



UNIVERSIDADE DE BRASÍLIA  
INSTITUTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO EM CIÊNCIAS

Jailton Correia Fraga Junior

**Desigualdades sociais e excelência na educação científica: uma busca por escolas que fazem a diferença a partir do Exame Nacional do Ensino Médio**

Brasília  
2023

Jailton Correia Fraga Junior

**Desigualdades sociais e excelência na educação científica: uma busca por escolas que fazem a diferença a partir do Exame Nacional do Ensino Médio**

Dissertação submetida ao Programa de Pós-Graduação em Educação em Ciências da Universidade de Brasília como requisito parcial para a obtenção do título de Mestre em Educação em Ciências

Orientador: Prof. Paulo Lima Junior, Doutor.

Brasília

2023

Ficha catalográfica elaborada automaticamente, com  
os dados fornecidos pelo(a) autor(a)

Cd      Correia Fraga Junior, Jailton  
         Desigualdades sociais e excelência na educação científica:  
         uma busca por escolas que fazem a diferença a partir do  
         Exame Nacional do Ensino Médio / Jailton Correia Fraga  
         Junior; orientador Paulo Roberto Menezes Lima Junior. --  
         Brasília, 2023.  
         95 p.

         Dissertação (Mestrado em Educação em Ciências) --  
         Universidade de Brasília, 2023.

         1. ENEM. 2. Bourdieu. 3. Sociologia da Educação. 4.  
         Eficácia Escolar. 5. Ciências da Natureza. I. Roberto  
         Menezes Lima Junior, Paulo, orient. II. Título.

Jailton Correia Fraga Junior

**Desigualdades sociais e excelência na educação científica: uma busca por escolas  
que fazem a diferença a partir do Exame Nacional do Ensino Médio**

O presente trabalho em nível de Mestrado foi avaliado e aprovado, em 22 de agosto de 2023, pela banca examinadora composta pelos seguintes membros:

Prof. Paulo Lima Junior, Dr.

Universidade de Brasília

Prof. Samuel Molina Schnorr, Dr.

Universidade de Brasília

Prof.a Maria Alice Nogueira, Dr.a

Universidade Federal de Minas Gerais

Esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Educação em Ciências conforme processo SEI 23106.093190/2023-12.

Brasília, 2023.

Aos meus pais Jailton e Licia, à minha irmã Sarah e à minha noiva Júlia.

## **AGRADECIMENTOS**

Gostaria de agradecer, primeiramente, a minha família pela criação e exemplo que me deram ao longo de minha criação. Se hoje cheguei até aqui, muito se deve aos esforços que vocês empreenderam em minha vida e em minha educação. Por esse e diversos outros motivos, meu muito obrigado a vocês: meu pai, Jailton Correia Fraga, minha mãe, Licia Helena Cavalcante Faria, e minha irmã, Sarah Faria Fraga.

Falando em família, não poderia deixar de agradecer também a minha noiva Júlia Amaral e Silva, quem esteve comigo desde meus estudos na Licenciatura em Física e segue ao meu lado em cada conquista, em cada degrau subido. Muito obrigado pelo seu apoio, pelo seu conforto, pelo seu carinho e pelo seu amor. Sem dúvidas, essa conquista tem suas mãos. Registro aqui meus mais sinceros agradecimentos, repletos de amor e admiração que tenho pela sua pessoa. Muito obrigado!

Gostaria de agradecer também aos amigos que fiz ao longo de minha vida e no meu curso de Mestrado. Obrigado pelas conversas, pelas trocas de experiência e pelo apoio ao longo do curso. Também gostaria de agradecer aos professores do Programa de Pós-Graduação em Educação em Ciências que passaram, de alguma forma, pela minha trajetória acadêmica no Mestrado. Não poderia me esquecer, é claro, das incríveis pessoas que compõem o grupo de pesquisa Perspectivas Críticas. Obrigado a cada um de vocês pelas conversas, pelos conselhos, pelo carinho e pela amizade.

Um agradecimento especial aos professores Maria Alice Nogueira e Samuel Molina Schnorr pelas ótimas contribuições ao meu trabalho, seja na Qualificação, seja na Defesa. Vou carregar seus ensinamentos ao longo de minha vida e carreira acadêmica. Um agradecimento especial também ao meu orientador, professor Paulo Lima Junior, com quem venho trabalhando e pesquisando há muito tempo. Seus ensinamentos, suas vivências e sua carreira acadêmica me inspiram e me motivam a ser um pesquisador e intelectual melhor. Essa conquista também se deveu muito ao seu trabalho. Muito obrigado!

## RESUMO

Desde meados do século XX, sistemas escolares de todo o mundo têm detectado que o desempenho escolar de seus estudantes é menor entre famílias socialmente vulneráveis e menos escolarizadas. Esse fato tem fundamentado diversas reflexões críticas sobre as contribuições da escola para a reprodução das desigualdades sociais. Nesta dissertação, investigamos o poder preditivo do nível socioeconômico (NSE) no desempenho de estudantes Brasileiros no Exame Nacional do Ensino Médio (ENEM) com especial atenção à área de Ciências da Natureza. O principal objetivo dessa investigação não é corroborar o resultado já bastante conhecido (de que há uma associação entre desempenho escolar e origem social), mas identificar as escolas e Unidades da Federação que mais fazem a diferença no desempenho dos seus estudantes. Em outras palavras, procuramos pelos casos concretos nos quais a contribuição da escola ao desempenho dos estudantes supera o que poderíamos razoavelmente esperar a partir de seu nível socioeconômico. Baseados na Sociologia da Educação de Pierre Bourdieu e em estudos sobre eficácia escolar, lançamos um olhar crítico sobre os *rankings* escolares e seu papel na luta que as instituições de ensino travam por reconhecimento. De fato, lideram esses rankings justamente as escolas e Unidades da Federação que atendem às famílias mais ricas e escolarizadas. No entanto, ao controlar os efeitos da origem social sobre os *rankings* escolares, mostramos que novas formas de classificar as escolas são possíveis. Tomando como ponto de partida as edições de 2015 a 2019 do ENEM, foi realizada uma Análise de Correspondência Múltipla Ajustada das informações socioeconômicas dos participantes. Essa análise permitiu determinar precisamente o NSE de todos os participantes do exame. Em seguida, o NSE e a dependência administrativa foram inseridos como variáveis explicativas em um modelo de Regressão Linear Multinível. A nota na prova de Ciências da Natureza foi explicada em 89% pelas variáveis do modelo. Apenas 5% do desempenho dos estudantes nessa área de conhecimento pode ser atribuída às escolas. No caso do ENEM como um todo, esses valores são de 88% e 8%, respectivamente. Portanto, escolas que fazem a diferença são um fenômeno raro no sistema educacional brasileiro, mas elas existem e podem ser identificadas. Controlando o efeito da origem social sobre o desempenho escolar, observamos que algumas escolas obtiveram desempenho muito acima do que era esperado nas provas de Ciências da Natureza e no ENEM como um todo. Com relação às regiões do país, a região Nordeste apresentou alguns estados que despontaram quando o efeito do NSE sobre o desempenho dos estudantes foi controlado. Esperamos, com essa pesquisa, mostrar que a maneira vigente de ranquear as escolas tende a privilegiar e celebrar instituições e regiões onde está localizada a classe dominante.

