mantivemos versões distintas do mesmo item a fim de testá-las quanto ao seu funcionamento diferencial (DIF) entre as edições do Enem. O Quadro 2 lista os itens considerados para as próximas etapas.

## QUADRO 2

### LISTA DE ITENS DO INSE COMPATIBILIZADOS ENTRE AS EDIÇÕES DO ENEM 2011-2017

(continua)

#	Item	Objeto do Item	Nº de Categorias	Edições
1	Q01_ABC	TV em cores	4	2011-17
2	Q02_ABC	TV por assinatura	2	2011-17
3	Q03_AB	Rádio	4	2011-14
4	Q04_1_ABC	Videocassete/DVD ou somente Aparelho de DVD	2	2011-17
5	Q04_2_AB	Videocassete/DVD	3	2011-14
6	Q04_3_C	Aparelho de DVD	2	2015-17
7	Q05_C	Forno micro-ondas	5	2015-17
8	Q06_C	Máquina de lavar louça	5	2015-17
9	Q07_ABC	Geladeira	4	2011-17
10	Q08_ABC	Freezer	4	2011-17
11	Q09_1_ABC	Máquina de lavar roupa [mencionando ou não o tanquinho]	4	2011-17
12	Q09_2_AB	Máquina de lavar roupa	4	2011-14
13	Q09_3_C	Máquina de lavar roupa (o tanquinho NÃO deve ser considerado)	5	2015-17
14	Q10_C	Máquina de secar roupa	5	2015-17
15	Q11_1_ABC	Carro / automóvel	4	2011-17
16	Q11_2_AB	Automóvel	4	2011-14
17	Q11_3_C	Carro	5	2015-17
18	Q12_C	Motocicleta	5	2015-17
19	Q13_1_ABC	Computador / Microcomputador	4	2011-17
20	Q13_2_AB	Microcomputador	4	2011-14
21	Q13_3_C	Computador	5	2015-17
22	Q14_ABC	Acesso a internet	2	2011-17
23	Q15_C	Quartos para dormir	5	2015-17
24	Q16_ABC	Banheiro	4	2011-17
25	Q17_1_A	Renda familiar [em faixas de salário mínimo]	11	2011
26	Q17_2_BC	Renda familiar [em faixas de salário mínimo]	17	2012-17
27	Q18_ABC	Telefone fixo	2	2011-17
28	Q19_ABC	Telefone celular	4	2011-17

## QUADRO 2

### LISTA DE ITENS DO INSE COMPATIBILIZADOS ENTRE AS EDIÇÕES DO ENEM 2011-2017

(conclusão)

#	Item	Objeto do Item	Nº de Categorias	Edições
29	Q20_ABC	Aspirador de pó	2	2011-17
30	Q21_1_ABC	Contratação de empregada doméstica/mensalista	2	2011-17
31	Q21_2_AB	Contratação de empregada mensalista [quantidade]	4	2011-14
32	Q21_3_C	Contratação de empregada doméstica [frequência]	4	2015-17
33	Q22_1_ABC	Escolaridade do pai	6	2011-17
34	Q22_2_AB	Escolaridade do pai	8	2011-14
35	Q22_3_C	Escolaridade do pai	7	2015-17
36	Q23_1_ABC	Escolaridade da mãe	6	2011-17
37	Q23_2_AB	Escolaridade da mãe	8	2011-14
38	Q23_3_C	Escolaridade da mãe	7	2015-17
39	Q24_C	Ocupação do pai	5	2015-17
40	Q25_C	Ocupação da mãe	5	2015-17

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Ao término desses procedimentos, constituímos uma base de dados com todas as inscrições no Enem entre 2011 e 2017, totalizando mais de 50 milhões de inscritos, distribuídos pelas sete edições conforme a Tabela 1. É preciso lembrar que, por ora, não nos preocupamos com os indivíduos que se inscreveram para mais de uma edição do Enem, uma vez que a estimação do nível socioeconômico será executada separadamente para cada ano. Ademais, ressaltamos que os quantitativos abaixo podem não corresponder exatamente ao dos microdados porque excluimos as inscrições dos participantes que não responderam o questionário.

## TABELA 1

### QUANTITATIVO DE INSCRIÇÕES NO ENEM 2011-2017

Ano	Quantitativo
2011	5.366.948
2012	5.791.065
2013	7.173.563
2014	8.722.247
2015	7.745.759
2016	8.627.367
2017	6.731.341
Total	50.158.290

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Além dos itens do construto NSE, incluímos na base de dados a média aritmética simples das quatro áreas de conhecimento avaliadas no Enem, um indicador de participação nas quatro provas e as chaves de ligação necessárias para cruzamentos com os microdados (máscara de identificação e ano). Em seguida, as sete bases foram exportadas, em formato “.csv”, para posterior abertura em *R*.

## 2.2 EXTRAÇÃO DA AMOSTRA

Em virtude do tamanho das bases de dados descritas na Tabela 1, extraímos uma amostra com o intuito de reduzir o quantitativo de inscritos e agilizar os procedimentos analíticos até a etapa de calibração dos itens. Com relação ao sorteio da amostra, uma de nossas preocupações foi garantir que houvesse um número mínimo de respostas para cada categoria dos itens do construto NSE. Estudos anteriores que se debruçaram sobre a influência do tamanho da amostra na estimação de traços latentes a partir do modelo de respostas graduais têm sugerido um tamanho mínimo de 500 a 1.000 respondentes, a depender do comprimento do instrumento e do número de dimensões captadas, para assegurar a estabilidade da calibração dos itens e das estimativas dos escores (Reise; Yu, 1990; De Ayala, 1994; Jiang; Wang; Weiss, 2016). Em nosso caso, o patamar mínimo de 1.000 respondentes por item é mais do que suficiente para asseverar a qualidade da amostra. Sabendo que uma das vantagens da TRI é a independência dos resultados em relação à amostra utilizada para a estimação dos parâmetros dos itens, buscamos obter uma amostra grande o suficiente para representar as variações existentes do traço latente que procuramos mensurar (Couto; Primi, 2011), ou seja, que cobrisse toda a escala de nível socioeconômico dos inscritos no Enem. Logo, julgamos adequada a extração de uma amostra probabilística, por meio de um plano de amostragem aleatória estratificada.

Para tanto, após alguns testes, chegamos a uma amostra de 507.500 inscritos (aproximadamente 1% do quantitativo total de inscritos), composta por 72.500 inscritos de cada edição entre 2011 e 2017. Visando tornar a amostra representativa de toda a escala do nível socioeconômico, dividimos cada edição do Enem em 11 estratos: dez grupos delimitados pelos decis de desempenho (sintetizado na média aritmética simples das notas nas quatro áreas de conhecimento) e um grupo composto pelos estudantes que não tiveram desempenho calculado (por motivos de abstenção ou eliminação). Essa escolha se embasa na forte associação existente entre nível socioeconômico e desempenho no teste cognitivo em questão, já evidenciada pela literatura (Travitzki, 2013; Silveira; Barbosa; Silva, 2015). Dos 11 estratos, foram amostrados aleatoriamente 7.500 inscritos do grupo sem desempenho, dos três menores e dos três maiores décimos; quanto aos quatro décimos intermediários, foram amostrados apenas 5.000 inscritos de cada (Tabela 2); essa diferença se justifica pela necessidade de aumentar a precisão das estimativas nas porções inicial e final da escala, geralmente sujeitas a maiores erros de estimação.

TABELA 2

## ESTRATOS E TAMANHO DA AMOSTRA ORIGINAL – ENEM 2011-2017

Estratos	Tamanho da amostra	
	por edição	geral
Sem desempenho	7.500	52.500
1° decil	7.500	52.500
2° decil	7.500	52.500
3° decil	7.500	52.500
4° decil	5.000	35.000
5° decil	5.000	35.000
6° decil	5.000	35.000
7° decil	5.000	35.000
8° decil	7.500	52.500
9° decil	7.500	52.500
10° decil	7.500	52.500
Total	72.500	507.500

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Antes de concluir essa etapa, garantimos que a amostra tivesse um número mínimo de 500 respostas para cada categoria dos itens<sup>8</sup>. Para assegurar a representatividade de categorias de itens de baixa frequência, adicionamos 1.214 casos à amostra original. Após esses procedimentos, concluímos a preparação de uma amostra final com 508.714 indivíduos representando distintas porções de desempenho das sete edições do Enem. Essa amostra foi empregada para todas as etapas seguintes, até o momento de estimação do nível socioeconômico para a população.

### 2.3 MATRIZ DE CORRELAÇÃO POLICÓRICA

Com a amostra final em mãos, partimos para a elaboração da matriz de correlação policórica – ideal para variáveis categóricas ordinais politômicas<sup>9</sup>. Essa análise permitiu verificar as correlações entre pares de itens, assim como se o conjunto do construto

<sup>8</sup> Embora a literatura não trate especificamente da representatividade por categoria dos itens, julgamos que, dado o tamanho da amostra e da importância de que o indicador fosse comparável ao longo do tempo, seria pertinente que todas as alternativas de resposta estivessem bem representadas para permitir ulteriores estudos sobre a qualidade dos itens. Para alcançar o patamar mínimo de 500 respostas, agrupamos categorias de baixíssima frequência de determinados itens.

<sup>9</sup> Para uma revisão das correlações com variáveis ordinais e não paramétricas, ver Bistaffa (2010).

apresentava um grau aceitável de coesão interna, já que se espera que itens associados ao mesmo traço latente sejam correlacionados. Considerando que há itens parcialmente comuns entre os anos, executamos cinco análises de correlações distintas: itens comuns a todas as edições de 2011 a 2017 (configuração ABC), itens da edição de 2011 (A), itens comuns às edições de 2012 a 2014 (B), itens comuns às edições de 2015 a 2017 (C), e todos os itens conjuntamente, independentemente de serem comuns ou não às edições (A, B e C). Por motivo de espaço, e de similitude nos resultados gerais, apresentaremos apenas duas matrizes de correlação policórica.

Na Tabela 3, se observa a matriz relativa aos itens comuns às sete edições do Enem (i.e., os itens apresentados com o sufixo “\_ABC” no Quadro 2) e, na Tabela 4, a matriz referente aos itens comuns às edições de 2015 a 2017 do Enem (i.e., os itens que continham a letra “C” no sufixo), em razão desta conter a maior quantidade de itens e também ter apresentado um item com comportamento anômalo, discutido adiante. Como sugerido por Pasquali (2012, p. 40), adotamos como referência o valor de  $\pm 0,30$  para estimar se há uma correlação de tamanho suficiente para validar o relacionamento entre variáveis e, assim, checar o grau de fatorabilidade da matriz para as análises posteriores. Por ora, o importante não é o valor exato das correlações entre os itens, e sim o padrão geral da análise; segundo Hair et al. (2005), se a inspeção visual da matriz não apresentar um grande quantitativo de correlações iguais ou superiores a 0,30, a análise fatorial, como uma técnica para evidenciar a existência de um traço latente, poderá ser inapropriada.

**TABELA 3**  
**MATRIZ DE CORRELAÇÃO POLICÓRICA DOS ITENS COMUNS AO ENEM 2011-2017**

	Q01	Q02	Q04_1	Q07	Q08	Q09_1	Q11_1	Q13_1	Q14	Q16	Q18	Q19	Q20	Q21_1	Q22_1	Q23_1
Q01	1,00															
Q02	0,58	1,00														
Q04_1	0,46	0,24	1,00													
Q07	0,49	0,39	0,25	1,00												
Q08	0,46	0,48	0,22	0,37	1,00											
Q09_1	0,49	0,51	0,27	0,40	0,50	1,00										
Q11_1	0,52	0,52	0,25	0,42	0,51	0,55	1,00									
Q13_1	0,55	0,56	0,27	0,41	0,48	0,52	0,59	1,00								
Q14	0,48	0,61	0,20	0,36	0,45	0,49	0,55	0,82	1,00							
Q16	0,59	0,52	0,26	0,50	0,47	0,47	0,61	0,56	0,51	1,00						
Q18	0,46	0,54	0,24	0,29	0,40	0,47	0,47	0,54	0,61	0,45	1,00					
Q19	0,50	0,40	0,21	0,40	0,35	0,35	0,43	0,47	0,43	0,47	0,19	1,00				
Q20	0,54	0,58	0,31	0,41	0,54	0,60	0,63	0,60	0,59	0,54	0,54	0,37	1,00			
Q21_1	0,46	0,55	0,28	0,39	0,47	0,39	0,56	0,49	0,41	0,57	0,42	0,27	0,49	1,00		
Q22_1	0,39	0,44	0,19	0,26	0,33	0,35	0,42	0,44	0,40	0,45	0,39	0,28	0,43	0,45	1,00	
Q23_1	0,33	0,38	0,14	0,23	0,29	0,30	0,40	0,41	0,37	0,42	0,33	0,27	0,36	0,43	0,66	1,00

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Nota: Estão destacados em laranja os coeficientes de correlação iguais ou superiores a 0,30.

De acordo com a Tabela 3, se observa que todos os itens, com exceção do Q04\_1 (videocassete/aparelho de DVD), apresentaram coeficientes de correlação iguais ou superiores a 0,30 com a maioria de seus pares. Essas correlações chegaram ao valor máximo de 0,82 entre os itens Q13\_1 (computador/microcomputador) e Q14 (acesso a internet). Com relação ao item que foge à regra, optamos por mantê-lo em nossas próximas análises, a fim de verificar se as correlações apresentadas, ainda que menores que 0,30, seriam suficientes para compor um fator com os demais.

Já na Tabela 4, vemos a matriz de correlação policórica dos itens comuns ao Enem 2015, 2016 e 2017 (configuração C), edições que, além de mais recentes, contêm a maior quantidade de itens do construto NSE. Há numerosos pares de itens com correlações superiores a 0,30, com exceção de apenas três: Q04\_1 (videocassete/aparelho de DVD), Q04\_3 (aparelho de DVD) e Q12 (motocicleta). É notório que o último apresentou correlações negativas com dez itens do questionário, além de ter apresentado correlações positivas próximas a zero com os outros itens. Acreditamos que esse comportamento anômalo decorre de distintas associações existentes entre possuir ou não uma motocicleta e o nível socioeconômico familiar do indivíduo – basta pensarmos que a motocicleta pode ser, ao mesmo tempo, um bem de luxo adquirido para fins de passeio por parte das camadas sociais abastadas, ou um instrumento de trabalho, obtido como um substituto dos eventualmente inacessíveis automóveis, por parte das camadas populares. Em resumo, o comportamento desse item sugere que, entre os respondentes do questionário do Enem, não possuir uma motocicleta significa ter um nível socioeconômico familiar maior do que possuir apenas uma, fenômeno que desvirtua o ordenamento previsto. Esse resultado indica que tal item, ao menos na configuração atual, poderá ser considerado inadequado para as análises posteriores, motivo pelo qual ele será eliminado após esta etapa.



TABELA 4

## MATRIZ DE CORRELAÇÃO POLICÓRICA DOS ITENS COMUNS AO ENEM 2015-2017

	Q01	Q02	Q04_1	Q04_3	Q05	Q06	Q07	Q08	Q09_1	Q09_3	Q10	Q11_1	Q11_3	Q12	Q13_1	Q13_3	Q14	Q15	Q16	Q17_2	Q18	Q19	Q20	Q21_1	Q21_3	Q22_1	Q22_3	Q23_1	Q23_3	Q24	Q25					
Q01	1,00																																			
Q02	0,60	1,00																																		
Q04_1	0,42	0,25	1,00																																	
Q04_3	0,42	0,25	1,00	1,00																																
Q05	0,53	0,53	0,28	0,28	1,00																															
Q06	0,40	0,46	0,24	0,24	0,43	1,00																														
Q07	0,54	0,41	0,30	0,30	0,45	0,38	1,00																													
Q08	0,48	0,47	0,24	0,24	0,49	0,41	0,41	1,00																												
Q09_1	0,50	0,53	0,22	0,22	0,60	0,44	0,45	0,51	1,00																											
Q09_3	0,51	0,56	0,22	0,22	0,62	0,47	0,42	0,53	1,00	1,00																										
Q10	0,35	0,36	0,21	0,21	0,39	0,48	0,32	0,49	0,59	0,64	1,00																									
Q11_1	0,53	0,53	0,22	0,22	0,56	0,48	0,46	0,50	0,58	0,60	0,39	1,00																								
Q11_3	0,53	0,53	0,22	0,22	0,56	0,48	0,46	0,50	0,58	0,60	0,39	1,00	1,00																							
Q12	0,04	-0,04	0,10	0,10	0,03	0,00	0,12	0,04	0,05	0,03	0,10	0,07	0,07	1,00																						
Q13_1	0,53	0,54	0,26	0,26	0,55	0,45	0,42	0,47	0,53	0,55	0,38	0,59	0,59	0,06	1,00																					
Q13_3	0,53	0,54	0,26	0,26	0,55	0,45	0,42	0,47	0,53	0,55	0,38	0,59	0,59	0,06	1,00	1,00																				
Q14	0,48	0,60	0,19	0,19	0,59	0,35	0,38	0,46	0,54	0,56	0,37	0,58	0,58	0,06	0,81	0,81	1,00																			
Q15	0,48	0,34	0,24	0,24	0,33	0,31	0,45	0,35	0,36	0,35	0,28	0,50	0,50	0,18	0,41	0,41	0,36	1,00																		
Q16	0,60	0,54	0,26	0,26	0,52	0,49	0,54	0,48	0,52	0,54	0,35	0,65	0,65	0,05	0,58	0,58	0,54	0,65	1,00																	
Q17_2	0,54	0,59	0,22	0,22	0,54	0,46	0,42	0,47	0,55	0,57	0,38	0,68	0,68	0,06	0,62	0,62	0,58	0,43	0,62	1,00																
Q18	0,46	0,55	0,18	0,18	0,50	0,38	0,35	0,41	0,50	0,53	0,28	0,50	0,50	-0,12	0,56	0,56	0,66	0,30	0,48	0,53	1,00															
Q19	0,55	0,44	0,32	0,32	0,43	0,22	0,42	0,40	0,44	0,44	0,32	0,49	0,49	0,15	0,51	0,51	0,49	0,49	0,50	0,51	0,32	1,00														
Q20	0,56	0,56	0,33	0,33	0,62	0,53	0,43	0,53	0,63	0,67	0,43	0,65	0,65	-0,01	0,60	0,60	0,59	0,40	0,57	0,61	0,54	0,44	1,00													
Q21_1	0,42	0,47	0,22	0,22	0,37	0,50	0,34	0,37	0,34	0,35	0,32	0,49	0,49	-0,04	0,42	0,42	0,32	0,32	0,56	0,50	0,36	0,25	0,43	1,00												
Q21_3	0,51	0,53	0,27	0,27	0,42	0,55	0,44	0,45	0,39	0,40	0,35	0,57	0,57	-0,03	0,48	0,48	0,38	0,43	0,65	0,57	0,41	0,32	0,47	0,99	1,00											
Q22_1	0,38	0,45	0,14	0,14	0,38	0,39	0,29	0,35	0,38	0,40	0,26	0,45	0,45	-0,06	0,45	0,45	0,41	0,28	0,48	0,50	0,39	0,33	0,45	0,42	0,47	1,00										
Q22_3	0,37	0,44	0,14	0,14	0,37	0,38	0,29	0,34	0,37	0,39	0,25	0,44	0,44	-0,05	0,44	0,44	0,40	0,28	0,47	0,49	0,38	0,33	0,44	0,41	0,46	1,00	1,00									
Q23_1	0,33	0,38	0,10	0,10	0,32	0,34	0,26	0,30	0,32	0,34	0,22	0,41	0,41	0,00	0,41	0,41	0,38	0,27	0,45	0,46	0,33	0,31	0,37	0,39	0,45	0,65	1,00									
Q23_3	0,33	0,37	0,10	0,10	0,31	0,34	0,26	0,30	0,32	0,33	0,22	0,41	0,41	0,00	0,41	0,41	0,37	0,27	0,44	0,45	0,32	0,30	0,37	0,39	0,44	0,65	1,00	1,00								
Q24	0,45	0,50	0,17	0,17	0,45	0,43	0,34	0,38	0,44	0,47	0,29	0,51	0,51	-0,10	0,51	0,51	0,48	0,29	0,54	0,58	0,46	0,39	0,50	0,47	0,54	0,67	0,67	1,00								
Q25	0,42	0,47	0,15	0,15	0,43	0,42	0,32	0,36	0,42	0,44	0,28	0,49	0,49	-0,07	0,49	0,49	0,46	0,28	0,52	0,57	0,43	0,37	0,46	0,47	0,55	0,57	0,57	0,70	0,69	1,00						

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Nota: Estão destacados em laranja claro os coeficientes de correlação iguais ou superiores a 0,30 e em laranja escuro os coeficientes negativos.