**Palavras-chave:** ENEM; Bourdieu; Sociologia da Educação; Eficácia Escolar; Ciências da Natureza.

## ABSTRACT

Since the mid-twentieth century, school systems around the world have detected that the academic performance of their students is lower among socially vulnerable and less educated families. This fact has informed several critical reflections on the contributions of schools to the reproduction of social inequalities. In this dissertation, we investigate the predictive power of socioeconomic status (SES) on the performance of Brazilian students in the National High School Exam (ENEM) with special attention to the area of Natural Sciences. The main objective of this investigation is not to corroborate the already well-known result (that there is an association between school performance and social origin), but to identify the schools and Federation Units that make a difference in the performance of their students. In other words, we are looking for concrete cases in which the school's contribution to student performance exceeds what we could reasonably expect from their socioeconomic status. Drawing on Pierre Bourdieu's Sociology of Education and studies on school effectiveness, we take a critical look at school rankings and their role in the struggle that educational institutions wage for recognition. In fact, these rankings are led by the schools and states where the wealthiest and most educated families are located. However, by controlling for the effects of social origin on school rankings, we show that new ways of ranking schools are possible. Taking the 2015-2019 editions of ENEM as a starting point, an Adjusted Multiple Correspondence Analysis of participants' socioeconomic information was performed. This analysis allowed to precisely determine the SES of all participants of the exam. Then, the SES and administrative dependence were inserted as explanatory variables in a Multilevel Linear Regression model. The score in the Natural Sciences test was explained by 89% of the variables in the model. Only 5% of the students' performance in Natural Sciences test can be attributed to the schools. In the case of ENEM as a whole, these values are 88% and 8%, respectively. Therefore, schools that make a difference are a rare phenomenon in the Brazilian education system, but they do exist and can be identified. Controlling for the effect of social origin on school performance, we observed that some schools performed well above what was expected in the Natural Sciences test and in ENEM as a whole. Regarding the regions of the country, the Northeast region presented some states that stood out when the effect of SES on student performance was controlled. We hope, with this research, to show that the current way of ranking schools tends to privilege and celebrate institutions and regions where the dominant class is located.

**Keywords:** ENEM; Bourdieu; Sociology of Education; School Effectiveness; Natural Sciences.

## LISTA DE ABREVIATURAS E SIGLAS

ACM	Análise de Correspondência Múltipla
ACP	Análise de Componentes Principais
AL	Alagoas
BA	Bahia
CH	Ciências Humanas
CN	Ciências da Natureza
EMI	Effectively Maintained Inequality
ENEM	Exame Nacional do Ensino Médio
FIES	Fundo de Financiamento Estudantil
IDH	Índice de Desenvolvimento Humano
IF	Instituto Federal
IFES	Instituições Federais de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
LC	Linguagens e Códigos
LDB	Lei de Diretrizes e Bases da Educação Nacional
MLH	Modelagem Linear Hierárquica
MMI	Maximally Maintained Inequality
MT	Matemática
NSE	Nível Socioeconômico
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
PCNEM	Parâmetros Curriculares Nacionais do Ensino Médio
PE	Pernambuco
PROUNI	Programa Universidade para Todos
RD	Redação
RS	Rio Grande do Sul
TRI	Teoria de Resposta ao Item
UF	Unidade Federativa
UNESCO	Organização das Nações Unidas para Educação, a Ciência e a Cultura

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
1.1	TRANSFORMAÇÕES DA EDUCAÇÃO BRASILEIRA	14
1.2	ENEM: HISTÓRIA E TRANSFORMAÇÕES	16
1.3	OBJETIVOS DA DISSERTAÇÃO	18
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>21</b>
2.1	A ESCOLA NÃO FAZ A DIFERENÇA	21
2.2	ESCOLAS QUE FAZEM A DIFERENÇA	30
2.3	O QUE É FAZER A DIFERENÇA?	35
<b>3</b>	<b>REVISÃO DA LITERATURA</b>	<b>40</b>
3.1	DISTRIBUIÇÃO DOS ARTIGOS AO LONGO DOS ANOS ANALISADOS	41
3.2	FREQUÊNCIA E ASSOCIAÇÃO ENTRE AS PALAVRAS	42
3.3	ENEM E QUESTÕES SOCIOECONÔMICAS	47
3.4	SÍNTESE DA REVISÃO	56
<b>4</b>	<b>MÉTODOS</b>	<b>58</b>
4.1	ANÁLISE DE CORRESPONDÊNCIA	59
<b>5</b>	<b>RESULTADOS</b>	<b>71</b>
5.1	DESEMPENHO MÉDIO POR ESCOLA E POR UNIDADE DA FEDERAÇÃO NO ENEM	71
5.2	RANKINGS ESCOLARES DE CIÊNCIAS DA NATUREZA E DO ENEM	73
5.3	MODELO DE REGRESSÃO LINEAR MULTINÍVEL	75
5.4	RANKINGS ESTADUAIS E ESTADOS QUE FAZEM A DIFERENÇA	81
5.5	SÍNTESE E DISCUSSÃO DOS RESULTADOS	84
<b>6</b>	<b>CONCLUSÃO</b>	<b>87</b>