## 2.4 VERIFICAÇÃO DO PRESSUPOSTO DA UNIDIMENSIONALIDADE

Uma vez atestada, por meio das correlações policóricas, a fatorabilidade da matriz dos itens do NSE, prosseguimos com análises fatoriais a fim de verificar os pressupostos da TRI, a começar pelo *pressuposto da unidimensionalidade*. Em linhas gerais, a análise fatorial envolve um conjunto de técnicas estatísticas que visam expressar um grande número de variáveis em um número mínimo de fatores. De acordo com Pasquali (2012), essas técnicas se baseiam na premissa de que uma série de variáveis observadas, passíveis de mensuração – em nosso caso, os itens dos questionários do Enem – pode ser explicada por um número menor de variáveis não observáveis, chamadas de variáveis-fonte ou *fatores*. Esses fatores podem ser de dois tipos: (1) o primeiro são as variáveis-fonte comuns, que representam o fato de que as variáveis observadas se relacionam entre si, isto é, são responsáveis pelas intercorrelações (ou covariância) entre essas variáveis; (2) e o segundo tipo de fator são as variáveis-fonte exclusivas, ou fatores únicos, de cada variável observável, que representam aquilo que a variável tem de específico, ou seja, que independe das correlações com outras variáveis – essas fontes de variação podem incluir variáveis não contempladas pela análise, como também os erros de medida. Nesse sentido, a suposição é de que, se as variáveis observadas se relacionam entre si, é porque existe uma causa comum que gera essa inter-relação. Identificar essa causa comum e como ela se expressa no comportamento das variáveis observadas é, precisamente, o objetivo da análise fatorial.

Para realizar essa análise, considerando apenas uma dimensão preponderante (o nível socioeconômico familiar), checaremos seus resultados por meio de correlações entre cada variável e o fator, da magnitude de variância total explicada por esse fator e do diagrama de declividade (*scree plot*). De forma similar à matriz de correlações, a análise fatorial também nos informará um valor que indica a magnitude da correlação, dessa vez não mais das variáveis entre si, e sim entre a variável e o fator, também chamada de *carga fatorial*. Novamente, consideraremos o critério de  $\pm 0,30$  para sugerir uma carga fatorial de valor minimamente aceitável para a constituição de um fator (Pasquali, 2012). Sabendo que, em Ciências Sociais, é virtualmente impossível isolar um construto de outros congêneres, não buscamos a *unidimensionalidade estrita* (a associação entre os itens estaria relacionada a um único fator), e sim a *unidimensionalidade essencial* (a associação entre os itens estaria relacionada predominantemente a um fator), de acordo com Schilling (2007).

Novamente, a análise fatorial com matriz de correlação policórica foi empregada separadamente para as configurações de item A (2011), B (2012-14) e C (2015-17), assim como somente para os itens comuns a essas três configurações (ABC)<sup>10</sup>; em adição, realizamos uma análise fatorial de informação completa, para todos os itens independentemente de serem comuns ou não entre as edições do Enem. Na Tabela 5, exibimos as cargas fatoriais, referentes

<sup>10</sup> Em virtude de limitações da análise fatorial por meio do pacote *psych*, que não aceita itens com mais de oito opções de resposta, os itens Q17\_1 e Q17\_2 (renda familiar), foram recodificados para conter apenas oito alternativas cada, exclusivamente para a análise fatorial.

ao fator preponderante, dos itens de cada uma dessas análises. Vê-se que a menor carga fatorial é 0,31 para o item Q04\_3\_C, que corresponde à posse de aparelho de DVD. Conforme já discutimos na seção anterior, as baixas correlações encontradas entre esses itens e os demais do questionário já sugeriria fraca associação entre o item e o construto. A despeito disso, os valores obtidos nessa análise não são suficientemente baixos para que os itens sejam descartados (na medida em que superam o limiar de 0,30 tomado como referência).

**TABELA 5**  
**CARGAS FATORIAIS DOS ITENS NO FATOR PREPONDERANTE – ENEM 2011-2017**

(continua)

Itens	Configuração dos Itens				
	A	B	C	ABC (Itens Comuns)	Todos os Itens
Q01_ABC	0,73	0,72	0,71	0,73	0,71
Q02_ABC	0,78	0,73	0,72	0,75	0,70
Q03_AB	0,47	0,43	-	-	0,43
Q04_1_ABC	0,48	0,44	0,35	0,37	0,37
Q04_2_AB	0,63	0,57	-	-	0,58
Q04_3_C	-	-	0,35	-	0,31
Q05_C	-	-	0,70	-	0,70
Q06_C	-	-	0,61	-	0,64
Q07_ABC	0,55	0,52	0,57	0,55	0,61
Q08_ABC	0,67	0,65	0,63	0,64	0,61
Q09_1_ABC	0,67	0,67	0,72	0,68	0,74
Q09_2_AB	0,67	0,67	-	-	0,73
Q09_3_C	-	-	0,74	-	0,77
Q10_C	-	-	0,52	-	0,49
Q11_1_ABC	0,79	0,77	0,80	0,76	0,81
Q11_2_AB	0,79	0,77	-	-	0,81
Q11_3_C	-	-	0,79	-	0,81
Q13_1_ABC	0,82	0,80	0,78	0,79	0,83
Q13_2_AB	0,82	0,80	-	-	0,85
Q13_3_C	-	-	0,78	-	0,87
Q14_ABC	0,76	0,73	0,73	0,75	0,83
Q15_C	-	-	0,54	-	0,81
Q16_ABC	0,74	0,72	0,78	0,75	0,78
Q17_1_A	0,76	-	-	-	0,75
Q17_2_BC	-	0,77	-	-	0,79
Q18_ABC	0,64	0,62	0,65	0,65	0,79
Q19_ABC	0,51	0,53	0,60	0,54	0,63

TABELA 5

## CARGAS FATORIAIS DOS ITENS NO FATOR PREPONDERANTE – ENEM 2011-2017

(conclusão)

Itens	Configuração dos Itens				
	A	B	C	ABC (Itens Comuns)	Todos os Itens
Q20_ABC	0,76	0,76	0,77	0,77	0,56
Q21_1_ABC	0,83	0,71	0,62	0,67	0,73
Q21_2_AB	0,83	0,71	-	-	0,73
Q21_3_C	-	-	0,72	-	0,74
Q22_1_ABC	0,64	0,63	0,65	0,59	0,71
Q22_2_AB	0,64	0,63	-	-	0,66
Q22_3_C	-	-	0,64	-	0,67
Q23_1_ABC	0,58	0,57	0,60	0,53	0,67
Q23_2_AB	0,58	0,57	-	-	0,62
Q23_3_C	-	-	0,59	-	0,61
Q24_C	-	-	0,70	-	0,63
Q25_C	-	-	0,68	-	0,71

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Nota: Para as colunas "A", "B", "C" e "ABC (Itens Comuns)", foram realizadas análises fatoriais com matriz de correlação policórica; para a coluna "Todos os itens", foi realizada análise fatorial de informação completa.

A Tabela 6 revela que, em todas as configurações, a proporção de variância total explicada por uma dimensão predominante sobre os itens do NSE é de, no mínimo, 43,6% entre os itens comuns de 2015 a 2017. Para todos os itens, independentemente de serem comuns ou não às edições do Enem, obtivemos uma proporção de 46,1%. Tendo em vista que a literatura aponta o limiar de 20% como um indicativo da predominância de um fator (Reckase, 1979), concluímos que os valores encontrados neste estudo superam, em muito, esse patamar, conferindo-nos segurança para continuar as análises com a suposição de que tratamos de apenas uma dimensão preponderante do traço latente.

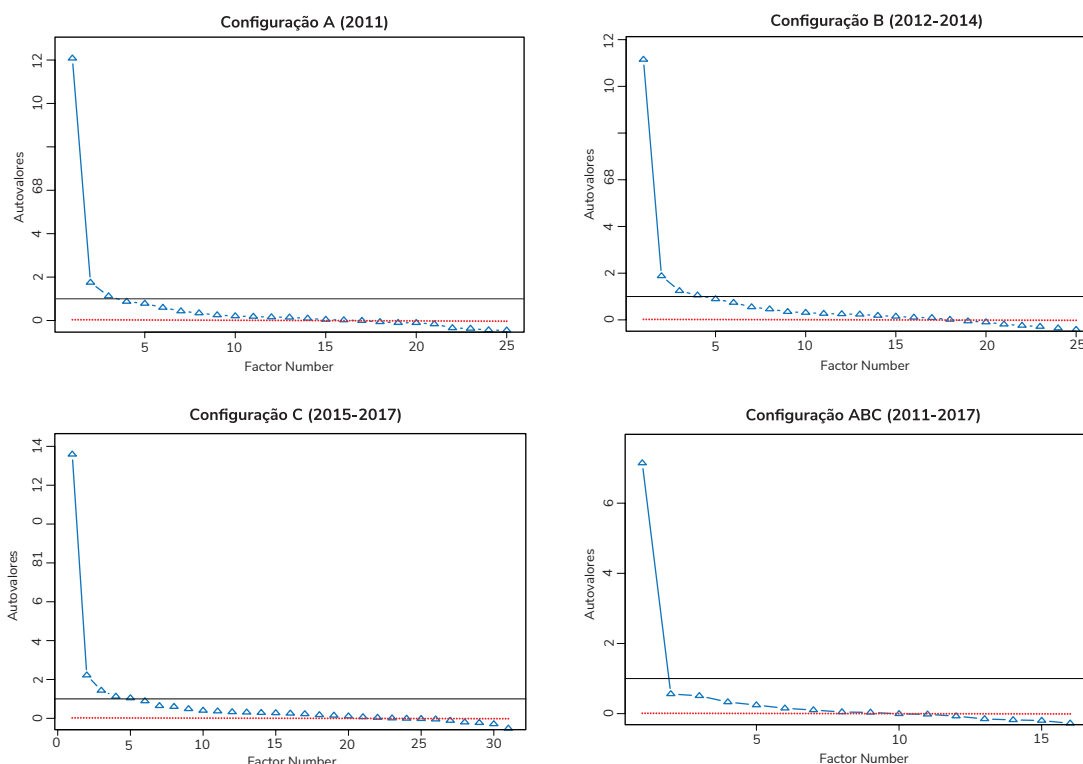
TABELA 6

## PROPORÇÃO DE VARIÂNCIA TOTAL EXPLICADA POR UM FATOR DO NSE – ENEM 2011-2017

Configuração	Ano	Proporção de Variância Explicada
A	2011	48,1%
B	2012-2014	44,3%
C	2015-2017	43,6%
ABC (Itens Comuns)	2011-2017	44,5%
Todos os Itens	2011-2017	46,1%

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Para reforçar esses achados, na Figura 2 são apresentados os diagramas de declividade relativos às quatro configurações de itens, elaborados através da análise fatorial por matriz de correlação policórica. Esses gráficos apresentam, no eixo-x, a quantidade possível de fatores existentes no conjunto dos dados; no eixo-y, plotam a magnitude dos autovalores (ou o quanto da variância é explicada pelo respectivo fator). Essa representação gráfica é particularmente interessante para ilustrar a importância relativa dos fatores. Quando há um fator preponderante no conjunto de dados, o diagrama de declividade tipicamente apresenta um ponto conspícuo seguido por uma queda vertiginosa que separa o primeiro fator dos demais. Esse padrão pode ser visualizado nos quatro gráficos apresentados.



**FIGURA 2**

**DIAGRAMA DE DECLIVIDADE, POR CONFIGURAÇÃO DOS ITENS DO NSE – ENEM 2011-2017**

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Na literatura, existe dissenso em torno da quantidade de valores a se reter por meio da análise gráfica de um diagrama como os apresentados acima: se, por um lado, alguns autores defendem que autovalores superiores a 1,00 indicam a relevância de um fator, outros mencionam que 0,70 já seria suficiente (Pasquali, 2012). Para resolver esse impasse, voltamos a afirmar: por estarmos buscando a unidimensionalidade essencial, a manifestação conspícua de um fator preponderante é suficiente para atestar a unidimensionalidade de nossas variáveis; além do mais, a proporção de variância explicada pelo primeiro fator nos dá segurança acerca dessa escolha. Para reforçar essa conclusão, voltaremos a esse ponto quando da verificação do pressuposto da independência local.

## 2.5 VALIDAÇÃO DO ORDENAMENTO DAS CATEGORIAS

Antes de continuarmos a verificação dos pressupostos da TRI, aplicamos um modelo politômico de respostas nominais, desenvolvido por Bock (1972). Segundo Thissen, Cai e Bock (2010), uma das utilidades desse modelo é validar empiricamente o ordenamento esperado das categorias de resposta de cada item do construto. Por não pressupor um ordenamento *a priori* das categorias, o modelo se presta à checagem da qualidade dos itens (Pinheiro; Costa; Cruz, 2010), na medida em que as questões do NSE se apresentam como variáveis categóricas ordinais em que se supõe que quanto maior a renda familiar ou a quantidade de bens de consumo, por exemplo, maior será o nível socioeconômico familiar de um indivíduo. Na prática, o modelo gera uma série de estimativas de parâmetros de discriminação para cada item  $i$  e categoria  $k$  ( $a_{i,k}$ ). Espera-se, para um item com categorias ordinais, que as estimativas desses parâmetros aumentem conforme aumente a categoria do item. Com isso, efetuamos a diferença entre  $a_{i,k}$  e  $a_{i,k-1}$ , a fim de verificar se o parâmetro de discriminação de uma determinada categoria tem um valor inferior ao da categoria anterior, ocorrência que indicaria algum problema de ordenamento das categorias. Embora os dados não tenham sido apresentados neste texto, nossas análises apontaram que essa situação não ocorreu em nenhum dos itens considerados<sup>11</sup>.

## 2.6 VERIFICAÇÃO DO PRESSUPOSTO DA INVARIÂNCIA

Ato contínuo, executamos as análises de função diferencial do item (DIF) entre as distintas edições do Enem, comparando singularmente cada uma delas, para testar o *pressuposto da invariância*. Em nosso caso, analisar o comportamento diferencial dos itens significa verificar se indivíduos que possuem o mesmo valor do Inse apresentam probabilidades distintas de assinalar determinada categoria das questões apresentadas. Para tanto, é preciso definir quais subpopulações se pretende comparar. Neste texto, em razão de termos compatibilizado itens de sete edições distintas do Enem, optamos por comparar o comportamento dos itens entre as edições, com duas finalidades: primeiro, verificar se é viável compor um indicador de nível socioeconômico a partir de itens apresentados a populações diferentes com intervalos temporais da magnitude de alguns anos; segundo, checar a validade da compatibilização de itens que, em um primeiro momento, tinham alterações nos enunciados e nas alternativas de respostas. Em outras palavras, caso um mesmo item apresente estimativas diferentes dos parâmetros em uma mesma escala entre as edições comparadas, o pressuposto da invariância não é atendido: diz-se que o item apresenta DIF (Glas, 2010). A ocorrência desse fenômeno pode comprometer a estimação de um indicador de nível socioeconômico familiar comparável ao longo dos anos.

<sup>11</sup> Vale, ainda, ressaltar que, nessa etapa do processo de construção do indicador, já havíamos excluído da base o item Q12 (Motocicleta), por ter apresentado problema de ordenamento na matriz de correlação policórica.

Existem diversos métodos para verificar DIF. Aqui, adotaremos aquele indicado por Mellenbergh (1989), o qual se baseia na inspeção visual das curvas características dos itens. A lógica dessa verificação é cotejar se há diferenças no padrão das curvas características dos itens entre os grupos, o que indicaria que o traço latente ( $\theta$ ) mensurado não varia somente em função dos parâmetros dos itens, mas também de características intrínsecas dos grupos (Andriola, 2000). Esse achado, por sua vez, enviesaria a estimação do nível socioeconômico familiar entre as distintas edições do Enem e colocaria em xeque o intuito de obter um indicador comparável ao longo do tempo. Assim, nesta etapa, são inspecionadas visualmente as curvas características de um mesmo item pelas sete edições do Enem, quando calibradas por meio de grupos múltiplos e garantindo a comparabilidade da escala. Com isso, para obter tal resultado, os parâmetros das questões serão estimados para cada grupo, pelo modelo de respostas graduais, e as questões que não apresentarem DIF, ou o apresentarem em um grau mínimo, terão seus parâmetros fixados enquanto as demais serão recalibradas; esses procedimentos são recomendados por Chalmers (2012). Para alcançar os resultados desejados, executamos seis rodadas de análise de invariância. A fim de que o texto não se torne demasiadamente extenso, exibiremos resultados selecionados da primeira análise, com a presença de todos os itens, para ilustrar alguns dos problemas encontrados e os encaminhamentos adotados.

Nas Figuras 3 e 4, observamos quatro curvas características dos itens, referentes a questões presentes nas sete edições do Enem. Para compreender melhor essas análises, esses gráficos devem ser assim interpretados: no eixo-x, dispomos de uma escala arbitrária contínua, que varia de -6 a 6 desvios-padrão em torno da média 0, expressando a variação do  $\theta$  que, no nosso caso, é o nível socioeconômico familiar; no eixo-y, temos a probabilidade de que os indivíduos de determinado  $\theta$  assinalem uma das categorias apresentadas (variando de 1 a 6, a depender da questão). Como se nota, o padrão geral das curvas é o progressivo deslocamento da esquerda à direita entre as categorias assinaladas, indicando que maiores quantidades de bens de consumo ou maiores níveis de renda familiar e escolaridade parental, por exemplo, estão correlacionados com maiores valores do índice socioeconômico – achado condizente com o ordenamento das categorias já validado até então. A contribuição de gráficos como estes é verificar em qual intervalo da escala de NSE as categorias dos itens variam, bem como se o padrão de resposta das categorias é similar entre as edições do Enem.