## 1 INTRODUÇÃO

A construção de indicadores educacionais a partir de exames padronizados vem sendo realizado e financiado mundialmente. No Brasil, esse movimento está ancorado na reforma neoliberal da educação e data da década de 1990. No nosso contexto nacional, o Exame Nacional do Ensino Médio (ENEM) tem sido utilizado para ranquear escolas e estados a partir do desempenho dos participantes. O ENEM foi criado em 1998 para avaliar o estudante ao término do Ensino Médio e, a partir de sua reestruturação em 2009, passou a apresentar a possibilidade de ser usado como porta de entrada (vestibular) para o Ensino Superior (INEP, 2020). Ainda que o ENEM pretenda democratizar o acesso ao Ensino Superior (MALUSÁ; ORDONES; RIBEIRO, 2015), os desempenhos obtidos nessa prova carregam influências de privilégios de classe (LIMA JUNIOR; FRAGA JUNIOR, 2021; KLEINKE, 2017; NASCIMENTO; CAVALCANTI; OSTERMANN, 2018), condicionando a entrada de uma parcela da população aos cursos universitários (GONÇALVES; RAMOS, 2019). Portanto, devemos olhar com suspeita para os *rankings* escolares que classificam as boas escolas com base no desempenho nesse certame.

Dada a influência da origem social em avaliações escolares, pesquisadores recomendam que não sejam utilizados os dados brutos para ranquear escolas, pois essa forma de classificar está assentada sobre uma premissa de que todo o desempenho alcançado por um estudante decorre apenas da escola (SOARES; ALVES; OLIVEIRA, 2001). Para contornar esse problema, investigadores vêm empregando técnicas estatísticas que consigam controlar efeitos da origem social dos estudantes e efeitos de composição social da escola (TRAVITZKI, 2013), com o objetivo de aferir o quanto cada escola efetivamente agrega aos seus alunos. Essas pesquisas podem ser classificadas como pesquisas do efeito-escola e baseiam-se na teoria do valor agregado (SOARES; ALVES; OLIVEIRA, 2001). Com o auxílio dessas técnicas estatísticas, podemos, portanto, olhar para o desempenho alcançados pelas escolas e pelas Unidades da Federação que se destacaram quando são retirados os efeitos de origem social do modelo.

Para localizarmos o discurso por trás da criação de indicadores educacionais, bem como o atrelamento da qualidade escolar ao desempenho em provas padronizadas, apresentamos, nas seções seguintes, uma breve história das

transformações da educação brasileira e do ENEM. Na última seção deste capítulo, trazemos os objetivos dessa pesquisa.

## 1.1 TRANSFORMAÇÕES DA EDUCAÇÃO BRASILEIRA

No Brasil, a relação entre desigualdades sociais e desigualdades escolares passou por transformações ao longo dos anos. Durante o período colonial, a educação era destinada apenas à classe dominante, visando a transmissão de costumes e valores para aqueles que dispunham de tempo para ser cultivados (KUJAWA; MARTINS; PATIAS, 2020). Esse modelo de educação jesuítica tradicional permeou a educação brasileira até o final da década de 1920 (LAUTÉRIO; NEHRING, 2012). Com isso, vemos que as desigualdades estavam presentes no acesso à educação, onde grande parcela da sociedade era analfabeta e se via excluída da escola (BARROSO FILHO, 2000).

Essa exclusão de pessoas menos favorecidas da escola também marcou o período imperial do Brasil. Com isso, a educação continuava a ser destinada apenas aos filhos oriundos da classe dominante, que contavam com aulas particulares sobre valores morais e conhecimentos aritméticos e linguísticos (KUJAWA; MARTINS; PATIAS, 2020). Um dos motivos que pautava a justificativa dessa exclusão da classe popular à educação era o pensamento de que essas pessoas seriam incapazes de ser civilizadas (GOUVÊA, 2007).

No período republicano, com o advento de políticas voltadas para a industrialização do país, por volta da década de 1930, houve uma preocupação com a instrução dos trabalhadores para formação de mão-de-obra técnica qualificada para assumir determinados postos de trabalho. Portanto, algumas pessoas da classe popular começaram a ser escolarizadas para o desenvolvimento e progresso do Brasil. Essa proposta tinha por projeto o uso da educação para exploração capitalista (KUJAWA; MARTINS; PATIAS, 2020). A partir da década de 1930, o tema educação ganha mais espaço nas discussões políticas, culminando em uma maior preocupação com a educação pública (GADOTTI, 1993).

Em 1961, é publicada a primeira Lei de Diretrizes e Bases da Educação Nacional (LDB) com a preocupação de educar todos os brasileiros (LAUTÉRIO; NEHRING, 2012). Nela, encontra-se a afirmação do direito à educação pública para todos. No entanto, ainda que a educação fosse pública, a obrigatoriedade abarcava apenas os 4 anos iniciais do primário, o suficiente para iniciar o processo de

alfabetização (*ibid*). Com a promulgação da LDB/61, a classe popular passou a ter acesso, por direito, à escola, ainda que por um período curto de tempo. Se, por um lado, o ensino primário teve seu acesso universalizado, por outro lado, as formas de exclusão foram transladas para o ensino secundário, cujo acesso, na década de 50, estava disponível a aproximadamente 1% da população (CASTRO; TIEZZI, 2002).

O crescimento do número de matrículas no Ensino Médio ocorre a partir de mudanças decorrentes da promulgação da segunda LDB, em 1971. Nela, a educação obrigatória passa a ser de 8 anos e o exame de passagem entre o primário e o secundário é eliminado (CASTRO; TIEZZI, 2002). Diante dessas mudanças, as matrículas no Ensino Médio de 1971 para 1980 mais que duplicam, saindo de cerca de 1,1 milhão para 2,8 milhões, respectivamente (*ibid*). Entretanto, ainda que tenha crescido o número de matrículas no Ensino Médio, as taxas de repetência eram elevadas. Nesse contexto da ditadura militar, ainda que tenha ocorrido uma grande expansão da educação básica, ela não foi acompanhada de investimentos em educação, contribuindo para um processo de precarização. Essa expansão tinha por objetivo a formação de recursos humanos (BARROSO FILHO, 2000). Ainda que houvesse mais oportunidades de acesso à escola, para a classe popular, terminar o Ensino Fundamental ainda era um desafio, onde poucos alcançavam o sucesso. Ou seja, o ensino secundário continuava a ser destinado a poucas pessoas (*ibid*). Com todas essas mudanças no período de escolarização obrigatória e acesso à educação dos meios populares, as desigualdades escolares evoluem da exclusão *da* escola para a exclusão *na* escola (KUJAWA; MARTINS; PATIAS, 2020).

Durante a década de 90, a educação brasileira sofre outra grande expansão, atingindo tanto o Ensino Fundamental, quanto o Ensino Médio (CASTRO; TIEZZI, 2002). Nesse período, os ideais neoliberais inspiraram uma educação voltada ao capitalismo (KUJAWA; MARTINS; PATIAS, 2020) e uma nova forma de geri-la: uma gestão empresarial (FREITAS, 2018). Nesse contexto, o foco da educação era a eficiência e a produtividade (KAJUWA; MARTINS; PATIAS, 2020), alinhado à lógica mercantilista (LAVAL, 2004). Dessa forma, entra em voga uma preocupação maior com a qualidade da educação ofertada e os meios para aferi-la.

Nessa lógica, está também o controle de gastos públicos com a educação. Para isso, algumas medidas são tomadas para conter os custos (como da evasão e da repetência), como políticas de correção de fluxo e classes de aceleração, alinhadas a uma visão de custo/benefício (FREITAS, 2002). Consequentemente, o Brasil figura-

se entre os países que menos investe em educação. Essa realidade pode ser visualizada ao comparar o salário inicial pago aos professores dos últimos anos da educação básica ao redor do mundo e no Brasil (LIMA JUNIOR; FRAGA JUNIOR, 2021). Com isso, temos um fenômeno interno à escola que altera seu funcionamento e passa a adiar a eliminação de estudantes das camadas populares. Ou seja, os alunos permanecem na escola, ainda que não tenham aprendido, e são eliminados posteriormente, em momento oportuno (FREITAS, 2007). Conseqüentemente, esse fenômeno faz com que:

O sistema de ensino, amplamente aberto a todos e, no entanto, estritamente reservado a alguns, consiga a façanha de reunir as aparências da 'democratização' com a realidade da reprodução que se realiza em um grau superior de dissimulação, portanto, com um efeito acentuado de legitimação social (BOURDIEU, 1998, p. 223).

Essas transformações que as desigualdades escolares sofreram continuam até os exames que dão acesso ao Ensino Superior. No caso do ENEM, mais de um terço do desempenho dos estudantes na prova de Ciências da Natureza pode ser determinado pela renda familiar e escolaridade dos pais (LIMA JUNIOR; FRAGA JUNIOR, 2021), afetando suas chances objetivas de ingresso no Ensino Superior. Já nos cursos de graduação, a probabilidade de conclusão de um curso também é sensível a privilégios sociais da família dos estudantes (KNOP; COLLARES, 2019).

## 1.2 ENEM: HISTÓRIA E TRANSFORMAÇÕES

A década de 90 trouxe grandes mudanças para a educação brasileira (CASTRO; TIEZZI, 2002). Nesse contexto, vemos uma expansão do Ensino Médio, ao mesmo tempo em que são publicados documentos normativos (como os Parâmetros Curriculares Nacionais do Ensino Médio – PCNEM) para delimitação de seu conteúdo e a forma de abordar o conhecimento nessa etapa. Além disso, a partir da publicação da LDB/96, o Ensino Médio começa a compor, juntamente com a Educação Infantil e o Ensino Fundamental, a educação escolar obrigatória no Brasil (LAUTÉRIO; NEHRING, 2012).

Os PCNEM foram elaborados a partir de princípios da Organização das Nações Unidas para Educação, a Ciência e a Cultura – UNESCO (LAUTÉRIO; NEHRING, 2012). A partir desses documentos normativos, vemos que os objetivos para a educação eram de se pautar menos numa educação mnemônica e trabalhar mais a contextualização e a interdisciplinaridade do conhecimento (MALUSÁ;

ORDONES; RIBEIRO, 2015). Nesse cenário, em 1998, cria-se o Exame Nacional do Ensino Médio (ENEM) (VIGGIANO; MATTOS, 2013), sob responsabilidade do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), com o objetivo de avaliar o desempenho do estudante ao fim do Ensino Médio, aferindo o desenvolvimento de competências e habilidades consideradas necessárias para o exercício da cidadania (ANDRIOLA, 2011). Além disso, o ENEM também foi criado como uma alternativa para substituir ou complementar as formas de ingresso no Ensino Superior (VIGGIANO; MATTOS, 2013).

Inicialmente, o ENEM era uma prova única, composta por 63 itens interdisciplinares e uma redação. Esses itens eram baseados em 5 competências e 21 habilidades, não sendo divididos por disciplinas, como os vestibulares tradicionais da época (CASTRO; TIEZZI, 2002). Em sua primeira edição, contou apenas com 115 mil inscrições, um baixo número de adesão se consideramos as 7 milhões de matrículas, aproximadamente, no Ensino Médio (*ibid*). No entanto, no princípio, o exame era pago, tornando-se gratuito para alunos oriundos de escolas públicas a partir de 2001. Com essa isenção da taxa de inscrição, os números de alunos aderindo ao certame cresceu significativamente, alcançando 1,2 milhão de inscritos presentes no mesmo ano (*ibid*).

Em 2005, é criado o Programa Universidade para Todos (PROUNI) (MALUSÁ; ORDONES; RIBEIRO, 2015). Com isso, estudantes que comprovassem carência e alcançassem bons desempenhos no ENEM poderiam concorrer a bolsas integrais e parciais de estudo em universidades privadas do Brasil (LAUTÉRIO; NEHRING, 2012). Apesar dessas diferenças na maneira de avaliar e as possibilidades de uso da nota no certame, o ENEM não conseguiu firmar-se como um substituto ao vestibular tradicional (LAUTÉRIO; NEHRING, 2012).

Em 2009, um novo ENEM é proposto, visando democratizar o acesso ao Ensino Superior e induzir a reestruturação dos currículos do Ensino Médio (MALUSÁ; ORDONES; RIBEIRO, 2015). Além disso, o exame passa a valer também como certificação de conclusão do Ensino Médio (INEP, 2020). Percebe-se uma lógica de como o ENEM pode cumprir o papel de regulador do cidadão que se espera ao final da educação básica (LOPES; LÓPEZ, 2010). Nesse mesmo período, foi criado o Sistema de Seleção Unificada (Sisu), com o objetivo de auxiliar no acesso aos cursos universitários das Instituições Federais de Ensino Superior (IFES) (MALUSÁ;

ORDONES; RIBEIRO, 2015). Diante dessas mudanças, o ENEM assume, também, um caráter avaliativo (LAUTÉRIO; NEHRING, 2012).