Na Figura 3, exibimos quatro gráficos referentes a questões que, já na primeira análise, apresentaram curvas características dos itens aceitáveis para um teste de invariância. Duas delas (Q07 – Geladeira e Q20 – Aspirador de pó) mostram que os itens são bastante comparáveis entre os anos, uma vez que as curvas referentes a cada edição do Enem têm um padrão bastante similar ao longo da escala do NSE, minimizando possíveis vieses do DIF. Quanto aos dois gráficos restantes (Q11\_1 - Carro/Automóvel e Q13\_1 – Computador/Microcomputador), os resultantes também nos deixam otimistas, uma vez que, mesmo com alterações importantes no enunciado<sup>12</sup>, os itens permaneceram compatíveis nos sete anos comparados, o que se expressa pela sobreposição das curvas em cada quadrante. Em outras

<sup>12</sup> Vale lembrar, entre 2011 e 2014, os itens Q11 e Q13 questionarem, respectivamente, quantos automóveis e quantos microcomputadores os respondentes possuíam em sua casa. Já entre 2015 e 2017, os mesmos itens interrogaram quantos carros e quantos computadores os respondente tinham em sua residência.

palavras, as alterações nos respectivos enunciados não geraram mudanças significativas no padrão geral das respostas.

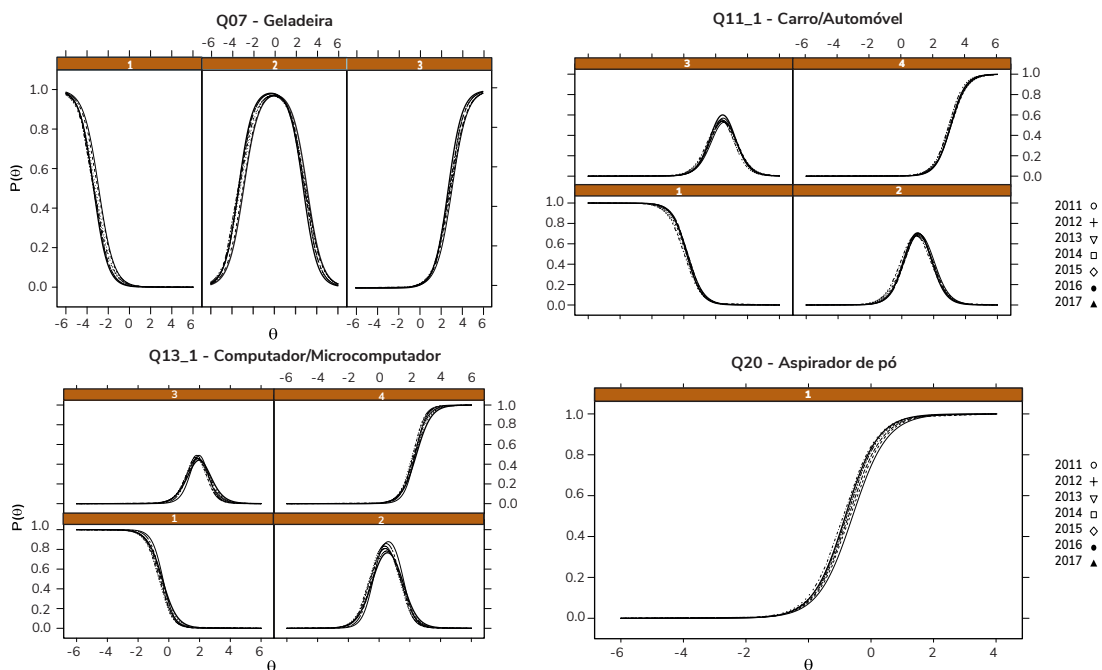


FIGURA 3

### CURVAS CARACTERÍSTICAS DE ITENS SELECIONADOS, POR EDIÇÃO DO ENEM 2011-2017

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Em contrapartida, com outros itens não encontramos resultados tão promissores. Na Figura 4, mostramos as curvas características de quatro itens que apresentaram DIF já na primeira rodada do teste. O primeiro deles, Q04\_1 (Videocassete/DVD), procurou compatibilizar dois itens bastante diferentes: entre 2011 e 2014, se o indivíduo possuía videocassete e/ou DVD em casa; de 2015 em diante, se o indivíduo possuía aparelho de DVD em casa. Como se vê no gráfico, ocorre um visível desencontro entre as curvas, separando, em especial, essas duas configurações do item<sup>13</sup>. O mesmo se aplica ao item Q21\_1 (Contratação de empregada doméstica), em que tentamos compatibilizar em um item dicotomizado uma questão apresentada entre 2011 e 2014 que interroga sobre a quantidade de empregados domésticos contratados, com outra de 2015 em diante sobre a frequência semanal do trabalho desses empregados. Novamente, o item apresentou forte DIF: um padrão de 2011 a 2013, outro de 2015 a 2017, e ainda um terceiro exclusivo de 2014<sup>14</sup>.

<sup>13</sup> O desencontro ano a ano das curvas do item Q04\_1 (videocassete/DVD) é um indicativo de obsolescência do item, fenômeno que sugere a possível eliminação deste do questionário. Esse processo de desgaste é conhecido na literatura como *DRIFT*, o qual pode ser entendido como uma tendência de divergência nos parâmetros do item associada ao tempo (Andrade; Laros; Gouveia, 2010).

<sup>14</sup> Não detectamos exatamente o motivo por trás de um DIF exclusivo para 2014; para dirimir eventuais problemas, resolvemos atribuir valores ausentes (*missing*) para esse item nessa edição.



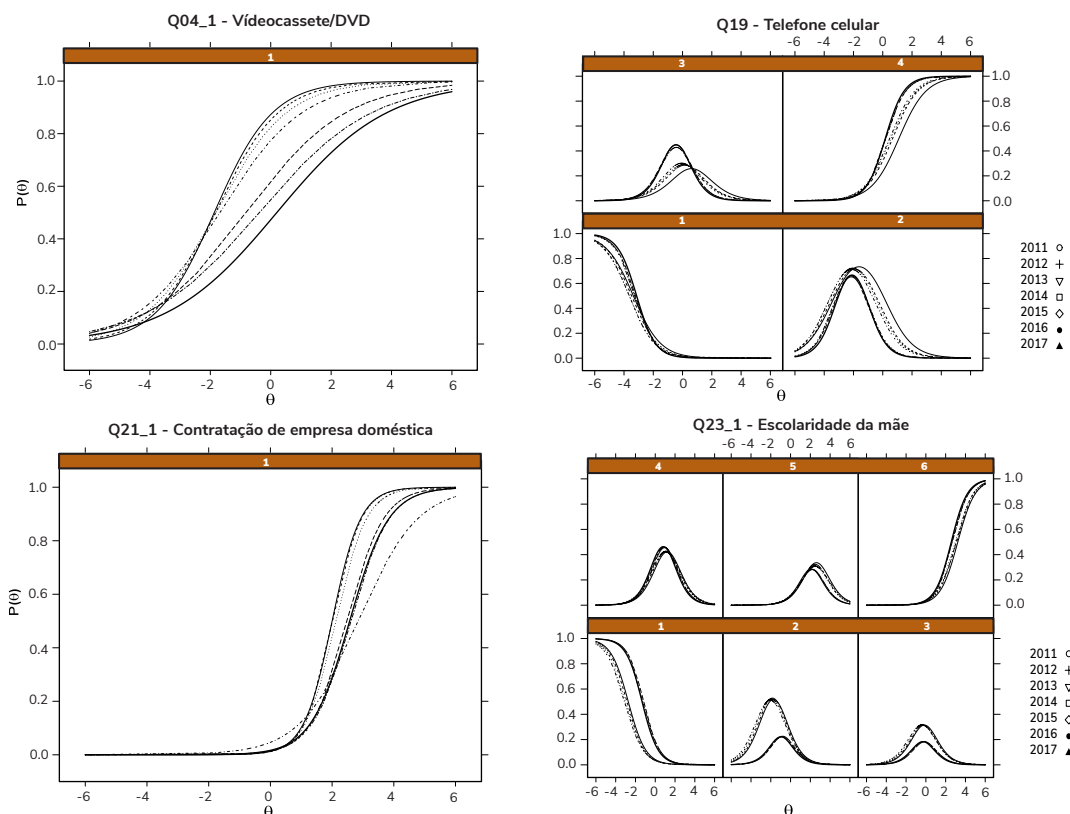


FIGURA 4

**CURVAS CARACTERÍSTICAS DE OUTROS ITENS SELECIONADOS,  
POR EDIÇÃO DO ENEM 2011-2017**

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

A respeito dessa mesma figura, percebemos a ocorrência de DIF em itens politômicos como Q19 (Telefone celular) e Q23\_1 (Escolaridade da mãe). Quanto ao primeiro, verifica-se um desencontro das curvas em várias das categorias apresentadas, sobretudo na categoria 3 (posse de dois telefones celulares em casa). Quanto ao segundo, também se percebeu discrepâncias no padrão das curvas, em especial nas três primeiras categorias – fruto de um reagrupamento de categorias originais que, como vemos, não foi feliz em seus resultados. Em situações como essas, nosso encaminhamento foi excluir os itens que não se mostraram compatíveis entre as sete edições do Enem e, em seguida, refazer a análise de invariância comparando as curvas características dos itens entre as edições do Enem de sua mesma configuração. Com isso, pudemos manter itens de construtos similares, porém em formatos diferentes (e.g. dois itens distintos sobre a escolaridade da mãe e do pai). Dado que estes foram apresentados a populações diferentes (*i.e.*, o conjunto de inscritos de cada edição do Enem em aplicações temporalmente distintas), não há empecilhos à manutenção desses itens em nossa base.

## 2.7 VERIFICAÇÃO DO PRESSUPOSTO DA INDEPENDÊNCIA LOCAL

Logo após, partimos para a verificação do *pressuposto da independência local*, o qual assume que, uma vez controlada pela dimensão comum aos itens, as respostas aos diferentes itens do questionário são localmente independentes (Andrade; Tavares; Valle, 2000, p. 17; Borsboom; Mellenbergh; Van Heerden, 2003, p. 208). Essa premissa é necessária para garantir a validade do modelo, dado que o seu não cumprimento indicaria que as respostas dos indivíduos dependem não apenas do nível socioeconômico em si, como também de fatores não observados (Embreston; Reise, 2000), violando, por consequência, a unidimensionalidade que se busca encontrar. Para checar esse ponto, adotamos a estatística  $Q_3$  de Yen (1993), índice calculado após a modelagem dos dados e que representa a correlação residual entre as questões, isto é, a correlação após a extração do traço latente mensurado. Sabendo que, quanto mais próximo de -1,00 ou 1,00 maior será a dependência entre os itens, tomamos como referência o critério sugerido por Ayala (2009) de que, para instrumentos com pelo menos 17 itens, um valor de corte razoável é  $|Q_3| > 0,20$ .

Para alcançar os resultados desejados, executamos o teste mencionado em três rodadas. O Quadro 3 exibe os valores do índice  $Q_3$  para os pares de itens que obtiveram dependências locais superiores a 0,20 e os respectivos encaminhamentos de cada rodada.

**QUADRO 3**  
**ANÁLISE DE INDEPENDÊNCIA LOCAL ENTRE PARES DE ITENS PELO ÍNDICE  $Q_3$  – ENEM 2011-2017**

Análises	Análise de independência local			Encaminhamentos
	Pares de itens		Índice $Q_3$	
1ª Rodada	Q13_1 (Computador)	Q14 (Internet)	0,20	Manter para nova verificação na próxima rodada
	Q15 (Quartos para dormir)	Q16 (Banheiros)	0,24	
	Q22_2 (Escolaridade do pai)	Q23_2 (Escolaridade da mãe)	0,41	Fusão no item Q26_1 (Escolaridade do pai ou da mãe)
	Q22_3 (Escolaridade do pai)	Q23_3 (Escolaridade da mãe)	0,38	Fusão no item Q26_2 (Escolaridade do pai ou da mãe)
	Q22_3 (Escolaridade do pai)	Q24 (Ocupação do pai)	0,28	Manter para nova verificação na próxima rodada
	Q23_3 (Escolaridade da mãe)	Q25 (Ocupação da mãe)	0,39	
	Q24 (Ocupação do pai)	Q25 (Ocupação da mãe)	0,27	
2ª Rodada	Q15 (Quartos para dormir)	Q16 (Banheiros)	0,22	Manter para nova verificação na próxima rodada
	Q26_2 (Escolaridade do pai ou da mãe)	Q27 (Ocupação do pai ou da mãe)	0,40	Excluir o item Q27 (Ocupação do pai ou da mãe)
3ª Rodada	Q15 (Quartos para dormir)	Q16 (Banheiros)	0,21	Manter os dois itens na base

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Em resumo, as análises empreendidas subsidiaram a fusão dos itens relativos à escolaridade do pai e da mãe em um único item – o que nos levou à criação das versões Q26\_1\_AB (2011-2014) e Q26\_2\_C (2015-2017) referentes aos itens de escolaridade mais alta da mãe ou do pai nas respectivas edições do ENEM –, e à exclusão dos itens Q24\_C e Q25\_C relativos à ocupação do pai e da mãe, respectivamente. Optamos por não proceder a nenhuma alteração com relação à acusação de dependência local entre os itens Q15 (Quartos para dormir) e Q16 (Banheiro) em razão, primeiro, da importância das duas questões para caracterização dos domicílios e, segundo, do fato de os valores  $Q_3$  terem caído de 0,24 na primeira rodada para 0,21 na terceira, estando bem próximo do liminar estabelecido para a verificação do pressuposto.

## 2.8 CALIBRAÇÃO DOS ITENS: ESTIMAÇÃO DOS PARÂMETROS E DO ERRO-PADRÃO

Uma vez finalizadas as etapas descritas até então, encaminhamos nosso indicador para os passos finais, os quais envolverão, neste momento, a estimação dos parâmetros e do erro-padrão dos itens de acordo com o modelo final da TRI. As análises anteriores ensejaram alterações nos itens originalmente elencados para a mensuração do nível socioeconômico familiar, tais como exclusões e fusões de itens, além do reagrupamento de categorias. A configuração final dos itens é apresentada no Quadro 4; para conhecer as categorias de cada item, indicamos a Tabela 1 (Apêndice C). Antes de ajustar o modelo de TRI, verificamos novamente a qualidade do conjunto de dados, por meio de uma análise fatorial de informação incompleta, e verificamos uma proporção de variância explicada, para todos os itens indicados no quadro, da ordem de 46,1%.

### QUADRO 4

#### LISTA FINAL DE ITENS UTILIZADOS PARA O CÁLCULO DO INSE – ENEM 2011-2017

(continua)

	Item	Objeto do Item	Nº de Categorias	Edições
1	Q01_ABC	TV em cores	4	2011-17
2	Q02_ABC	TV por assinatura	2	2011-17
3	Q03_AB	Rádio	4	2011-14
4	Q04_3_C	Aparelho de DVD	2	2015-17
5	Q05_C	Forno micro-ondas	3	2015-17
6	Q06_C	Máquina de lavar louça	2	2015-17
7	Q07_ABC	Geladeira	3	2011-17
8	Q08_ABC	Freezer	3	2011-17
9	Q09_1_ABC	Máquina de lavar roupa [mencionando ou não o tanquinho]	3	2011-17
10	Q10_C	Máquina de secar roupa	2	2015-17

#### QUADRO 4

##### LISTA FINAL DE ITENS UTILIZADOS PARA O CÁLCULO DO INSE – ENEM 2011-2017

(conclusão)

	Item	Objeto do Item	Nº de Categorias	Edições
11	Q11_1_ABC	Carro / automóvel	4	2011-17
12	Q13_1_ABC	Computador / Microcomputador	4	2011-17
13	Q14_ABC	Acesso a internet	2	2011-17
14	Q15_C	Quartos para dormir	5	2015-17
15	Q16_ABC	Banheiro	4	2011-17
16	Q17_1_A	Renda familiar [em faixas de salário mínimo]	11	2011
17	Q17_2_BC	Renda familiar [em faixas de salário mínimo]	17	2012-17
18	Q18_ABC	Telefone fixo	2	2011-17
19	Q19_ABC	Telefone celular	2	2011-17
20	Q20_ABC	Aspirador de pó	2	2011-17
21	Q21_2_AB	Contratação de empregada mensalista [quantidade]	2	2011-13
22	Q21_3_C	Contratação de empregada doméstica [frequência]	4	2015-17
23	Q26_1_AB	Escolaridade mais elevada do pai ou da mãe	8	2011-14
24	Q26_2_C	Escolaridade mais elevada do pai ou da mãe	7	2015-17

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Nota: Foram atribuídos valores ausentes ao item Q21\_2\_AB para o Enem 2014.

Conforme explanamos anteriormente, adotamos o modelo de respostas graduais de Samejima (1969) para a calibração dos itens, cujos parâmetros foram estimados pelo método de máxima verossimilhança marginal<sup>15</sup>. Esse modelo gera curvas características de itens que apresentam a distribuição probabilística da marcação de cada categoria ao longo de uma escala de NSE com média 0 e desvio-padrão 1, denominada escala (0,1). Na Figura 5, temos um painel com todas as curvas características dos 24 itens considerados no modelo final – para analisar cada curva separadamente, indicamos a Figura 1 (Apêndice A). Para a interpretação desses gráficos<sup>16</sup>: o deslocamento das curvas no eixo-x representa o parâmetro *b* de uma determinada categoria que, como discutimos anteriormente, está associado à “dificuldade” de cada categoria dos itens (i.e., ao limiar de  $\theta$  associado à probabilidade de um indivíduo marcar determinada categoria); já a inclinação das curvas está relacionada ao parâmetro *a*, ou seja, à capacidade de as categorias dos itens discriminarem os indivíduos ao longo da escala. Desse modo, percebe-se que o painel da Figura 5 exibe itens que cobrem a escala de NSE com características bastante distintas: enquanto alguns itens são altamente discriminativos (Q14 – Acesso a internet), outros são pouco (Q03 – Rádio); ao passo que alguns itens cobrem porções

<sup>15</sup> Para conhecer detalhes do processo de estimação, ver Andrade, Tavares e Valle (2000).

<sup>16</sup> Recomendamos a leitura de Pasquali (2013) para se aprofundar na interpretação das curvas características dos itens, em especial para questões dicotômicas.

mais elevadas da escala (Q08 – Freezer), outros abarcam porções menos elevadas (Q19 – Telefone celular). Logo, o que essas curvas nos mostram é como a diversidade de itens nos fornece informações variadas sobre o *continuum* da escala do indicador, permitindo que sejam estimados os níveis socioeconômicos familiares de indivíduos mais ou menos privilegiados em termos de renda, escolaridade e acesso a bens e serviços. Essa conclusão será explorada mais adiante quando atentarmos à curva de informação total do questionário (Gráfico 2).