As principais mudanças entre o Novo ENEM (que entra em vigor a partir de 2009) e o antigo ENEM dizem respeito à mudança na estrutura da prova. Hoje, o certame é composto de 2 provas e uma redação aplicadas em 2 dias. Os conhecimentos passam a ser avaliados a partir de matrizes para cada área do conhecimento (MALUSÁ; ORDONES; RIBEIRO, 2015). São elas: Linguagens, Códigos e suas Tecnologias; Matemática e suas Tecnologias; Ciências da Natureza e suas Tecnologias; e Ciências Humanas e suas Tecnologias, totalizando 180 questões objetivas (45 questões por área). Além disso, a forma de corrigir o exame passou a adotar a Teoria de Resposta ao Item (TRI), possibilitando a comparação de desempenho entre edições, já que os itens são formulados para medir competências dos candidatos (VIGGIANO; MATTOS, 2013).

Outros marcos importantes do ENEM são: em 2010, é utilizado como pré-requisito para o Fundo de Financiamento Estudantil (Fies); em 2013, passa a valer como critério para seleção no programa Ciências sem Fronteiras; em 2014, algumas universidades portuguesas passam a aceitá-lo como porta de acesso para o Ensino Superior português; em 2020, passa a ser ofertado em uma nova modalidade: o ENEM Digital, aplicado em lugares específicos definidos pelo INEP (INEP, 2020).

### 1.3 OBJETIVOS DA DISSERTAÇÃO

Tendo em vista a grande desigualdade social que estrutura nossa sociedade e suas instituições, podemos colocar em dúvida o emprego dos *rankings* que, feitos a partir de notas no ENEM, são apresentados publicamente como se fossem indicadores confiáveis de qualidade escolar. Uma vez que os desempenhos obtidos na prova de Ciências da Natureza do ENEM carregam influências da origem social dos candidatos (LIMA JUNIOR; FRAGA JUNIOR, 2021), a informação comunicada no *ranking* contribui para dissimular a contribuição escolar para a reprodução das desigualdades sociais.

Pensando nisso, esta dissertação trata de sugerir uma forma alternativa de pensar o ranqueamento: uma classificação que tente subtrair o efeito da composição social da escola no desempenho em Ciências da Natureza e no ENEM, exaltando as instituições e Unidades da Federação que fazem a diferença, dada a origem social média de sua população. Essa forma alternativa não pretende nem poderia ser uma

solução, em última instância, para a avaliação da qualidade da educação, em geral, e da educação em Ciências, em particular. Afinal, dado o caráter performativo de toda avaliação escolar, a qualidade educacional é uma ficção. Ela não está dada na natureza das coisas, mas é construída e definida a partir das lutas e disputas entre os agentes presentes no campo educacional. O ranqueamento alternativo proposto aqui tem por objetivo problematizar a maneira dominante de pensar as diferenças interescolares. Para isso, é apresentada uma proposta viável que explicita os interesses objetivamente atendidos pela proposta vigente.

Diante de tudo isso, esta dissertação pretende responder à seguinte questão de pesquisa:

- *Controlando o efeito da origem social sobre o ENEM, quais escolas e Unidades da Federação fazem a diferença na educação em Ciências no Brasil?*

Nesta dissertação, designam-se como *escolas que fazem a diferença* aquelas cujo desempenho médio dos seus estudantes está muito acima do esperado dada sua origem social. Da mesma forma, os *estados que fazem a diferença* são aqueles que se sobressaem acima do esperado pela posição social de seus habitantes.

Baseando-nos nas contribuições de Pierre Bourdieu (1998, 2007, 2020) à sociologia da educação, e com o auxílio de métodos estatísticos de Análise de Correspondência Múltipla (GREENACRE, 2007) e de Regressão Linear Multinível (HOX, 2010), investigamos, em todo o território nacional, quais escolas e quais Unidades da Federação fizeram a diferença na prova de Ciências da Natureza, em específico, e no ENEM, como um todo, no período de 2015 a 2019. Esperamos, com essa discussão, contribuir para um olhar alternativo sobre o uso dos *rankings* baseados no ENEM e sobre a qualidade do ensino de Ciências.

Esta dissertação de mestrado está composta da seguinte maneira: no primeiro capítulo, o problema de pesquisa é introduzido e contextualizado. Aqui, foram elencadas algumas transformações pelas quais a educação brasileira atravessou ao longo dos anos, bem como uma breve história do ENEM. No segundo capítulo, apresentamos o referencial teórico que orienta essa pesquisa, bem como uma discussão sobre o que é fazer a diferença. No terceiro capítulo, são apresentados: (i) uma revisão de artigos cujo título contém a palavra ENEM (e suas variações) com auxílio da análise estatística de mineração de texto; e (ii) os resultados de artigos que tratam de questões socioeconômicas e ENEM. No quarto capítulo, explicamos o

método de análise aqui empregado para construção do indicador da posição social individual e média da escola a partir de variáveis socioeconômicas. Por fim, nos capítulos cinco e seis, apresentamos os resultados com suas discussões e as considerações finais, respectivamente. Esperamos, com essa pesquisa, contribuir para a Educação em Ciências no Brasil a partir de uma discussão teórica sobre o que é excelência na educação científica, além de propor uma maneira diferente de olhar para os desempenhos escolares no ENEM. Além disso, visamos apresentar o poder que as ferramentas de análise estatística, quando aplicadas aos dados educacionais, proporcionam aos pesquisadores da área.

## 2 REFERENCIAL TEÓRICO

### 2.1 A ESCOLA NÃO FAZ A DIFERENÇA

Desde a década de 1960, sabemos que a origem social dos indivíduos exerce um efeito sobre seus desempenhos escolares (COLEMAN, 1968). No entanto, esse efeito passou por transformações ao longo dos anos e assume distintas formas. Por exemplo, ao estudar a expansão do Ensino Superior irlandês, Raftery e Hout (1993) lançam mão do conceito de Desigualdade Mantida ao Máximo (do inglês *Maximally Maintained Inequality – MMI*), onde afirmam que as classes populares terão acesso aos níveis de ensino que passaram por expansão quando as classes mais abastadas já tiverem saturado esse nível. Com isso, ocorre uma translação da desigualdade de acesso ao próximo nível educacional (BOURDIEU, 1998).