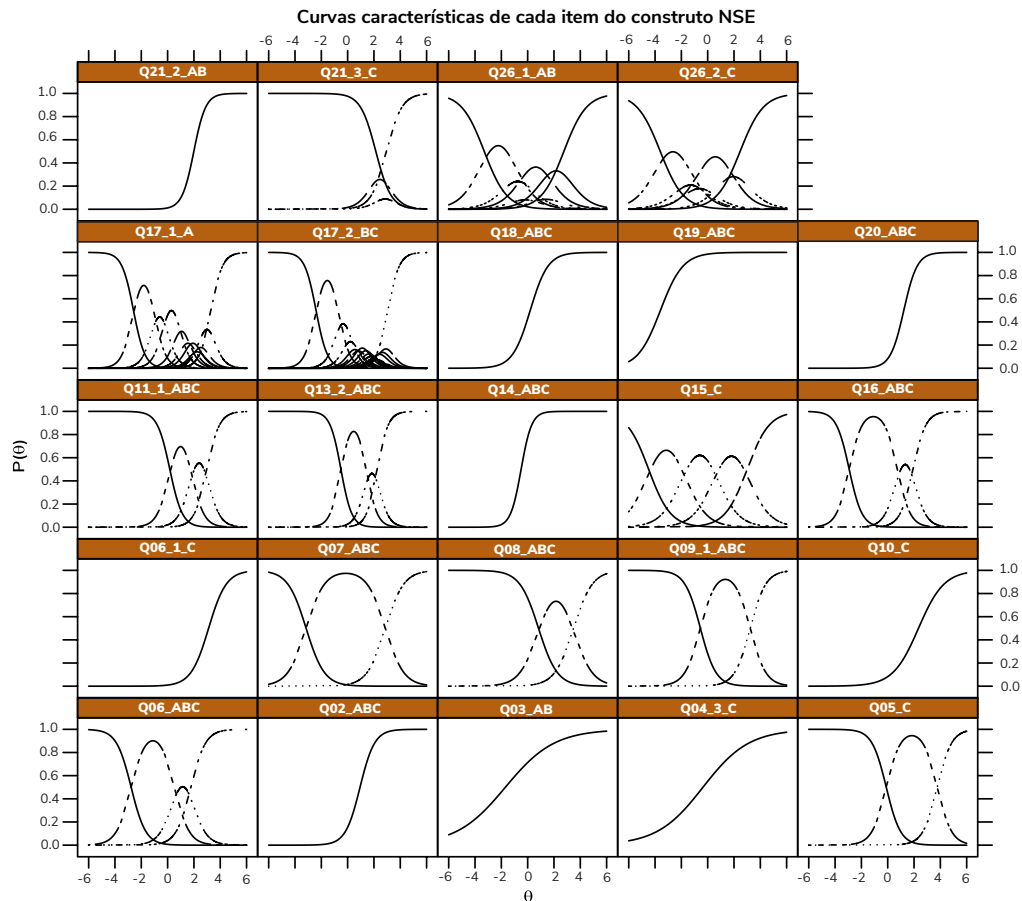


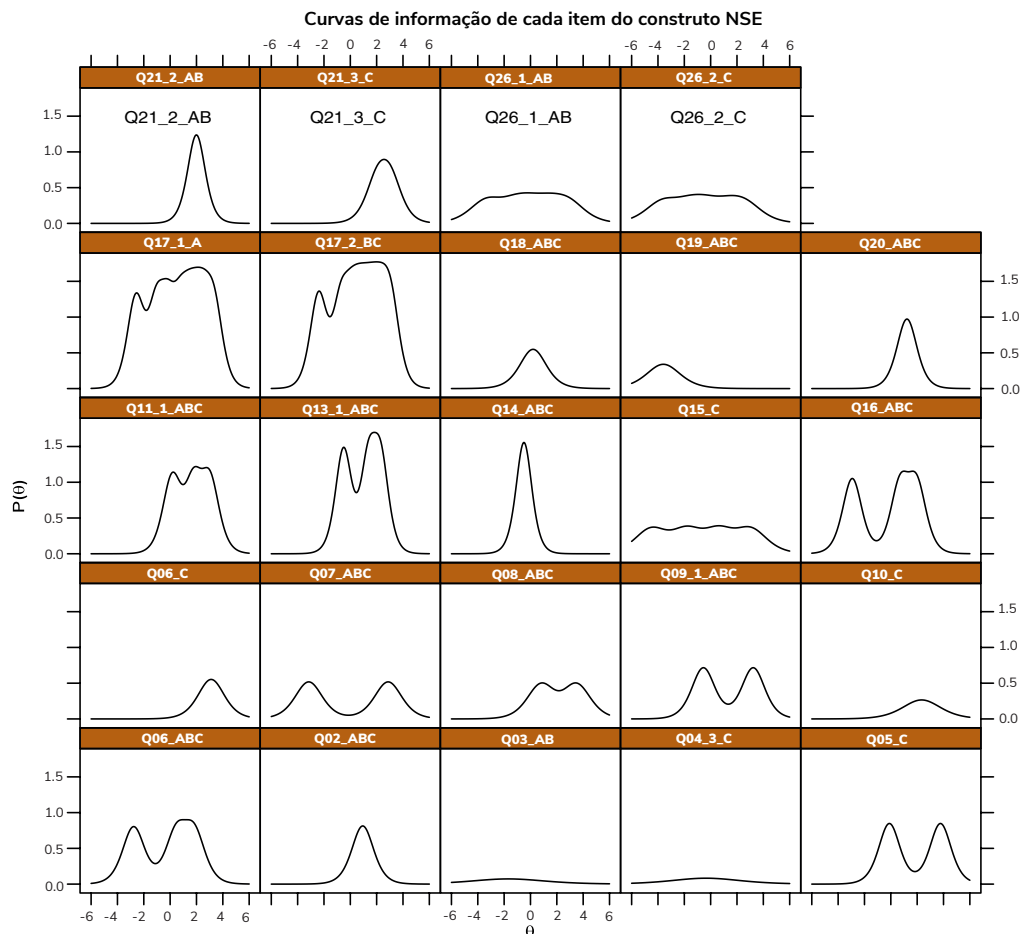
FIGURA 5

**CURVAS CARACTERÍSTICAS DOS ITENS DO INSE, SEGUNDO O MODELO DE RESPOSTAS GRADUAIS – ENEM 2011-2017**

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Em adição aos gráficos acima, apresentamos outro painel com as curvas de informação de cada item (Figura 6). Nestas, é apresentada a magnitude de informação (eixo-y) fornecida por cada item em função da posição na escala do NSE (eixo-x). Na prática, o que se percebe é quanto os itens contribuem para o cômputo do indicador e como essa contribuição se distribui ao longo da escala. Se, de um lado, itens como Q03 (Rádio) e Q04\_3 (Aparelho de DVD) são pouco informativos, de outro, itens como Q11\_1 (Carro/Automóvel) e Q13\_1 (Computador/Microcomputador) são bastante informativos; enquanto alguns itens (Q14 –

Acesso a internet e Q20 – Aspirador de pó) são informativos de porções específicas da escala, outros itens (Q17\_1 e Q17\_2 – Renda Familiar) são informativos de praticamente todo o espectro observado. Novamente, esses dados apontam para um papel de complementação dos itens do NSE, de modo que o conjunto deles seja capaz de cobrir, com diferentes graus de informação, toda a escala. Cabe ressaltar que, aqui, dispomos de um modelo teórico que varia de -6 a 6 desvios-padrão em torno da média 0; porém, em uma distribuição normal, sabe-se que 99,73% dos dados encontram-se entre -3 e 3 desvios-padrão<sup>17</sup>.



**FIGURA 6**

**CURVA DE INFORMAÇÃO DOS ITENS DO INSE, SEGUNDO O MODELO DE RESPOSTAS GRADUAIS – ENEM 2011-2017**

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Para finalizar, comentaremos brevemente sobre os erros-padrão. Assim como no processo de estimação dos parâmetros, o erro-padrão é calculado para os parâmetros

<sup>17</sup> Com efeito, optamos por apresentar os gráficos em uma escala ampliada, de -6 a 6, para a melhor visualização das curvas.

$a_i$  e  $b_{i,k}$ . Segundo nossos cálculos, o erro-padrão dos itens variou de 0,00 a 0,03, tendo uma média de 0,0079; vale salientar que o erro-padrão se apresentou baixo devido à quantidade de respostas a cada item ser bastante elevada. Apenas duas categorias de itens apresentaram erros-padrão de 0,03 no parâmetro  $b$ , quais sejam: Q15 (item “Quartos para dormir”/categoria “Sim, um”) e Q19 (item “Telefone celular”/categoria “Sim, um ou mais”).

## 2.9 ESTIMAÇÃO DO NÍVEL SOCIOECONÔMICO DOS INDIVÍDUOS

Com os itens calibrados após a etapa anterior, fixamos os parâmetros dos itens estimados no modelo de respostas graduais para posterior aplicação do mesmo método para a estimação do escore ou traço latente ( $\theta$ ), isto é, do nível socioeconômico familiar dos indivíduos. O primeiro passo, porém, foi garantir que as bases de dados com a população de inscritos nas sete edições do Enem apresentassem os itens em sua configuração final após todas as alterações efetuadas na base amostral. Assim, estendemos às bases populacionais todas as modificações aplicadas sobre os itens da amostra, tais como exclusões e fusões de questões, bem como reagrupamento de categorias, em suas versões finais exibidas na Tabela 1 (Apêndice C).

Em seguida, utilizamos os parâmetros fixados dos 24 itens do Inse para calcular o nível socioeconômico de cada indivíduo das edições do Enem 2011-2017. Evidentemente, cada edição do Enem contou com um conjunto específico de itens, a depender das configurações A, B e C discutidas até então. Como resultado desse processo de estimação, o escore dos indivíduos é apresentado em uma escala (0,1), ou seja, de média 0 e desvio-padrão 1 do grupo de inscritos do Enem de todas as edições utilizado na amostra de calibração dos itens. É preciso ter em mente que essa escala é intervalar e possui um zero arbitrário (que não significa a ausência da dimensão mensurada). Pelo fato de a média ser igual a 0, existirão valores negativos para o indicador, o que é pouco conveniente do ponto de vista da interpretação da medida. Logo, duas providências precisaram ser tomadas: a transformação da escala para outra mais comunicável e inteligível visando à sua interpretação; e a alocação dos indivíduos em uma mesma escala comparável ao longo do tempo.

Primeiramente, definimos arbitrariamente que os valores do Inse serão transformados para uma escala de média 50 e desvio-padrão 10. Essa operação foi realizada, em um primeiro momento, com os inscritos do Enem 2011, aqui definidos como o grupo de referência por ser o primeiro ano da série histórica<sup>18</sup>. Para tanto, aplicamos a seguinte transformação linear no  $\theta$  desses inscritos:

$$\theta_{(50,10)} = 10 * \left( \frac{\theta_{(0,1)} - \text{média}_{\text{grupo ref.}(0,1)}}{\text{desvio\_padrão}_{\text{grupo ref.}(0,1)}} \right) + 50$$

<sup>18</sup> Essa escolha se deu pelo interesse original de elaborar um painel dos inscritos do Enem, sendo 2011 a linha de base para as análises longitudinais. Esses resultados não serão apresentados neste texto.

Além disso, é indicado também colocar os parâmetros dos itens na nova escala:

$$a_{(50,10)} = \frac{a_{(0,1)}}{10 / \text{desvio\_padrão}_{\text{grupo ref.}(0,1)}}$$

$$b_{(50,10)} = 10 * \left( \frac{b_{(0,1)} - \text{média}_{\text{grupo ref.}(0,1)}}{\text{desvio\_padrão}_{\text{grupo ref.}(0,1)}} \right) + 50$$

Especificamente para o Enem 2011, dispomos dos seguintes valores:  $\text{média}_{\text{grupo ref.}} = -0,114659$  e  $\text{desvio\_padrão}_{\text{grupo ref.}} = -0,9111155$ . Esses valores devem ser aplicados tanto para a transformação da escala do Enem 2011, quanto para a transformação da escala das demais edições do Enem, mantendo-se a comparabilidade com o grupo de referência. Portanto, basta aplicar a fórmula (2) tomando-se a média e o desvio-padrão do grupo de referência. Mais informações obtidas após a estimação do nível socioeconômico e transformação da escala para todos os inscritos serão apresentadas na seção Resultados.

## 2.10 SEGMENTAÇÃO E INTERPRETAÇÃO DA ESCALA

Por mais valioso que seja dispor de um indicador de nível socioeconômico expresso em escala contínua, a fim de utilizá-lo como variável de controle em modelagens estatísticas, entendemos que, do ponto de vista teórico, a apresentação do NSE em grupos ordinais facilita a interpretação sociológica dos estratos sociais. Em vista disso, optamos por utilizar a técnica de análise de agrupamento hierárquico (*hierarchical cluster*) de Ward (1963), para segmentar a escala contínua em oito categorias ordenadas. De acordo com Hair et al. (2005), essa técnica é utilizada para classificar indivíduos em grupos de modo a maximizar as semelhanças dentro dos grupos e também as diferenças entre eles.

Sabendo que é possível dividir a escala em uma quantidade arbitrária de agrupamentos, optamos por oito grupos para prover uma discriminação relativamente elevada entre eles, com auxílio do coeficiente de aglomeração e do gráfico de dendrograma (Gráfico 1). Ao mesmo tempo, desejávamos que o agrupamento se diferenciasse da escala de sete grupos adotada oficialmente na versão original do Inse pelo Inep (2014), para evitar confusão entre os dois indicadores (o nosso, em nível individual; o do Inep, em nível escolar). Em adição a isso, caso pesquisadores optem por trabalhar com um número menor de agrupamentos, é possível agregar alguns deles com base na similaridade expressa no mesmo gráfico, prestando-se



atenção aos ramos comuns a partir dos quais se derivam os agrupamentos (caixas laranja)<sup>19</sup>. A título de ilustração, é possível agrupar os níveis 2 e 3 ou os níveis 7 e 8 da escala. Com isso, reduz-se a quantidade de grupos, o que, por um lado, facilita a análise e interpretação, mas, por outro, reduz a discriminação interna de cada agrupamento, com perdas de informações a serem ponderadas por cada pesquisador.

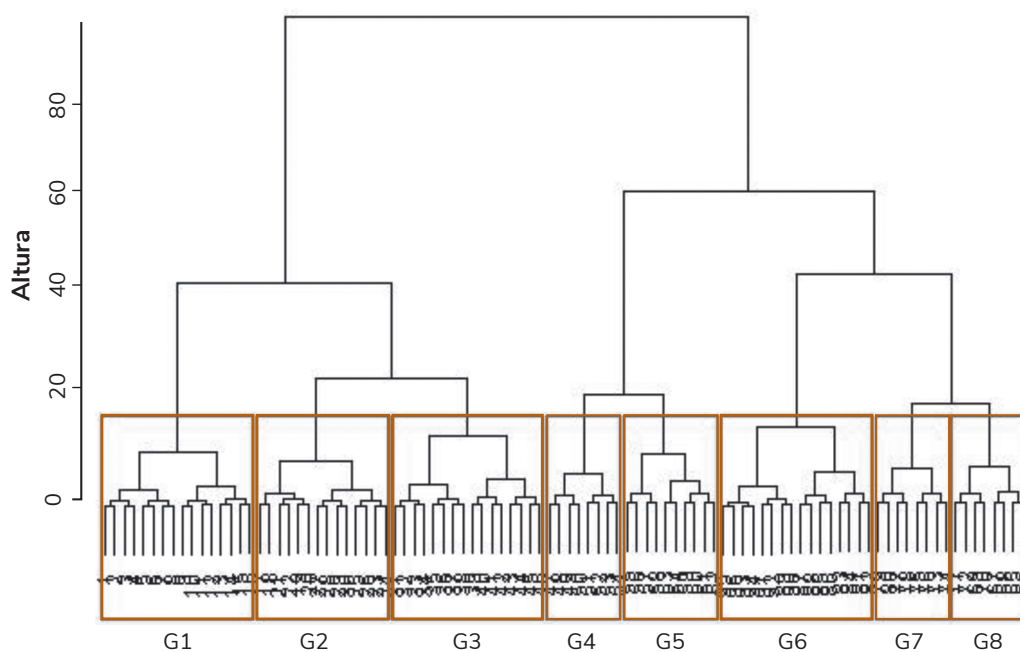


GRÁFICO 1

REPRESENTAÇÃO DENDROGRÂMICA DOS AGRUPAMENTOS HIERÁRQUICOS DO INSE (GRUPOS 1 A 8) – ENEM 2011-2017

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Logo após a segmentação da escala, encaminhamo-nos para a sua interpretação, etapa que corresponde à alocação dos itens ao longo da escala de nível socioeconômico com a finalidade de possibilitar uma leitura sociológica dos grupos. Para auxiliar nesta etapa, classificamos os itens de bens de consumo e serviços em três categorias – elementares, complementares e suplementares –, cujos detalhes são fornecidos no Apêndice B. Em seguida, adaptamos a técnica discutida por Barros, Borgatto e Oliveira (2017), de modo que, após a formação dos grupos definidos pelas faixas de Inse, foram calculadas as médias das probabilidades referentes a cada item/categoria em cada grupo e, a partir dessa média, a categoria foi descrita na faixa em que ela apresentava maior probabilidade de ocorrência (Tabela 1 do Apêndice C). Com base nas probabilidades médias, os grupos

<sup>19</sup> Ressalta-se que reagrupamentos arbitrários também são permitidos, embora eles não estejam respaldados na técnica utilizada.

foram descritos de forma que cada um equivalesse a um nível da escala, explicitando-se, portanto, a ideia de ordem e hierarquia. Para a descrição da escala e o posicionamento dos itens/categorias nos grupos formados na análise de agrupamento, foram utilizados três critérios, tal como descrito a seguir:

- critério 1: probabilidades maiores ou iguais a 0,60;
- critério 2: caso não se atenda ao critério anterior, consideram-se probabilidades maiores ou iguais a 0,30 quando a probabilidade de uma categoria próxima estiver entre 0,50 e 0,60;
- critério 3: caso não se atenda aos critérios anteriores, verifica-se se a soma das probabilidades na junção de categorias próximas é maior que 0,60.

Para esclarecer as medidas adotadas e visualizar os resultados, recomendamos uma leitura atenta da Tabela 1 (Apêndice C). Nela, é possível ver como cada categorias dos itens do NSE se distribuem em termos de probabilidade média entre os oito níveis da escala. Estão sombreadas em verde as probabilidades maiores ou iguais a 0,60 (critério 1) e de amarelo as probabilidades maiores ou iguais a 0,30 (critério 2); além disso, as categorias atribuídas para cada grupo estão contornadas com uma linha preta, podendo representar qualquer um dos três critérios, a depender do caso. Por fim, acrescenta-se que a descrição dos níveis da escala do NSE será apresentada na seção seguinte.

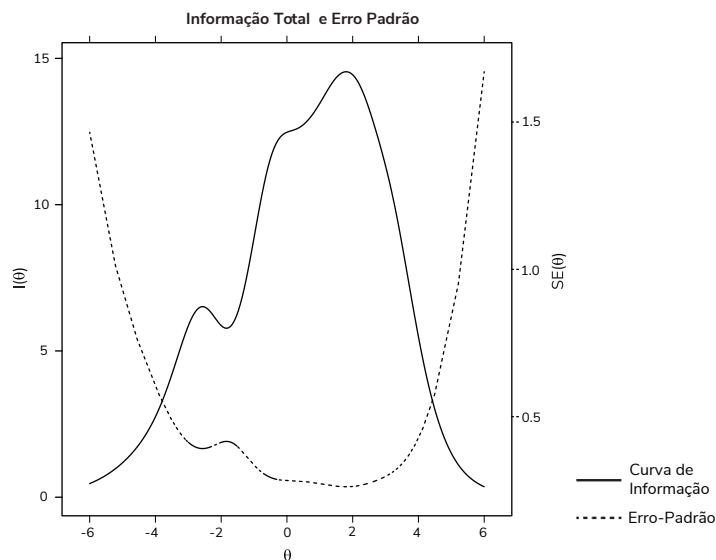
### 3 RESULTADOS

---

Discutiremos, nesta seção, os resultados do Inse dos inscritos do Enem 2011-2017, com o intuito de explorar o produto das etapas descritas anteriormente, apresentar critérios de validação do indicador e exemplificar algumas de suas aplicações. Salvo quando mencionado, os resultados desta seção foram obtidos a partir da amostra das sete edições do Enem – é válido lembrar que calibramos os parâmetros dos itens a partir da amostra justamente para permitir que os itens sejam alocados em uma mesma escala. Tanto na amostra quanto na população, consideraremos nas próximas páginas apenas os indivíduos que tenham respondido a pelo menos cinco questões do construto NSE.

Para iniciar, o Gráfico 2 ilustra as curvas de informação e do erro-padrão do modelo empregado, o que nos permite concluir que a modelagem dos nossos itens permite cobrir, com relativa qualidade, uma escala entre -3 e +3 desvios-padrão a partir da média; conforme cai a função de informação – geralmente nos valores extremos da distribuição –, cresce o erro-padrão, fenômeno que expressa o crescimento na imprecisão da medida para indivíduos pertencentes a estratos socioeconômico muito pobres ou muito ricos. De acordo com estudos anteriores, o modelo de respostas graduais é o que melhor fornece informações cobrindo a maior extensão da escala do indicador, possivelmente por extrair mais informações

de itens politômicos como os que compõem questionários como o utilizado – esses achados foram encontrados por pesquisadores que utilizaram outras bases de dados (Soares, 2005; Alves; Soares, 2009) ou que mesclaram dados do Enem com os Saeb (Alves; Soares; Xavier, 2014; Barros, 2016). É preciso dizer que repetimos a análise fatorial de informação completa para os 24 itens do NSE, após todos os ajustes descritos, e obtivemos uma proporção de variância explicada de 46,1%.

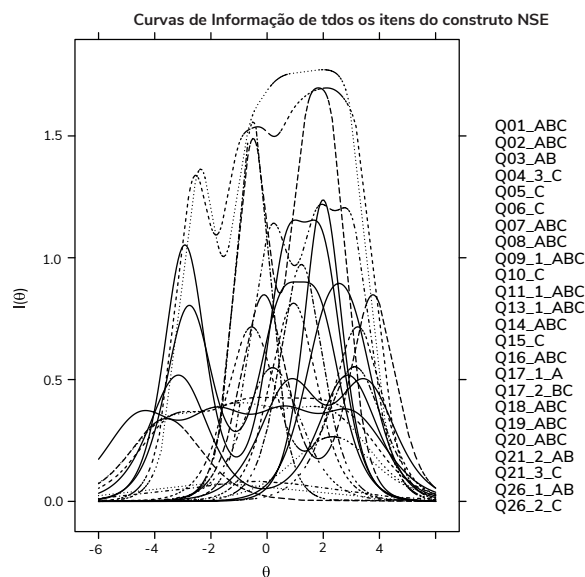


**GRÁFICO 2**

**CURVA DE INFORMAÇÃO E ERRO-PADRÃO DO MODELO DO INSE – ENEM 2011-2017**

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

É igualmente valioso explorar as curvas de informação sobrepostas de todos os itens (Gráfico 3) – em razão da quantidade de itens, não esperamos que o leitor interprete individualmente cada uma das curvas; chamamos atenção, contudo, para as duas curvas mais conspícuas, representando os itens de renda familiar (Q17\_1\_A e Q17\_2\_BC), os quais cobrem grande extensão da escala com maior magnitude de informação. Esse gráfico nos permite visualizar quais intervalos da escala do indicador estão menos cobertos por itens e, assim, sugerir futuras alterações no questionário para sanar essas lacunas. Vê-se, por exemplo, que o trecho da escala entre -1 e -3 desvios-padrão necessita de mais itens informativos. Isso significa que, em futuras modificações no instrumento de coleta dos dados, é indicado que se empenhe mais esforços na tentativa de obter informações sobre candidatos relativamente mais pobres. Uma leitura atenciosa da escala interpretada (Quadro 5) provê indícios de quais características socioeconômicas são típicas de indivíduos pertencentes a esse segmento da escala e, assim, fornece pistas sobre como aprimorar o questionário do Enem para a melhor aferição do construto.



**GRÁFICO 3**

### CURVAS DE INFORMAÇÃO SOBREPOSTAS DE TODOS OS ITENS DO INSE – ENEM 2011-2017

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Resultados gerais dos cálculos empreendidos nesta seção estão descritos na Tabela 7. Por definição, a média e o desvio-padrão do Inse no Enem 2011 são, respectivamente, 50,0 e 10,0. Já os demais anos, alocados na mesma escala, orbitam em torno desses mesmos valores. Nas três últimas edições, observa-se um aumento na amplitude dos valores, tanto por causa da redução dos valores mínimos quanto pelo aumento dos valores máximos, o que pode estar indicando aumento da disparidade socioeconômica entre os inscritos.

**TABELA 7**

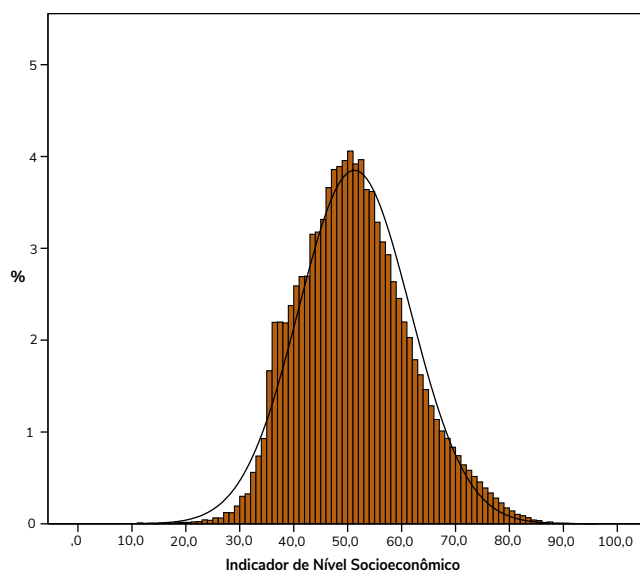
### ESTATÍSTICAS DESCRITIVAS DO INSE, SEGUNDO A EDIÇÃO DO ENEM 2011-2017

	2011	2012	2013	2014	2015	2016	2017
<b>Média</b>	50,00	51,14	51,20	50,98	50,73	50,03	49,89
<b>Desvio-padrão</b>	10,00	10,02	9,81	9,73	10,14	9,99	10,27
<b>Mínimo</b>	13,54	13,83	13,83	13,83	11,29	11,29	11,29
<b>1º quartil</b>	42,72	43,95	44,31	44,09	43,60	43,03	42,69
<b>Mediana</b>	49,36	50,70	50,76	50,55	50,19	49,54	49,29
<b>3º quartil</b>	56,09	57,32	57,28	57,07	56,96	56,17	56,20
<b>Máximo</b>	91,29	90,84	90,84	90,66	95,09	95,09	95,09
<b>Amplitude</b>	77,76	77,01	77,01	76,83	83,81	83,81	83,81
<b>Quantitativo</b>	5.366.948	5.791.065	7.173.563	8.722.247	7.745.734	8.627.367	6.731.340

Fonte: Elaboração própria com base de dados do Enem 2011-2017.

Nota: O quantitativo se refere apenas aos indivíduos com pelo menos cinco itens considerados no cômputo do indicador.

Para visualizar graficamente a distribuição percentual dos inscritos do Enem 2011-2017, o Gráfico 4 ilustra um histograma elaborado a partir da amostra. Nele, observa-se uma curva em forma de sino com dispersão em torno da média aproximada de 50,0. Evidentemente, o formato dessa curva é uma consequência da modelagem empregada, a qual termina por normalizar a distribuição do nível socioeconômico.

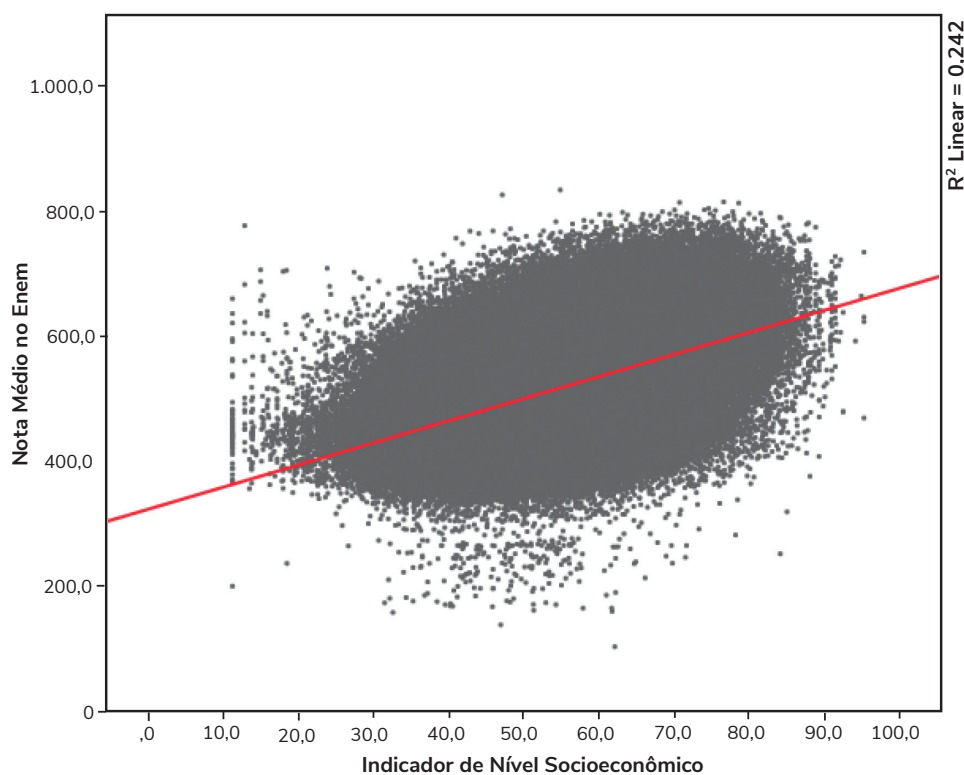


**GRÁFICO 4**

**HISTOGRAMA DA DISTRIBUIÇÃO PERCENTUAL DO INSE – ENEM 2011-2017**

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

A seguir, apresentaremos um conjunto de critérios de validação do indicador, a fim de checar a robustez da medida sintética aqui calculada. De partida, o Gráfico 5 exibe um gráfico de dispersão da nota média no Enem em função do nível socioeconômico do inscrito, no qual se percebe a associação existente entre os dois indicadores. Ademais, a análise de correlação linear de Pearson revelou uma correlação de magnitude 0,49 entre nível socioeconômico e proficiência média no Enem ( $N = 455.783$ ). Essa informação é importante porque confirma a qualidade da amostra extraída para a estimação dos parâmetros dos itens. A análise de correlação foi reproduzida para o total de inscritos de cada edição do Enem com resultados similares (dados que aqui não foram apresentados) e corrobora achados da literatura acerca da associação entre habilidades cognitivas e o contexto socioeconômico dos estudantes (Sirin, 2005; Soares; Andrade, 2006; Chudgar; Luschei, 2009; Matos; Rodrigues, 2016; Ferrão et al., 2018).



**GRÁFICO 5**

**DISPERSÃO DA NOTA MÉDIA NO ENEM EM FUNÇÃO DO INSE – 2011-2017**

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Nota: A “nota média no Enem” representa a média aritmética das notas nas quatro provas objetivas do exame.

Também verificamos a consistência externa do Inse comparando com uma miríade de indicadores sociais e econômicos de fontes diversas (Tabela 8). Para permitir correlações com a medida de NSE escolar calculada a partir do Saeb e Enem, com base em Alves, Soares e Xavier (2014), empilhamos as bases do Enem 2013 e 2015 e calculamos a média aritmética do Inse por escola, resultando em uma correlação de 0,89. Também comparamos, em agregados municipais<sup>20</sup>, o Inse do Enem 2011 com diversos indicadores obtidos a partir do Censo Demográfico 2010 (IDHM, RDPC, Renocup, PMPOB, PPOB e taxa de analfabetismo; vide tabela), tabulados por Pnud, Ipea e FJP (2013), encontrando correlações que variaram, em módulo, de 0,84 a 0,92, consideradas de elevada magnitude. Por fim, também comparamos em nível municipal o Inse do Enem 2017 com o Indicador de Oportunidades da Educação Brasileira (IOEB), índice sintético que agrega fatores relativos a insumos e resultados educacionais para estimar a qualidade das oportunidades educacionais de crianças e jovens em idade escolar (CPL, 2017), resultando em uma correlação de 0,70.

<sup>20</sup> O NSE municipal foi calculado utilizando a média aritmética simples do NSE individual de todos os inscritos do Enem 2011, residentes no respectivo município.

TABELA 8

## CORRELAÇÕES LINEARES DE PEARSON ENTRE O INSE DE DISTINTAS EDIÇÕES DO ENEM E UM CONJUNTO DE INDICADORES SOCIOECONÔMICOS – 2011/2013/2015/2017

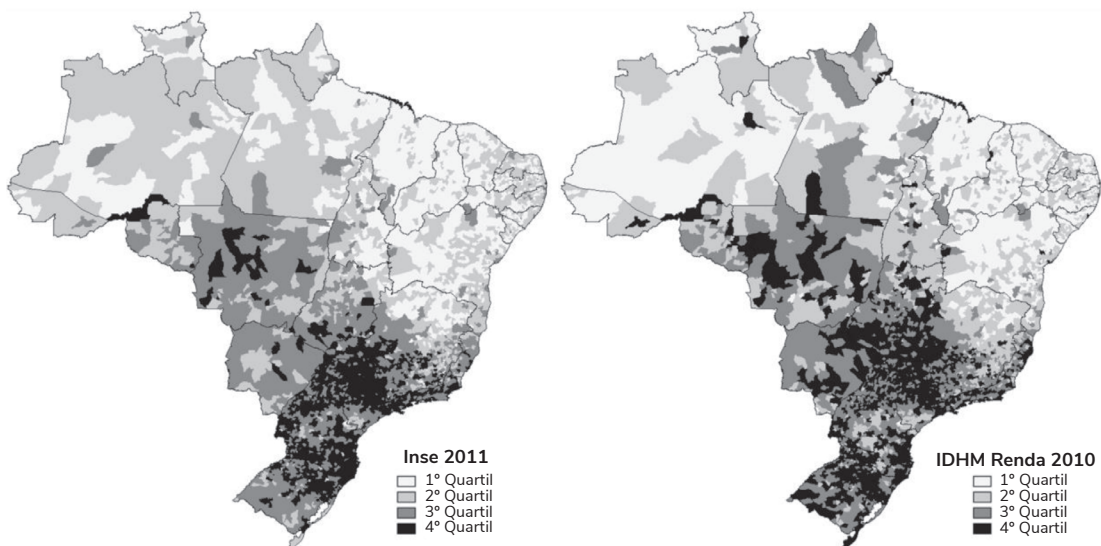
Correlações lineares de Pearson, bivariadas	Fonte	Ref.	Valor <sup>2</sup>
<b>Correlação com Inse do Enem 2013/15 (médio, escolar)</b>			
Medida de NSE das escolas a partir de Saeb e Enem	PB 2005/2011; Saeb 2003/2011; Enem 2007/2011 (Inep).	Alves; Soares; Xavier, 2014	0,89
<b>Correlação com Inse do Enem 2011 (médio, municipal)</b>			
IDHM-Renda - Renda municipal per capita	Censo Dem. 2010 (IBGE)	Pnud; Ipea; FJP, 2013	0,90
RDPC - Renda domiciliar per capita			0,87
Renocup - Rendimento médio dos ocupados com 18 anos ou mais			0,85
PMPOB - Proporção de pobres (indivíduos com renda domiciliar per capita ≤ R\$ 140,00 mensais)			-0,88
PPOB - Proporção de vulneráveis à pobreza (indivíduos com renda domiciliar per capita ≤ R\$ 255,00 mensais)			-0,92
Taxa de analfabetismo da população de 25 anos ou mais)			-0,84
<b>Correlação com Inse do Enem 2017 (médio, municipal)</b>			
IOEB - Índice de Oportunidades da Educação Brasileira	Inep/IBGE <sup>1</sup>	CLP, 2017	0,70

Fonte: Elaboração própria.

<sup>1</sup> 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 EM (IBGE).

<sup>2</sup> Correlações significantes a  $p < 0,01$ .

Para visualizar graficamente a correlação existente entre o Inse 2011 e o IDHM-Renda 2010, a Figura 5 ilustra, em nível municipal, o valor dos dois indicadores repartidos em quartis. Com essa imagem, o intuito é mostrar as similaridades no padrão geral dos dois mapas: os municípios constantes nos maiores quartis do Inse também são aqueles que ostentam os maiores valores de IDHM-Renda. É interessante notar que, enquanto o indicador calculado por Pnud, Ipea e FJP (2013) utiliza dados do questionário da amostra do Censo Demográfico, devidamente planejado para conferir resultados representativos em nível municipal com construtos voltados à aferição precisa da renda média mensal dos indivíduos, em reais de 1º de agosto de 2010, o indicador que discutimos neste texto se baseia na informação prestada por indivíduos, particularmente jovens egressos ou concluintes do ensino médio, que, ao se inscreverem no Enem, preenchem uma série de itens sobre a posse de bens de consumo, contratação de empregada doméstica, escolaridade parental e somente uma questão relativa à medida direta de renda. Apesar das distinções entre os dois levantamentos, os resultados não deixam dúvida a respeito da robustez da mensuração socioeconômica por meio do Inse ora apresentado e reforçam seu valor para os estudos e pesquisas educacionais.

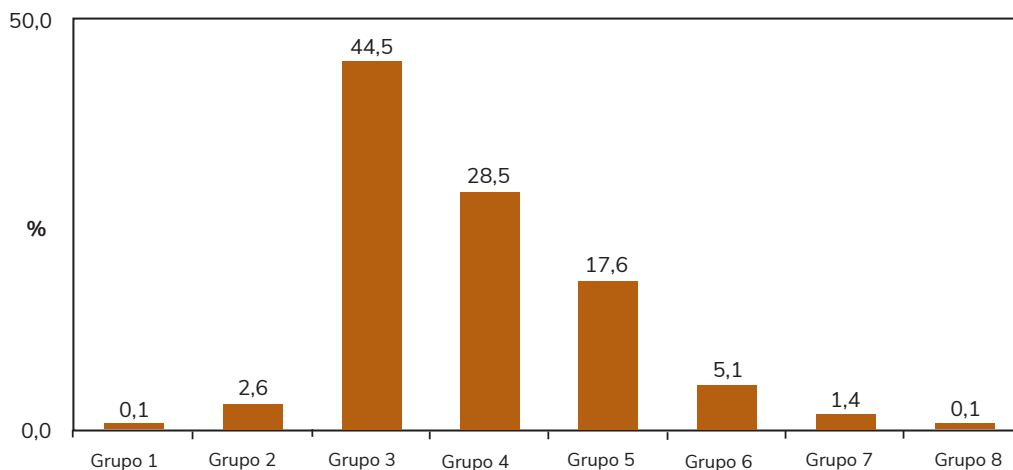


**FIGURA 5**

**CARTOGRAMA DE DISTRIBUIÇÃO DO INSE 2011 E DO IDHM-RENDA 2010 EM QUARTIS, POR MUNICÍPIO**

Fonte: Elaboração própria com base em dados do Enem 2011 e do Pnud, Ipea e FJP (2010).