Segundo o modelo da Desigualdade Efetivamente Mantida (do inglês *Effectively Maintained Inequality – EMI*), proposto por Lucas (2001), as desigualdades tendem a se manter por meio da destinação de certas instituições e cursos mais prestigiados para certa parcela da população. Essa mesma conclusão é enunciada por Bourdieu e Passeron (2014) quando, ao analisar o Ensino Superior francês, verificam que as chances objetivas de acesso a certos cursos variam de acordo com a origem social dos estudantes, onde ocorre um fenômeno de *relegação*. Isto é, cursos menos prestigiados tendem a ser ocupados, em sua maioria, por estudantes oriundos de famílias da classe popular. Além disso, as chances objetivas de conclusão de um curso superior também são sensíveis à origem social dos estudantes. No Brasil, filhos de pais mais escolarizados e com renda familiar elevada apresentam maior chance de concluir um curso de graduação (KNOP; COLLARES, 2019).

Nesta dissertação, baseamo-nos nas contribuições de Pierre Bourdieu (2007,2020) à sociologia da educação. Para compreendermos seu olhar para a contribuição da escola à reprodução das desigualdades sociais, apresentaremos os conceitos de espaço social, posição social, *habitus*, capital, investimento escolar, arbitrário cultural e atos de instituição.

#### 2.1.1 Espaço social e tipos de capital

Crítico da tradição economicista, que classifica os agentes sociais de acordo com variáveis estritamente financeiras, Bourdieu (1998) sugere que o espaço social seja concebido como um sistema dinâmico de lutas e disputas entre atores (individuais

ou institucionais) que são classificados de acordo com a posse de certos tipos de capital desigualmente distribuídos. Capital é entendido como uma fonte de poder, um recurso acumulável, uma energia social efetivamente utilizável que pode resultar em benefícios a seus possuidores (BOURDIEU, 1986, 2007). Além do mais, os tipos de capital podem ser reconvertidos (BOURDIEU, 1986), ou seja, transformados em outro tipo de recurso.

A saber, Bourdieu (1986) destaca três tipos de capital: (1) econômico, (2) social e (3) cultural. A distribuição desigual desses capitais em dado momento representa a estrutura do campo social (BOURDIEU, 1986), opondo dominantes e dominados. O capital econômico está mais diretamente relacionado ao dinheiro e todos os bens aos quais ele dá acesso. O capital social é definido como o conjunto de relações influentes acumuladas (*ibid*). Por último, o capital cultural é tudo aquilo que distingue as pessoas mais cultas de cada formação social. Ele pode incluir competências, gostos e maneiras de agir, bens culturais e certificados escolares (*ibid*). Vale lembrar que o que é considerado como capital em uma formação social não está completamente determinado, mas é objeto de luta e disputa entre os agentes sociais.

O capital cultural pode existir em três estados (BOURDIEU, 1986): (1) institucionalizado, (2) objetivado e (3) incorporado. No estado institucionalizado, o capital cultural diz respeito aos diplomas e títulos escolares acumulados por uma pessoa. No estado objetivado, ele se refere à posse de artefatos valorizados pela cultura dominante, tais como livros eruditos e obras de arte. Já o estado incorporado do capital cultural está relacionado às habilidades, gostos e maneiras de agir que contribuem para que uma pessoa seja percebida como semelhante às mais cultas de sua formação social. Podem ser considerados exemplos de capital cultural incorporado: o domínio culto da língua, a capacidade de distinguir e apreciar arte erudita, preferência por determinadas alimentos, roupas e práticas culturalmente valorizadas.

De todas as formas e estados do capital, o cultural incorporado pode operar livremente no sistema educacional mesmo quando condições educacionais muito semelhantes são garantidas a todos os cidadãos (BOURDIEU, 1998). Relacionado a essa informação, no Brasil, devido à desigualdade social que estrutura a sociedade e as instituições nela presentes, a variável renda se mostra muito influente nos destinos escolares, seja ao ingressar no Ensino Superior, seja ao concluí-lo (KLEINKE, 2017; KNOP; COLLARES, 2019; LIMA JUNIOR; FRAGA JUNIOR, 2021). Hoje, é sabido que

as vantagens culturais têm se mostrado menos rentáveis e transmissíveis que no passado. Além disso, uma condição econômica mais abastada vem se mostrando crucial para o sucesso escolar (NOGUEIRA, 2021).

Ainda que se tenha uma percepção do que está relacionado aos tipos de capital (cultural, econômico, social) e o que (não) é valorizado, essa percepção não é imutável, mas varia a depender dos contextos e das épocas. Mais especificamente, no caso do capital cultural, aquilo que distingue e, portanto, classifica os dominantes e os dominados, não é dado *a priori*. Trazendo essa reflexão para o caso de nossa pesquisa, as instituições escolares que são celebradas e que, de certa forma, tendem a distinguir seus alunos por meio dos certificados escolares, podem não ser valorizadas no futuro. Com isso, vemos que os certificados escolares carregam toda a volatilidade que as distinções de nobreza cultural carregam.

Podemos observar, então, como esse fato serve como pano de fundo para a compulsão pelos ranqueamentos e aferição de qualidade escolar no contexto das reformas neoliberais na educação. Ora, se o que distingue uma escola excelente de outra não excelente não está na natureza da escola, mas é fruto das relações de força travadas entre essas e outras escolas, os *rankings* podem ser vistos como uma ferramenta para controlar a volatilidade dos certificados escolares e institucionalizar uma diferença arbitrária, celebrando e mostrando quais escolas são excelentes e quais não são. Infelizmente, como será visto nesta pesquisa, a forma usual de ranquear (a partir de dados brutos) tende a celebrar as instituições frequentadas por filhos da classe dominante.

### **2.1.2 *Habitus* e posição social**

Ao desenvolver uma ciência das práticas, Bourdieu assume um posicionamento de rejeição tanto do olhar objetivista quanto do olhar subjetivista (BOURDIEU, 2009). Ou seja, as práticas individuais não resultam de ações rigidamente determinadas por um meio social externo, como se os indivíduos fossem marionetes da estrutura social. Tampouco os indivíduos possuem total consciência de suas práticas, como seres totalmente conscientes de suas tomadas de decisão, livres de influências externas. Superando essa dicotomia, as práticas são vistas como um produto de uma influência tanto externa, porém não mecânica; como de uma escolha pessoal, porém não totalmente consciente. Dessa forma, as práticas dos agentes

sociais são entendidas como resultado de um princípio organizador e unificador: o *habitus* (BOURDIEU, 2009).

*Habitus* é um princípio que orienta as práticas de acordo com as condições objetivas da classe à qual os indivíduos pertencem (BOURDIEU, 2009). As diferentes experiências de êxito e fracasso, sempre mais ou menos dependentes da origem social dos agentes, são incorporadas a eles, passando a orientar suas ações futuras. Segundo Bourdieu (2009), a partir da socialização em determinada posição no espaço social, o indivíduo incorporaria o *habitus* que corresponde a essa posição. Dessa forma, nossas aspirações e maneiras de agir tendem a exprimir, de maneira aproximada, nossa origem social.

O *habitus* pode ser pensado, então, como um instinto, um “senso do jogo”, que orienta as ações individuais sem jamais determiná-las. Por exemplo, no caso de famílias de elite empresarial, a escolha do estabelecimento de ensino pode se dar intuitivamente pela qualidade da clientela e por ambientes que tragam uma continuação para a educação iniciada em casa (NOGUEIRA, 2002). No caso das famílias de classe média, essa escolha pode ser auxiliada pela divulgação de desempenhos obtidos em avaliações institucionais (NOGUEIRA; RESENDE; VIANA, 2015). As diferentes tomadas de posição frente ao estabelecimento de ensino dos filhos por parte das famílias decorrem da posse desigual de recursos econômicos e culturais. Conseqüentemente, essas tomadas de posição tendem a refletir e reproduzir essas mesmas desigualdades (RESENDE; NOGUEIRA; NOGUEIRA, 2011).

A partir de análises de correspondência, Bourdieu encontra relações entre práticas, capital e posição social. Uma das características da análise de correspondência é buscar relações de associação em detrimento das relações de predição. Na análise de correspondência, os indivíduos são posicionados em relação aos demais com base em distinções referentes às dimensões do mapa. Portanto, no caso do espaço social, os agentes sociais encontram-se posicionados relativamente uns aos outros com base na distribuição desigual dos tipos de capital (dimensões da estrutura social). Nesse sentido, os indivíduos são agrupados e posicionados por (dis)similaridade entre as variáveis empregadas na construção do mapa (LIMA JUNIOR; FRAGA JUNIOR, 2021). Quanto maior a similaridade, mais próximas estão as variáveis e, conseqüentemente, as pessoas que ali estão representadas (e vice-versa). Por ser um espaço multidimensional, podemos entender que os tipos de capital

não se reduzem somente à questão econômica, mas também às questões culturais e sociais. Como exemplo da visualização desse espaço produzido a partir de análises de correspondência, apresentamos a Figura 7, onde Bourdieu distingue, no espaço, os agentes de acordo com seu volume de capital e, mais especificamente, os tipos de capital que mais possuem. Além disso, é possível observar também as práticas que estão mais próximas desses agentes, bem como a inclinação política dessas pessoas.

Figura 1 – Espaço das posições sociais e espaço dos estilos de vida



Fonte: retirado do livro Razões Práticas (BOURDIEU, 2008, p. 20).

Vale salientar que o espaço social não é anterior às variáveis que empregamos para perceber esse espaço (LIMA JUNIOR, 2023). Ou seja, as práticas culturais e escolares não são somente resultado de uma desigualdade pré-estabelecida, mas participam ativamente da construção dessas desigualdades. Sendo assim, não faz sentido perguntar se uma escola obtém bom resultado no ENEM por receber um público discente de origem social abastada ou se, por obter um bom desempenho no ENEM, uma escola passa a receber um público discente mais

abastado. Dentro dessa perspectiva, a causalidade ocorre a partir das relações recíprocas de associação das variáveis (LIMA JUNIOR, 2023).

### 2.1.3 Origem social e investimento escolar

As pessoas tendem a assumir instintivamente certas práticas que vão permitir que adquiram novos tipos de capital ou aumentem seu patrimônio. Um exemplo de prática que varia de acordo com a origem social é o investimento escolar (BOURDIEU, 2007). Estratégias de investimento escolar apresentam uma relação com as probabilidades objetivas e as esperanças subjetivas de sucesso na escola (BOURDIEU, 2009). Dessa relação, adquirimos, mesmo que inconscientemente, um senso daquilo que é possível alcançarmos e aquilo que não está ao nosso alcance. Por exemplo, uma criança que, havendo crescido em um ambiente onde os jovens e adultos estudam até o Ensino Médio e assumem profissões técnicas ou elementares, tende a interiorizar que o que está ao seu redor é o que lhe é possível. Tudo o que estiver fora de sua realidade seria carregado de uma incerteza de sucesso. No entanto, isso não impede que as pessoas adquiram disposições que não condigam com sua posição. Ou seja, o *habitus* não impossibilita que surja uma postura ativista, crítica e transformadora com relação ao destino social provável que se apresenta aos indivíduos.

Segundo Bourdieu, esse senso de êxito ou fracasso seria fruto da incorporação do *habitus* de classe. Como dito anteriormente, o *habitus* incorpora, nas pessoas, um passado acumulado. Ou seja, ele tende a incorporar as experiências das gerações anteriores, contribuindo para a manutenção da posição social de um grupo ou classe. Com isso, as experiências de familiares e pessoas próximas tendem a criar uma ideia sobre as posições até onde é possível chegar. Dito de outra forma, as pessoas adquirem um senso de suas possibilidades objetivas de ser, ter ou fazer nos mais diversos campos (BOURDIEU, 2009).

Temos aí uma relação de predisposição e predestinação: enquanto se dispõem a agir de maneira que suas chances objetivas se realizem, os agentes se predestinam ao que essas chances lhes reservam. Dessa relação dialética entre *habitus* e chances prováveis surge a *causalidade do provável*. Portanto, o futuro social não é dado, mas é fruto da interação entre o *habitus* e as chances objetivas. Consequentemente, as práticas resultam de um encontro entre um indivíduo predisposto e um mundo presumido (BOURDIEU, 1998). Essas chances objetivas

variam de acordo com a posição social e influenciam tanto o ingresso no Ensino Superior (GONÇALVES; RAMOS, 2019; LIMA JUNIOR; FRAGA JUNIOR, 2021), como sua conclusão (KNOP; COLLARES, 2019; LIMA JUNIOR; OSTERMANN; REZENDE, 2013, 2018). Além disso, há também uma variação, segundo a origem social, nas probabilidades de ingresso em cursos de maior prestígio social, sendo estes ocupados por herdeiros das famílias socialmente dominantes (BOURDIEU; PASSERON, 2014).