Finalmente, exploraremos alguns resultados considerando as categorias ordinais geradas a partir da segmentação da escala do Inse. O Gráfico 6 ilustra um gráfico de barras com a frequência relativa dos oito níveis da escala na amostra do Enem. Como se vê, o Grupo 3 contém 44,5% dos inscritos, seguido pelo Grupo 4 (28,5%) e pelo Grupo 5 (17,6%). Embora pouco representativos do conjunto de inscritos, os Grupos 1 e 8, com apenas 0,1% cada, têm peculiaridades socioeconômicas que não podem ser ignoradas, por representarem, qualitativamente, os extremos da distribuição mensurada pelo nosso indicador.



**GRÁFICO 6**

**FREQUÊNCIA RELATIVA DOS GRUPOS DO INSE – ENEM 2011-2017**

Fonte: Elaboração própria com base em dados do Enem 2011-2017.



A fim de possibilitar a interpretação sociológica dos níveis da escala, o Quadro 5 apresenta a descrição de cada um dos oito grupos. É preciso ter em mente que, pelo fato de o modelo ser probabilístico, a descrição da escala não é unívoca em delimitar todas as possibilidades dos indivíduos que se encontram em determinadas faixas do indicador; pelo contrário, os oito grupos contêm indivíduos que, tipicamente, apresentam as características descritas. Importa relembrar que a descrição dos grupos se deu a partir das probabilidades expostas na Tabela 1 (Apêndice C).

**QUADRO 5**  
**INTERPRETAÇÃO SOCIOLÓGICA DA ESCALA DO INDICADOR DE NÍVEL**  
**SOCIOECONÔMICO FAMILIAR – ENEM 2011-2017**

(continua)

Nível	Intervalo	Descrição
Grupo 1	(0 - 20]	Este grupo é composto pelos indivíduos de menor nível socioeconômico familiar. Tipicamente, seus pais ou não frequentaram escola, ou estudaram até o 5º ano do ensino fundamental, e o conjunto dos habitantes da residência não possui renda mensal regular. Em seus domicílios, dispõem somente de alguns bens elementares como um quarto para dormir e, na melhor das hipóteses, uma geladeira e telefone celular; não possuem banheiro em suas residências, tampouco um bem complementar ou suplementar.
Grupo 2	(20 - 34]	Neste grupo, de nível socioeconômico bastante baixo, os indivíduos pertencem a famílias que recebem, no máximo, um salário-mínimo mensal, e cujos pais estudaram até o 5º ano do ensino fundamental, mas não necessariamente o completaram. Em seus domicílios, é possível encontrar uma quantidade mínima de bens elementares (um ou dois quartos para dormir, um banheiro, uma televisão em cores, uma geladeira e telefone celular); podem ou não possuir um rádio e não dispõem de bens complementares ou suplementares.
Grupo 3	(34 - 50]	Em geral, os indivíduos alocados neste grupo vivem com uma renda familiar de até 1,5 salário-mínimo e seus pais têm provavelmente o ensino fundamental como maior nível de escolaridade. Seus domicílios dispõem de um mínimo de bens elementares (até dois quartos para dormir, um banheiro, uma televisão em cores, uma geladeira, rádio e telefone celular); podem ou não possuir algum bem complementar como aparelho de DVD e máquina de lavar roupa; poucos entre estes tendem a possuir computador e acesso à internet.
Grupo 4	(50 - 58]	Tipicamente, as famílias dos indivíduos que pertencem a este grupo têm renda mensal variando de 1 a 2,5 salários-mínimos e o ensino médio como o maior nível de escolaridade, tendo parcela considerável de seus pais completado os estudos no ensino fundamental. Seus domicílios têm bens elementares (dois ou três quartos para dormir, um banheiro, televisão em cores, geladeira) e uma quantidade mínima de bens complementares (forno micro-ondas, máquina de lavar roupa, computador e acesso à internet); podem ou não possuir carro, aparelho de DVD e de telefone fixo, mas ainda não apresentam bens suplementares.
Grupo 5	(58 - 68]	Neste grupo, os indivíduos provêm de famílias que recebem entre 1,5 e 5 salários-mínimos mensais, e cujos pais completaram pelo menos o ensino médio. Em suas residências, estão presentes uma quantidade razoável de bens elementares (dois ou três quartos para dormir, um ou dois banheiros, uma ou duas televisões em cores, uma geladeira) e uma quantidade mínima de bens complementares (aparelho de DVD, computador, acesso à internet, forno micro-ondas, máquina de lavar roupa, telefone fixo e somente um carro); ademais, podem ou não possuir TV por assinatura e alguns bens suplementares como freezer e aspirador de pó.

## QUADRO 5

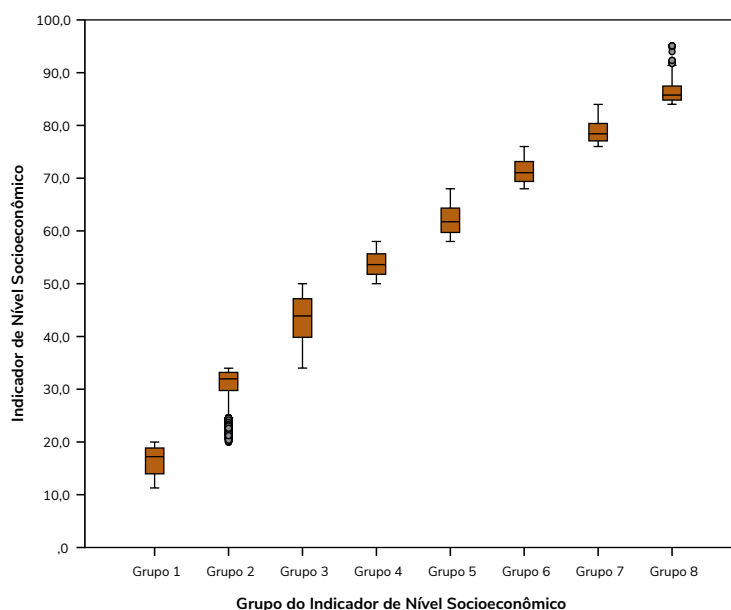
### INTERPRETAÇÃO SOCIOLÓGICA DA ESCALA DO INDICADOR DE NÍVEL SOCIOECONÔMICO FAMILIAR – ENEM 2011-2017

(conclusão)

Nível	Intervalo	Descrição
Grupo 6	(68 - 76]	Este grupo abarca indivíduos de famílias que vivem com renda mensal variando entre 3 e 10 salários-mínimos e com pais que, tendo concluído o ensino médio, podem ter concluído a graduação ou mesmo a pós-graduação. Além de uma quantidade maior de bens de consumo elementares (três quartos para dormir, pelo menos dois banheiros e duas televisões), em suas residências tipicamente se encontra bens complementares (um ou dois carros, aparelho de DVD, máquina de lavar roupa, acesso à internet) e alguns bens suplementares (freezer e aspirador de pó); podem ou não contratar empregada doméstica e, se o fazem, na frequência de uma ou duas vezes por semana.
Grupo 7	(76 - 84]	Este grupo contém indivíduos de elevados valores de nível socioeconômico familiar, com renda mensal igual ou superior a 10 salários-mínimos; seus pais completaram pelo menos a graduação, tendo uma boa parte concluído a pós-graduação. Suas residências apresentam no mínimo três quartos para dormir, três banheiros, uma ou duas geladeiras, pelo menos dois carros e contêm ao menos três televisores e três computadores, além de um mínimo dos demais bens elementares e complementares. Quanto aos bens suplementares, possuem freezer, aspirador de pó, e podem ou não possuir máquina de secar roupa e de lavar louça. De modo geral, contratam empregada doméstica uma ou mais vezes por semana.
Grupo 8	(84 - 100]	No grupo de maior nível socioeconômico familiar, encontram-se indivíduos cujos pais tipicamente completaram a pós-graduação e que vivem com rendas familiares superiores 20 salários-mínimos. Seus domicílios apresentam pelo menos quatro quartos para dormir e três banheiros. Como bens de consumo, possuem quantidades mais elevadas de bens elementares (três televisores, por exemplo) e complementares (três computadores, três carros), e uma quantidade mínima de bens suplementares. Provavelmente contratam empregada doméstica para uma frequência de no mínimo cinco vezes por semana.

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Para visualizar como os níveis se diferem ao longo da escala do indicador, o Gráfico 7 ilustra o boxplot da distribuição do NSE segundo o grupo. Essa figura permite observar que os grupos se distinguem sem sobreposição entre as porções extremas da distribuição correspondente, fruto da técnica adotada para a segmentação dos níveis.



**GRÁFICO 7**

### BOXPLOT DO INSE SEGUNDO OS NÍVEIS SEGMENTADOS DA ESCALA – ENEM 2011-2017

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

## CONSIDERAÇÕES FINAIS

Neste texto, procuramos apresentar as motivações, o referencial teórico, os procedimentos metodológicos e alguns dos resultados e aplicações do Indicador de Nível Socioeconômico (Inse) dos inscritos do Enem 2011-2017, construído a partir das respostas ao questionário que acompanha a inscrição nas respectivas edições do exame, sobre uma miríade de características socioeconômicas da família e dos domicílios dos candidatos. É preciso reiterar que apesar de o indicador ora apresentado ser inédito, ele é tributário de esforços de pesquisadores que, dentro e fora do Inep, têm envidado esforços na geração de uma medida válida, robusta e confiável da condição socioeconômica do corpo discente brasileiro. Nesta última seção, comentamos algumas contribuições e limitações do indicador.

Acreditamos que o maior mérito dos procedimentos aqui empregados reside em não apenas construir um indicador de relevância social e utilidade prática para pesquisas educacionais, como também gerar insumos para pensar a melhoria dos questionários socioeconômicos do Enem e como torná-los ainda mais proveitosos para o estudo do nível socioeconômico de seus participantes. Como vimos ao longo do texto, os resultados intermediários produzidos a partir da TRI permitem levantar uma série de informações em profundidade sobre cada item que compõe o construto e sobre o indicador como um todo.

Embora não tenhamos desenvolvido uma análise longitudinal do índice, é possível deduzir que, em contraste com indicadores socioeconômicos tradicionalmente calculados a partir

da renda corrente (particularmente aqueles derivados dos rendimentos do trabalho), o Inse não é uma medida indicada para aferição de variações das condições socioeconômicas em um curto intervalo de tempo, precisamente por ter grande parte de seu construto baseado em medidas patrimoniais que refletem a renda permanente (*i.e.* o estoque, o acúmulo), que, ao contrário da renda corrente (*i.e.* o fluxo), não responde tão rapidamente às variações de rendimento. Essa particularidade do Inse tem seu valor por hipoteticamente garantir uma medida mais estável, menos sujeita à volatilidade do mercado.

Outra contribuição do Inse é permitir a agregação dos resultados para os níveis escolar, municipal, estadual, regional e nacional, de acordo com o quantitativo de inscritos do Enem oriundos de cada instituição escolar ou ente da Federação. No entanto, por mais que tenhamos obtido elevadas correlações em nível escolar e municipal, nunca é demais asseverar que a metodologia do indicador não é desenhada para se obter resultados representativos para esses níveis de agregação. A implicação disso é que pesquisadores devem ter cautela ao empregar o Inse para tratar de desigualdades entre municípios ou regiões, por exemplo. Conforme temos ressaltado ao longo do texto, o Inse foi projetado para se obter resultados em nível individual e, portanto, como uma medida sintética pessoal da condição socioeconômica familiar, referida aos inscritos no Enem e não ao alunado brasileiro em seu conjunto. Para discussões em nível escolar, recomendamos o indicador de nível socioeconômico proposto pelo Inep desde 2014, o qual foi pensado para fornecer uma medida socioeconômica agregada da escola a partir dos questionários do Enem e do Saeb, contemplando, assim, estudantes de diferentes anos/séries escolares em um único indicador.

Originalmente, o objetivo de um índice sintético da posição social dos estudantes foi contextualizar o desempenho destes nas avaliações e exames em larga escala. Com a consolidação do indicador, entendemos que seus usos podem e devem transcender essa aplicação. Mais do que uma variável de contexto para a proficiência em testes cognitivos, o Inse pode ser empregado com fins contextuais para estudos que explorem outras variáveis – a título de ilustração: aprovação, reprovação, abandono ou conclusão de uma etapa de ensino; transição de uma etapa de ensino à etapa seguinte; desenvolvimento de habilidades não cognitivas ou socioemocionais; pode também ser utilizado para finalidades variadas, tais como a focalização de recursos orçamentários e/ou insumos educacionais, uma referência para a construção de novos indicadores, a investigação das desigualdades socioeconômicas em distintos níveis de agregação geográfica. Suas aplicações são amplas, sobretudo em sociedades como a brasileira, em que o caráter difuso das disparidades nas mais diversas dimensões sociais obrigam pesquisadores a sempre considerarem os efeitos da estratificação sobre a conformação das oportunidades de vida.

Convidamos pesquisadores interessados em desenvolver suas próprias análises com os indicadores aqui produzidos a adquirir os microdados desta pesquisa. Estes são compostos por sete bases de dados (de 2011 a 2017), em formato “.csv”, contendo a lista de inscritos no Enem do período analisado, acompanhados dos itens do construto do Inse, tal como utilizado para o modelo final da TRI, os escores do indicador calculado nas escalas original (média = 0 e desvio-padrão = 1) e transformada (média = 50 e desvio-padrão = 10),

os erros-padrão do indicador e, finalmente, o grupo correspondente na escala interpretada. Como chave de ligação, fornecemos a máscara dos microdados públicos do Enem, permitindo que o pesquisador cruze os microdados desta pesquisa com os do Enem divulgados no Portal do Inep<sup>21</sup>. Em conjunto, disponibilizamos como metadado um dicionário de variáveis. Com isso, incentivamos os pesquisadores interessados a não apenas realizarem suas próprias análises sobre o Inse, como também a o empregar como indicador contextual em suas próprias pesquisas. Por fim, estamos abertos a prover detalhes de todos os resultados intermediários de cada etapa deste estudo, aqui omitidos por limitação de espaço, bem como a tirar dúvidas que por ventura permaneçam.

## REFERÊNCIAS

---

ALLAIRE, J. J. et al. *Rmarkdown: dynamic documents for R*. 2018. Disponível em: <<https://CRAN.R-project.org/package=rmarkdown>>. Acesso em: 2 abr. 2018.

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 Política Pública e Educação*, Rio de Janeiro, v. 22, n. 84, p. 671-703, jul./set. 2014.

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. *Teoria da resposta ao item: conceitos e aplicações*. São Paulo: ABE, 2000. Disponível em: <<http://egov.ufsc.br/portal/sites/default/files/livrotri.pdf>>. Acesso em: 6 fev. 2019.

ANDRADE, J. M.; LAROS, J. A.; GOUVEIA, V. V. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. *Avaliação Psicológica*, São Paulo, v. 9, n. 3, p. 421-435, 2010.

ANDRIOLA, W. B. Funcionamento diferencial dos itens (DIF): estudo com analogias para medir o raciocínio verbal. *Psicologia: Reflexão e Crítica*, Porto Alegre, v. 13, n. 3, p. 475-483, 2000.

AYALA, R. J. *The theory and practice of item response theory*. New York, NY: The Guilford Press, 2009.

---

<sup>21</sup> Disponível em: <<http://portal.inep.gov.br/web/guest/microdados>>. Acesso em: 19 jul. 2018.

- BARROS, G. T. de F. *Procedimentos para a construção de indicadores por meio da Teoria de Resposta ao Item: a criação de uma medida de nível socioeconômico familiar*. 2016. 163 f. Dissertação (Mestrado Profissional em Métodos e Gestão da Avaliação) – Programa de Pós-Graduação em Métodos e Gestão em Avaliação, Universidade Federal de Santa Catarina, Florianópolis, 2016.
- BARROS, G. T. de F.; BORGATTO, A. F.; OLIVEIRA, A. S. de. Proposta de segmentação de uma escala de TRI utilizando o nível socioeconômico. *Examen: Pesquisa em Avaliação, Certificação e Seleção*, Brasília, v. 1, n. 1, p. 72-94, jul./dez. 2017.
- BISTAFFA, B. C. *Incorporação de indicadores categóricos ordinais em modelos de equações estruturais*. 2010. 130 f. Dissertação (Mestrado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2010.
- BOCK, R. D. Estimating item parameters and latent ability when responses are scored in two or more latent categories. *Psychometrika*, New York, v. 37, n. 1, p. 29-51, Mar. 1972.
- BOLLEN, K. A.; GLANVILLE, J. L.; STECKLOV, G. Socioeconomic status and class in studies of fertility and health in developing countries. *Annual Review of Sociology*, Palo Alto, n. 27, p. 153-185, 2001.
- BOLLEN, K.; LENNOX, R. Conventional wisdom on measurement: a structural equation on perspective. *Psychological Bulletin*, Washington, D.C., v. 110, n. 2, p. 305-314, 1991.
- BORSBOOM, D.; MELLENBERGH, G. J.; VAN HEERDEN, J. The theoretical status of latent variables. *Psychological Review*, Washington, D.C., v. 110, n. 2, p. 203-219, 2003.
- BOURDIEU, P. *Razões práticas: sobre a teoria da ação*. Campinas: Papirus, 1996.
- BOURDIEU, P. *A distinção: crítica social do julgamento*. 2. ed. Porto Alegre: Zouk, 2007.
- BOURDIEU, P. A escola conservadora: as desigualdades frente à escola e à cultura. In: NOGUEIRA, M. A.; CATANI, A. (Org.). *Pierre Bourdieu: escritos de educação*. 16. ed. Petrópolis: Vozes, 2015. p. 43-72.
- BOURDIEU, P.; PASSERON, J. C. *A reprodução: elementos para uma teoria do sistema de ensino*. 7. ed. Petrópolis: Vozes, 2014.
- BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Indicadores educacionais: notas técnicas: Indicador de Nível Socioeconômico das Escolas de Educação Básica (Inse)*. Brasília, 2014. Disponível em: <<http://portal.inep.gov.br/web/guest/indicadores-educacionais>>. Acesso em: 22 abr. 2016.
- BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Microdados: Enem: 2011-2017*. Brasília, [s.d.]. Disponível em: <<http://portal.inep.gov.br/microdados>>. Acesso em: 14 out. 2019.

- 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 LIDERANÇA PÚBLICA (CLP). *Nota técnica: índice de oportunidades educacionais do Brasil – IOEB*. São Paulo: Metas Avaliação e Proposição de Políticas Sociais, 2017. Disponível em: <<http://conteudo.clp.org.br/metodologia-ioeb>>. Acesso em: 30 abr. 2019.
- CHALMERS, P. Mirt: a multidimensional item response theory package for the R environment. *Journal of Statistical Software*, Innsbruck, v. 48, n. 6, p. 1-29, 2012.
- CHUDGAR, A.; LUSCHEI, T. F. National income, income inequality, and the importance of schools: a hierarchical cross-national comparison. *American Educational Research Journal*, Washington, D.C., v. 46, n. 3, p. 626-658, 2009.
- COLEMAN, J. S. et al. *Equality of Educational Opportunity*. Washington, DC: US Department of Health, Education and Welfare, 1966.
- CONNELL, R.W. et al. *Making the difference: schools, families and social division*. Sydney; London; Boston: George Allen & Unwin Australia, 1982.
- COSTA, F. M. da. *O modelo de Samejima no estudo da relação NSE x desempenho via teoria da resposta ao item*. 2010. 64 f. Dissertação (Mestrado em Estatística) – Programa de Pós-Graduação em Matemática e Estatística, Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, 2010.
- COUTO, G.; PRIMI, R. Teoria de resposta ao item (TRI): conceitos elementares dos modelos para itens dicotômicos. *Boletim de Psicologia*, São Paulo, v. 51, n. 134, p. 1-15, 2011.
- DE AYALA, R. J. The influence of multidimensionality on the graded response model. *Applied Psychological Measurement*, [S.l.], v. 18, n. 2, p. 155-170, June 1994.
- EDWARDS, J. R. The fallacy of formative measurement. *Organizational Research Methods*, [S.l.], v. 14, n. 2, p. 370-388, 2011.
- EDWARDS, J. R.; BAGOZZI, R. P. On the nature and direction of relationships between constructs and measures. *Psychological Methods*, [S.l.], v. 5, n. 2, p. 155-174, 2000.
- EMBRETSON, S. E.; REISE, S. *Item response theory for psychologists*. Mahwah, NJ: Erlbaum Publishers, 2000.

FERRÃO, M. E. et al. Estudo longitudinal sobre eficácia educacional no Brasil: comparação entre resultados contextualizados e valor acrescentado. *DADOS: Revista de Ciências Sociais*, Rio de Janeiro, v. 61, n. 4, p. 265-300, 2018.

GIDDENS, A.; SUTTON, P. W. *Conceitos essenciais da sociologia*. São Paulo: Editora Unesp, 2016.

GLAS, C. A. W. Testing fit to IRT models for polytomously scored items. In: NERING, M. L.; OSTINI, R. (Ed.). *Handbook of polytomous item response theory models*. New York; London: Routledge, 2010. p. 185-208.

HAIR, J. F. et al. *Análise multivariada de dados*. 5. ed. Porto Alegre: Bookman, 2005.

HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. *Fundamentals of item response theory*. Newbury Park; London; New Delhi: SAGE Publications, 1991.

HIRANO, S. *Castas, estamentos e classes sociais: introdução ao pensamento sociológico de Marx e Weber*. 3. ed. São Paulo: Editora da Unicamp, 2002.

JIANG, S.; WANG, C.; WEISS, D. J. Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in Psychology*, [S.l.], n. 7, p. 1-10, 2016.

KAMAKURA, W. A.; MAZZON, J. A. *Estratificação socioeconômica e consumo no Brasil*. São Paulo: Blucher, 2013.

MACHADO, D. C. de O. *Análise de fatores associados ao desempenho escolar de alunos do quinto ano do ensino fundamental com base na construção de indicadores*. 2014. 227 f. Dissertação (Mestrado Profissional em Métodos e Gestão em Avaliação) – Programa de Pós-Graduação em Métodos e Gestão em Avaliação, Universidade Federal de Santa Catarina, Florianópolis, 2014.

MATOS, D. A. S.; RODRIGUES, E. C. Indicadores educacionais e contexto escolar: uma análise das metas do Ideb. *Estudos em Avaliação Educacional*, São Paulo, v. 27, n. 66, p. 662-688, 2016.

MELLENBERGH, G. J. Item bias and item response theory. *International Journal of Educational Research*, [S.l.], v. 13, n. 2, p. 127-143, 1989.

MILIBAND, R. Análise de classes. In: GIDDENS, A.; TURNER, J. (Org.). *Teoria social hoje*. São Paulo: Editora Unesp, 1999. p. 471-502.

NOGUEIRA, M. A.; NOGUEIRA, C. M. M. *Bourdieu & a educação*. 4. ed. Belo Horizonte: Autêntica Editora, 2016.



PASQUALI, L. *Análise fatorial para pesquisadores*. Brasília: UnB, 2012.

PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. 5. ed. Petrópolis: Vozes, 2013.

PATTO, M. H. S. *A produção do fracasso escolar: história de submissão e rebeldia*. São Paulo: Casa do Psicólogo, 1999.

PINHEIRO, I. R.; COSTA, F. R.; CRUZ, R. M. Modelo nominal da teoria de resposta ao item: uma alternativa. *Avaliação Psicológica*, Porto Alegre, v. 9, n. 3, p. 437-447, 2010.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO (Pnud); INSTITUTO DE PESQUISA ECONÔMICA APLICADA (Ipea); FUNDAÇÃO JOÃO PINHEIRO (FJP). *O Índice de Desenvolvimento Humano Municipal Brasileiro*. Brasília, 2013. (Atlas do Desenvolvimento Humano no Brasil 2013).

R CORE TEAM. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. 2018. Disponível em: <<https://www.R-project.org/>>. Acesso em: 2 abr. 2018.

RECKASE, M. D. Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics*, [S.l.], v. 4, n. 3, p. 207-230, 1979.

REISE, S. P.; YU, J. Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement*, [S.l.], n. 27, p. 133-144, 1990.

REVELLE, W. *Psych: procedures for psychological, psychometric, and personality research*. 2018. Disponível em: <<https://CRAN.R-project.org/package=psych>>. Acesso em: 2 abr. 2018.

SAMEJIMA, F. *Estimation of latent ability using a response pattern of graded scores*. Richmond, VA: Psychometric Society, 1969. (Psychometric Monograph, n. 17). Available in: <<https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf>>. Access in: 11 Oct. 2019.

SARTES, L. M. A.; SOUZA-FORMIGONI, M. L. O. Avanços na psicometria: da teoria clássica dos testes à teoria de resposta ao item. *Psicologia: Reflexão e Crítica*, Porto Alegre, v. 26, n. 2, p. 241-250, 2013.

SCHILLING, S. G. The role of psychometric modeling in test validation: an application of multidimensional item response theory. *Measurement: Interdisciplinary Research and Perspectives*, [S.l.], v. 5, n. 2-3, p. 93-106, Aug. 2007.

SENKEVICS, A. S.; MACHADO, T. de S.; OLIVEIRA, A. S. *A cor ou raça nas estatísticas educacionais: uma análise dos instrumentos de pesquisa do Inep*. Brasília, 2016. (Texto para Discussão, n. 41).

SILVEIRA, F. L.; BARBOSA, M. C. B.; SILVA, R. Exame Nacional do Ensino Médio (Enem): uma análise crítica. *Revista Brasileira de Ensino de Física*, São Paulo, v. 37, n. 1, mar. 2015.

SIRIN, S. R. Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research*, [S.l.], v. 75, n. 3, p. 417-453, 2005.

SOARES, F. J.; MAMBRINI, J. Medida do nível socioeconômico de estudantes em pesquisas educacionais. In: REUNIÃO REGIONAL DA ABE, 35.; REUNIÃO REGIONAL DA SOBRAPO, 2003, Florianópolis. *Anais...* Florianópolis, 2003.

SOARES, J. F.; ALVES, M. T. G. Efeitos de escolas e municípios na qualidade do ensino fundamental. *Cadernos de Pesquisa*, São Paulo, v. 43, n. 149, p. 492-517, 2013.

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

SOARES, T. M. Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no Simave-2002. *Estudos de Avaliação Educacional*, São Paulo, n. 28, p. 103-124, 2003.

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

SOARES, T. M.; SOUZA, R. C.; PEREIRA, V. R. Métodos alternativos no Critério Brasil para construção de indicadores sócio-econômicos: teoria da resposta ao item. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 36., 2004, São João Del Rei. *Anais...* Rio de Janeiro: Sobrapo, 2004. Disponível em: <<http://www.din.uem.br/sbpo/sbpo2004/pdf/arq0089.pdf>>. Acesso em: 23 ago. 2015.

STAVENHAGEN, R. Estratificação social e estrutura de classes. In: STAVENHAGEN, R. *Estrutura de classes e estratificação social*. Rio de Janeiro: Zahar Ed., 1973. p. 133-170.

THISSEN, D.; CAI, L.; BOCK, R. D. The nominal categories item response model. In: NERING, M. L.; OSTINI, R. (Ed.). *Handbook of polytomous item response theory models*. New York; London: Routledge, 2010. p. 43-75.

TILLÉ, Y.; MATEI, A. *Sampling: survey sampling*. 2016. Disponível em: <<https://CRAN.R-project.org/package=sampling>>. Acesso em: 2 abr. 2018.

TRAVITZKI, R. *Enem: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar*. 2013. 320 f. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2013.

WARD, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, [S.l.], n. 58, p. 236-244, 1963.

WEININGER, E. B. Fundamentos de uma análise de classe de Pierre Bourdieu. In: WRIGHT, E. O. (Org.). *Análise de classe: abordagens*. Petrópolis: Vozes, 2015. p. 97-132.

WICKHAM, H. The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, Innsbruck, v. 40, n. 1, p. 1-29, 2011.

XIE, Y. et al. *Knitr*: a general-purpose package for dynamic report generation in R. 2018. Disponível em: <<https://rdr.io/cran/knitr/>>. Acesso em: 2 abr. 2018.

YEN, W. M. Scaling performance assessments: strategies for managing local item dependence. *Journal of Educational Measurement*, Washington, v. 30, n. 3, p. 187-213, 1993.





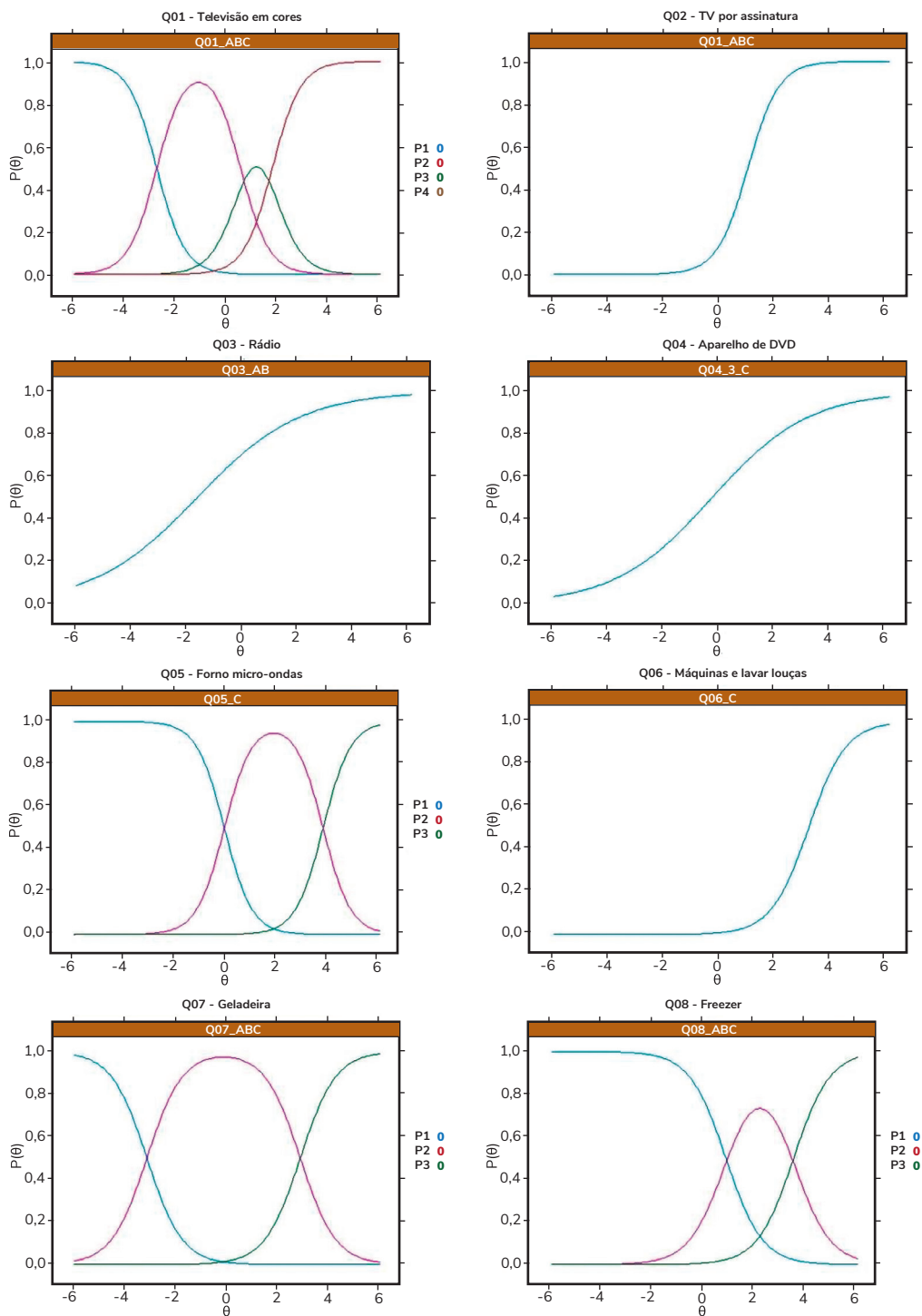
.....

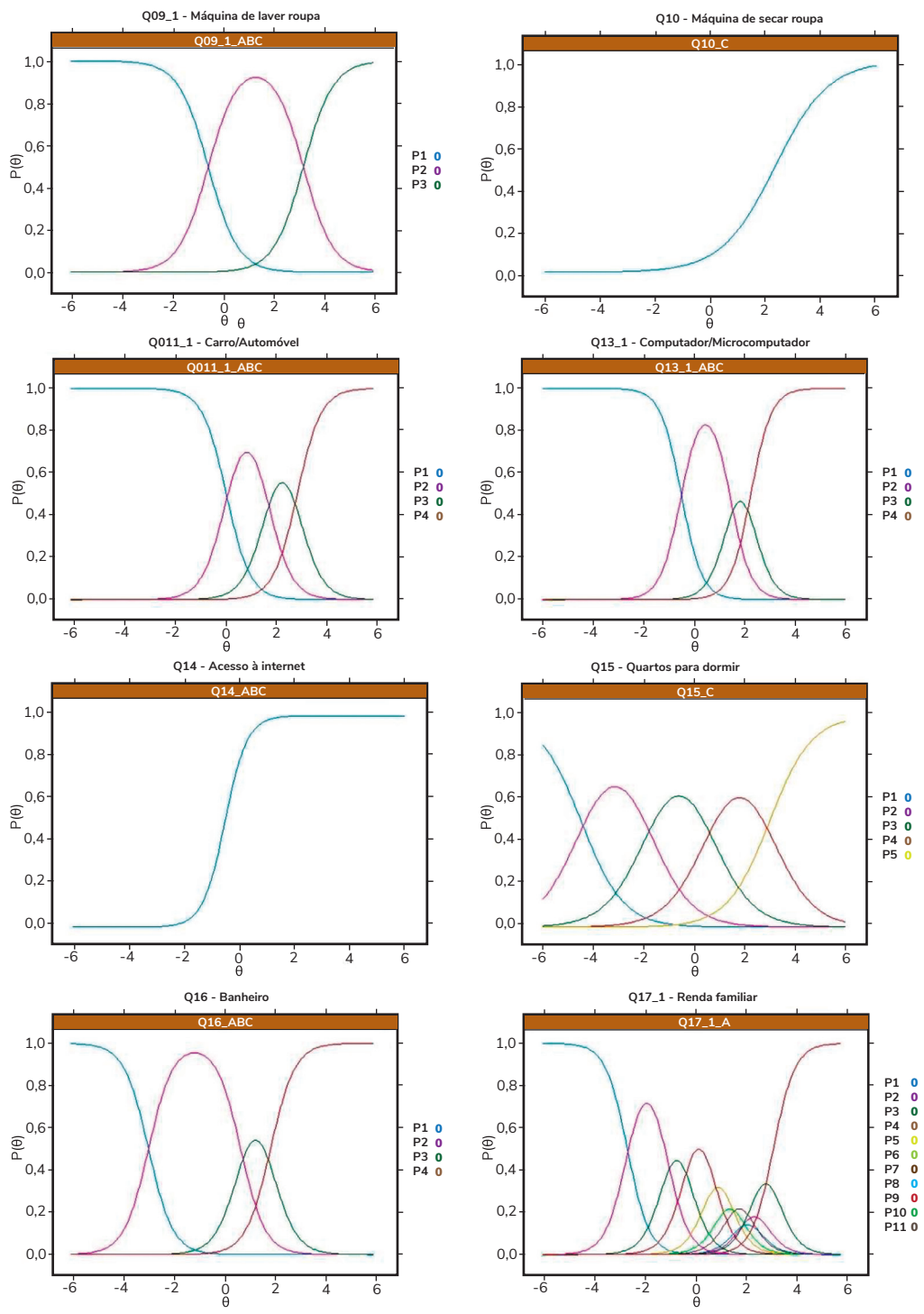
# APÊNDICE

.....

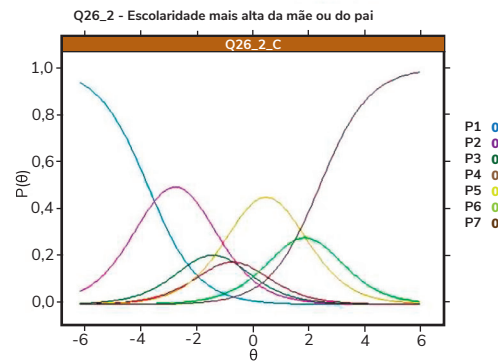
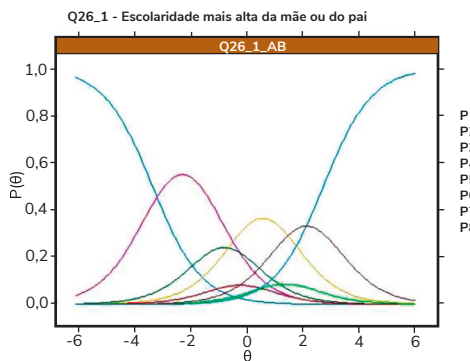
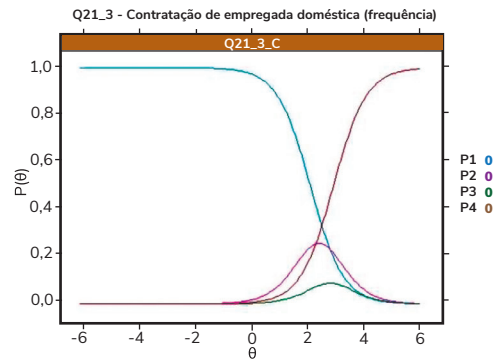
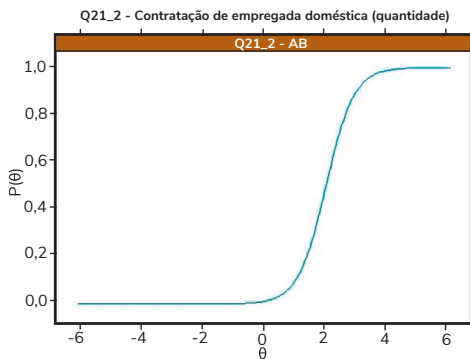
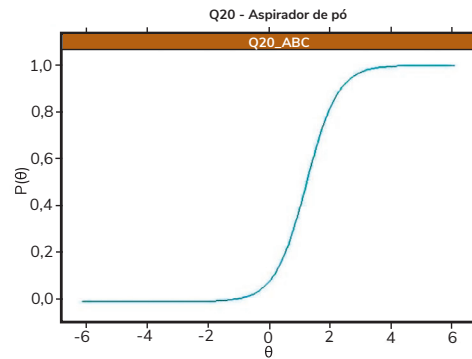
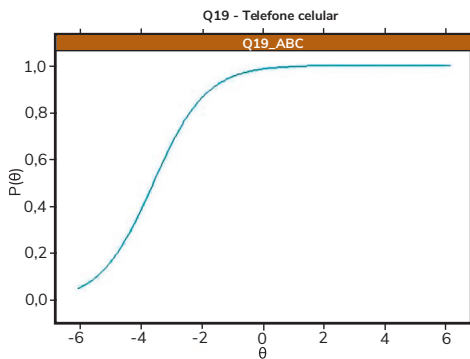
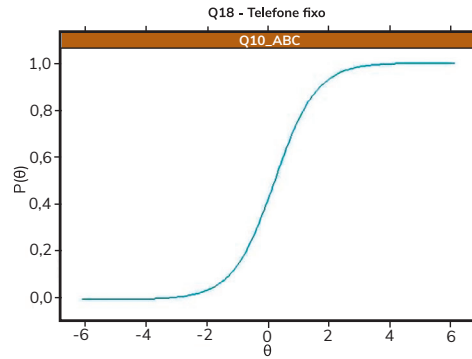
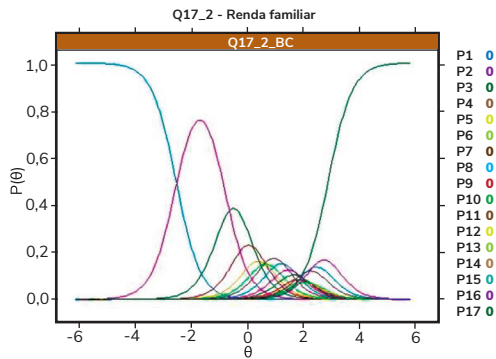


## APÊNDICE A – CURVAS CARACTERÍSTICAS DOS ITENS, POR QUESTÃO – ENEM 2011-2017









## APÊNDICE B – CLASSIFICAÇÃO DOS BENS DE CONSUMO E CONTRATAÇÃO DE SERVIÇOS

Para auxiliar na interpretação da escala do nível socioeconômico, classificamos empiricamente os bens de consumo e a contratação de serviço de empregada doméstica entre **bens elementares**, **complementares** e **suplementares**, com base na probabilidade de posse ou contratação deles em diferentes níveis da escala, tomando-se como referência probabilidades iguais ou superiores a 0,60. Seguem abaixo os critérios de classificação:

- Bens elementares: aqueles que se tornam frequentes entre indivíduos dos Grupos 1 e 2 do NSE – no geral, indivíduos de renda familiar de até um salário-mínimo e com pais que estudaram, no máximo, até os anos iniciais do Ensino Fundamental;
- Bens complementares: aqueles que se tornam frequentes entre indivíduos dos Grupos 3, 4 e 5 do NSE – no geral, contêm indivíduos de renda familiar de um a cinco salários-mínimos e com pais que estudaram, no máximo, até o Ensino Médio;
- Bens suplementares: aqueles que se tornam frequentes entre indivíduos dos Grupos 6, 7 e 8 do NSE – de modo geral, indivíduos de renda familiar superior a cinco salários-mínimos e com pais que estudaram até a graduação ou pós-graduação.

Esses critérios ensejaram a classificação exposta no Quadro 5.

**QUADRO 1**  
**CLASSIFICAÇÃO DOS BENS DE CONSUMO E CONTRATAÇÃO DE SERVIÇOS**

Elementares	Complementares	Suplementares
Televisão a cores	TV por assinatura	Freezer
Rádio	Aparelho de DVD	Máquina de lavar louça
Geladeira	Forno micro-ondas	Máquina de secar roupa
Quartos para dormir	Máquina de lavar roupa	Aspirador de pó
Banheiro	Carro	Contratação de empregada doméstica
Telefone celular	Computador	
	Acesso à internet	
	Telefone fixo	

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

## APÊNDICE C – DISTRIBUIÇÃO DAS ALTERNATIVAS DE CADA ITEM DO NSE PELOS NÍVEIS SEGMENTADOS DA ESCALA

TABELA 1

DISTRIBUIÇÃO DAS ALTERNATIVAS DE CADA ITEM DO NSE PELOS NÍVEIS SEGMENTADOS DA ESCALA – ENEM 2011-2017

(continua)

Itens do Inse	Níveis da Escala							
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8
<b>Q01 - TV em cores</b>								
1 - Não.	0,77	0,27	0,04	0,00	0,00	0,00	0,00	0,00
2 - Sim, uma.	0,23	0,72	0,86	0,60	0,27	0,08	0,02	0,00
3 - Sim, duas.	0,00	0,01	0,09	0,33	0,48	0,35	0,15	0,03
4 - Sim, três ou mais.	0,00	0,00	0,01	0,07	0,25	0,57	0,83	0,96
<b>Q02 - TV por assinatura em casa</b>								
1 - Não.	1,00	1,00	0,95	0,76	0,43	0,15	0,05	0,01
2 - Sim.	0,00	0,00	0,05	0,24	0,57	0,85	0,95	0,99
<b>Q03 - Rádio em casa</b>								
1 - Não.	0,73	0,56	0,38	0,26	0,18	0,12	0,09	0,05
2 - Sim, um ou mais.	0,27	0,44	0,62	0,74	0,82	0,88	0,91	0,95
<b>Q04_3 - Aparelho de DVD em casa</b>								
1 - Não.	0,86	0,74	0,56	0,41	0,30	0,21	0,15	0,09
2 - Sim.	0,14	0,26	0,44	0,59	0,70	0,79	0,85	0,91
<b>Q05 - Forno micro-ondas</b>								
1 - Não.	1,00	0,97	0,75	0,33	0,10	0,02	0,01	0,00
2 - Sim, um.	0,00	0,03	0,25	0,67	0,89	0,94	0,87	0,50
3 - Sim, dois ou mais.	0,00	0,00	0,00	0,00	0,01	0,04	0,12	0,49
<b>Q06 - Máquina de lavar louça</b>								
1 - Não.	1,00	1,00	1,00	0,98	0,95	0,85	0,66	0,29
2 - Sim, uma ou mais.	0,00	0,00	0,00	0,02	0,05	0,15	0,34	0,71
<b>Q07 - Geladeira</b>								
1 - Não.	0,62	0,21	0,04	0,01	0,00	0,00	0,00	0,00
2 - Sim, uma.	0,38	0,79	0,96	0,97	0,92	0,79	0,57	0,24
3 - Sim, duas ou mais.	0,00	0,00	0,01	0,02	0,08	0,21	0,43	0,76
<b>Q08 - Freezer (independente ou segunda porta da geladeira)</b>								
1 - Não.	1,00	0,98	0,89	0,67	0,40	0,18	0,07	0,02
2 - Sim, um.	0,00	0,02	0,10	0,31	0,56	0,72	0,68	0,40
3 - Sim, dois ou mais.	0,00	0,00	0,00	0,01	0,04	0,10	0,24	0,59

(continuação)

Itens do Inse	Níveis da Escala							
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8
<b>Q09_1 - Máquina de lavar roupa</b>								
1 - Não.	0,99	0,93	0,59	0,20	0,06	0,02	0,00	0,00
2 - Sim, uma.	0,01	0,07	0,41	0,79	0,91	0,88	0,71	0,31
3 - Sim, duas ou mais.	0,00	0,00	0,00	0,01	0,03	0,10	0,28	0,69
<b>Q10 - Máquina de secar roupa (independente ou em conjunto com a máquina de lavar roupa)</b>								
1 - Não.	1,00	0,99	0,96	0,89	0,78	0,60	0,42	0,20
2 - Sim, uma ou mais.	0,00	0,01	0,04	0,11	0,22	0,40	0,58	0,80
<b>Q11_1 - Automóvel / Carro</b>								
1 - Não.	1,00	0,99	0,85	0,44	0,13	0,03	0,01	0,00
2 - Sim, um.	0,00	0,01	0,14	0,52	0,66	0,41	0,14	0,02
3 - Sim, dois	0,00	0,00	0,01	0,04	0,18	0,46	0,51	0,18
4 - Sim, três ou mais.	0,00	0,00	0,00	0,00	0,02	0,11	0,34	0,80
<b>Q13_1 - Computador / Microcomputador</b>								
1 - Não.	1,00	0,98	0,64	0,13	0,02	0,00	0,00	0,00
2 - Sim, um.	0,00	0,02	0,36	0,80	0,64	0,23	0,05	0,01
3 - Sim, dois	0,00	0,00	0,01	0,06	0,26	0,44	0,23	0,03
4 - Sim, três ou mais.	0,00	0,00	0,00	0,01	0,07	0,32	0,72	0,96
<b>Q14 - Acesso à internet</b>								
1 - Não.	1,00	0,98	0,65	0,13	0,02	0,00	0,00	0,00
2 - Sim.	0,00	0,02	0,35	0,87	0,98	1,00	1,00	1,00
<b>Q15 - Quartos para dormir</b>								
1 - Não.	0,25	0,06	0,01	0,00	0,00	0,00	0,00	0,00
2 - Sim, um.	0,62	0,54	0,23	0,07	0,03	0,01	0,00	0,00
3 - Sim, dois.	0,12	0,36	0,59	0,52	0,33	0,16	0,08	0,02
4 - Sim, três.	0,01	0,04	0,16	0,37	0,54	0,60	0,52	0,27
5 - Sim, quatro ou mais.	0,00	0,00	0,01	0,04	0,10	0,22	0,40	0,70
<b>Q16 - Banheiro</b>								
1 - Não.	0,74	0,20	0,02	0,00	0,00	0,00	0,00	0,00
2 - Sim, um.	0,26	0,80	0,93	0,70	0,33	0,09	0,02	0,00
3 - Sim, dois.	0,00	0,00	0,05	0,26	0,50	0,41	0,17	0,03
4 - Sim, três ou mais.	0,00	0,00	0,01	0,04	0,17	0,51	0,81	0,97
<b>Q17_1 - Renda mensal da família</b>								
1 - Nenhuma renda.	0,86	0,30	0,02	0,00	0,00	0,00	0,00	0,00
2 - Até 1 SM.	0,13	0,61	0,37	0,05	0,01	0,00	0,00	0,00
3 - Entre 1 e 1,5 SM.	0,00	0,07	0,37	0,21	0,05	0,01	0,00	0,00
4 - Entre 1,5 e 2 SM.	0,00	0,01	0,20	0,48	0,27	0,06	0,01	0,00

(continuação)

Itens do Inse	Níveis da Escala							
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8
<b>Q17_1 - Renda mensal da família</b>								
5 - Entre 2 e 5 SM.	0,00	0,00	0,03	0,17	<b>0,29</b>	<b>0,15</b>	0,04	0,00
6 - Entre 5 e 7 SM.	0,00	0,00	0,01	0,05	<b>0,17</b>	<b>0,17</b>	0,06	0,01
7 - Entre 7 e 10 SM.	0,00	0,00	0,00	0,02	0,11	<b>0,20</b>	0,12	0,02
8 - Entre 10 e 12 SM.	0,00	0,00	0,00	0,01	0,04	<b>0,12</b>	0,11	0,02
9 - Entre 12 e 15 SM.	0,00	0,00	0,00	0,00	0,03	0,12	<b>0,17</b>	0,05
10 - Entre 15 e 30 SM.	0,00	0,00	0,00	0,00	0,02	0,12	<b>0,29</b>	0,19
11 - Acima de 30 SM.	0,00	0,00	0,00	0,00	0,01	0,05	<b>0,20</b>	<b>0,71</b>
<b>Q17_2 - Renda mensal da família</b>								
1 - Nenhuma renda.	<b>0,91</b>	<b>0,39</b>	0,04	0,00	0,00	0,00	0,00	0,00
2 - Até 1 SM.	0,09	<b>0,57</b>	<b>0,52</b>	0,10	0,02	0,00	0,00	0,00
3 - Mais de 1 e até 1,5 SM.	0,00	0,03	<b>0,28</b>	<b>0,25</b>	0,06	0,01	0,00	0,00
4 - Mais de 1,5 e até 2 SM.	0,00	0,01	0,09	<b>0,22</b>	<b>0,10</b>	0,02	0,00	0,00
5 - Mais de 2 e até 2,5 SM.	0,00	0,00	0,04	<b>0,14</b>	<b>0,11</b>	0,03	0,01	0,00
6 - Mais de 2,5 e até 3 SM.	0,00	0,00	0,02	0,10	<b>0,13</b>	0,04	0,01	0,00
7 - Mais de 3 e até 4 SM.	0,00	0,00	0,01	0,08	<b>0,16</b>	<b>0,08</b>	0,02	0,00
8 - Mais de 4 e até 5 SM.	0,00	0,00	0,01	0,04	<b>0,13</b>	<b>0,10</b>	0,03	0,00
9 - Mais de 5 e até 6 SM.	0,00	0,00	0,00	0,02	0,09	<b>0,11</b>	0,04	0,01
10 - Mais de 6 e até 7 SM.	0,00	0,00	0,00	0,01	0,06	<b>0,10</b>	0,05	0,01
11 - Mais de 7 e até 8 SM.	0,00	0,00	0,00	0,01	0,04	<b>0,08</b>	0,05	0,01
12 - Mais de 8 e até 9 SM.	0,00	0,00	0,00	0,00	0,03	<b>0,07</b>	0,05	0,01
13 - Mais de 9 e até 10 SM.	0,00	0,00	0,00	0,00	0,02	<b>0,06</b>	0,06	0,01
14 - Mais de 10 e até 12 SM.	0,00	0,00	0,00	0,00	0,02	0,09	<b>0,11</b>	0,03
15 - Mais de 12 e até 15 SM.	0,00	0,00	0,00	0,00	0,02	0,08	<b>0,13</b>	0,05
16 - Mais de 15 e até 20 SM.	0,00	0,00	0,00	0,00	0,01	0,06	<b>0,15</b>	0,08
17 - Mais de 20 SM.	0,00	0,00	0,00	0,00	0,01	0,08	<b>0,30</b>	<b>0,80</b>
<b>Q18 - Telefone fixo</b>								
1 - Não.	<b>1,00</b>	<b>0,97</b>	<b>0,80</b>	<b>0,46</b>	0,21	0,07	0,03	0,01
2 - Sim.	0,00	0,03	0,20	<b>0,54</b>	<b>0,79</b>	<b>0,93</b>	<b>0,97</b>	<b>0,99</b>
<b>Q19 - Telefone celular</b>								
1 - Não.	<b>0,48</b>	0,17	0,04	0,01	0,00	0,00	0,00	0,00
2 - Sim, um ou mais.	<b>0,52</b>	0,83	0,96	0,99	1,00	1,00	1,00	1,00
<b>Q20 - Aspirador de pó em casa</b>								
1 - Não.	<b>1,00</b>	<b>1,00</b>	<b>0,98</b>	<b>0,85</b>	<b>0,55</b>	0,21	0,06	0,01
2 - Sim.	0,00	0,00	0,02	0,15	<b>0,45</b>	<b>0,79</b>	<b>0,94</b>	<b>0,99</b>

(conclusão)

Itens do Inse	Níveis da Escala							
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8
<b>Q21_2 - Contrata empregada doméstica</b>								
1 - Não.	1,00	1,00	1,00	0,98	0,86	0,53	0,19	0,03
2 - Sim, uma ou mais.	0,00	0,00	0,00	0,02	0,14	0,47	0,81	0,97
<b>Q21_3 - Contrata empregada doméstica</b>								
1 - Não.	1,00	1,00	0,99	0,96	0,84	0,58	0,29	0,07
2 - Sim, um ou dois dias por semana.	0,00	0,00	0,01	0,03	0,09	0,21	0,24	0,10
3 - Sim, três ou quatro dias por semana.	0,00	0,00	0,00	0,00	0,02	0,05	0,08	0,05
4 - Sim, pelo menos cinco dias por semana.	0,00	0,00	0,00	0,01	0,04	0,15	0,38	0,77
<b>Q26_1 – Maior escolaridade do Pai ou Mãe</b>								
1 - Não estudou.	0,56	0,22	0,06	0,02	0,01	0,00	0,00	0,00
2 - Da 1ª à 4ª série do EF (antigo primário).	0,37	0,53	0,34	0,14	0,06	0,02	0,01	0,00
3 - Da 5ª à 8ª série do EF (antigo ginásio).	0,04	0,14	0,23	0,17	0,09	0,04	0,02	0,01
4 - EM (antigo 2º grau) incompleto.	0,01	0,03	0,07	0,07	0,05	0,02	0,01	0,00
5 - EM (antigo 2º grau).	0,01	0,06	0,21	0,35	0,33	0,22	0,13	0,04
6 - Ensino Superior incompleto.	0,00	0,01	0,02	0,06	0,08	0,08	0,05	0,02
7 - Ensino Superior.	0,00	0,01	0,05	0,13	0,24	0,32	0,30	0,17
8 - Pós-graduação.	0,00	0,00	0,02	0,06	0,13	0,28	0,47	0,75
<b>Q26_2 - Maior escolaridade do Pai ou Mãe</b>								
1 - Nunca estudou.	0,48	0,17	0,04	0,01	0,01	0,00	0,00	0,00
2 - Não completou a 4ª série/5º ano do EF.	0,40	0,46	0,24	0,09	0,04	0,02	0,01	0,00
3 - Completou a 4ª série/5º ano, mas não completou a 8ª série/9º ano do EF.	0,06	0,16	0,19	0,10	0,05	0,02	0,01	0,00
4 - Completou a 8ª série/9º ano do EF, mas não completou o EM.	0,03	0,09	0,17	0,14	0,08	0,04	0,02	0,01
5 - Completou o EM, mas não completou a Faculdade.	0,02	0,10	0,29	0,44	0,41	0,29	0,17	0,06
6 - Completou a Faculdade, mas não completou a Pós-graduação.	0,00	0,01	0,05	0,14	0,23	0,28	0,24	0,12
7 - Completou a Pós-graduação.	0,00	0,01	0,03	0,08	0,18	0,36	0,55	0,80

Fonte: Elaboração própria com base em dados do Enem 2011-2017.

Nota: Estão destacados em verde valores iguais ou superiores a 0,50; em amarelo, valores iguais ou superiores a 0,30.





(CC) BY-NC

VENDA PROIBIDA