A partir da percepção dessas trajetórias (im)prováveis, as famílias de distintas classes sociais tendem a assumir aspirações diferentes quanto à escolarização de seus filhos. Essas aspirações estarão mais ou menos ajustadas à dependência entre o sucesso na vida e a escola (famílias cuja posição social depende mais exclusivamente da escola tenderão a investir mais pesadamente na escolarização de seus filhos). Começando pelas famílias de classe popular (classe que menos detém capital cultural e econômico), Bourdieu (2007) classifica as ações dessas pessoas como orientadas por uma “escolha do necessário”. Ou seja, as estratégias assumidas aqui priorizam o imediato, aquilo que precisam para sua sobrevivência. Diante disso, essas famílias geralmente não podem se dar ao luxo de investir em uma trajetória escolar mais prolongada de seus filhos, priorizando uma formação básica para uma rápida inserção no mercado de trabalho. Dessa forma, a tendência observada e esperada é que os filhos de origem popular estudem apenas a educação compulsória (ou estudem um pouco acima do que seus pais estudaram). Essa tendência foi encontrada em pesquisa sobre os meios populares realizada por Zago (2000), onde foi relatada as dificuldades que alguns estudantes passaram durante sua vida escolar e os casos daqueles que conseguiram obter um certificado escolar além do que seus pais possuíam. Um panorama das pesquisas sobre as camadas populares pode ser encontrado em Massi, Muzzeti e Suficier (2017) e em Nogueira, Romanelli e Zago (2000). Para uma visão do estado do conhecimento sobre a relação família-escola, confira Nogueira e Resende (2022).

Do outro lado, as famílias culturalmente ou economicamente dominantes (detentoras do maior volume de capital cultural e econômico, respectivamente) possuem uma relação diferente com a escola. As famílias culturalmente dominantes, que dependem dos diplomas e títulos escolares para a manutenção de sua posição privilegiada, tendem a valorizar a trajetória escolar dos filhos. Já no caso das famílias economicamente dominantes, sua posição social não depende prioritariamente da

escola. Portanto, espera-se que os filhos dessa parcela da sociedade estudem apenas para assumir um cargo ou manter o prestígio da família. As práticas das famílias da classe dominante são orientadas por um sentimento de distinção das demais classes, um diletantismo, resultando na busca por práticas que os distingam, seja em matéria de culinária, música, literatura, esporte, entre outros (BOURDIEU, 2007).

Em pesquisa sobre a escolarização dos filhos de elites empresariais em Minas Gerais, Nogueira (2002) observa que os estabelecimentos de ensino cursados, tanto na educação básica, quanto no Ensino Superior, são em sua grande maioria da rede privada, contrapondo algumas ideias que sugerem uma passagem da rede privada para a rede pública de ensino na saída do Ensino Médio e ingresso na universidade. Sobre os critérios para escolha dos estabelecimentos de ensino, Nogueira (2002) observa que a qualidade da clientela possui maior peso que a qualidade da escola, optando por instituições que considerem refletir os padrões da família. No que concerne ao tipo de curso superior escolhido, as evidências mostraram uma certa inclinação para áreas que possuem relação com a empresa familiar. Além disso, esses filhos também trabalham nas empresas de parentes (quando não dos pais), deixando o papel escolar como secundário por talvez não ser útil para o mundo do trabalho (NOGUEIRA, 2002).

Por fim, entre os dois extremos, encontram-se as famílias de classe média. Geralmente essas são as pessoas que devem seu relativo conforto à escola. Portanto, seu posicionamento frente à escola é diferente das demais famílias, onde seus filhos tendem a ser criados e incentivados a manter um estudo e dedicação intensa às matérias escolares. Esse posicionamento deve-se ao fato de que a classe média se encontra entre duas lutas (que geralmente não são percebidas): uma luta contra a desclassificação, ou seja, o medo de descender e se tornar classe popular; e uma luta pela reclassificação, ou seja, por uma ascensão social. Além disso, outras práticas correspondem às famílias de classe média, como o controle de natalidade (malthusianismo), a renúncia de prazeres imediatos tendo em vista benefícios futuros (ascetismo) e uma busca por conhecimento e imitação de práticas culturais dominantes (boa vontade cultural) (BOURDIEU, 2007).

Ao estudar como se dá a conversão do capital cultural em capital escolar nas camadas médias, Nogueira (2000) elenca algumas estratégias de excelência adotadas por essas famílias, como a escolha dos estabelecimentos de ensino, as informações sobre o sistema educativo, estadias no exterior, entre outras. É nesse

grupo onde encontram-se os indivíduos que tendem a apresentar a menor porcentagem de atraso escolar, alcançando, na idade, ou até mesmo antes dela, o Ensino Superior (NOGUEIRA, 2000). Além disso, entre os participantes da pesquisa, a universidade pública federal foi o destino de todos os estudantes. O que varia é, portanto, sua trajetória escolar durante a educação básica, sendo realizada totalmente ou em parte nas redes públicas e privadas de ensino. Com relação ao tipo de estabelecimento de ensino, é possível observar como as escolhas são orientadas, mesmo que inconscientemente, para os estabelecimentos que possuem uma certa tradição e excelência conhecida socialmente (NOGUEIRA, 2000). Complementando os resultados acima, encontramos trabalhos onde são discutidas mais a fundo o tema de escolha dos estabelecimentos de ensino e fatores que influenciam nessa escolha e nas estratégias escolares das famílias (NOGUEIRA; RESENDE; VIANA, 2015; PAULA; NOGUEIRA, 2018; RESENDE; NOGUEIRA; NOGUEIRA, 2011).

Vale dizer que, ainda que aplicadas em campos distintos, todas as estratégias dos agentes de diferentes classes sociais possuem uma relação entre si, tendo em vista que são fruto de um mesmo princípio. Dessa forma, a escolha do necessário, a boa vontade cultural e o diletantismo são o que definem as práticas das classes popular, média e dominante, respectivamente, tanto no mercado escolar quanto nos outros mercados sociais (BOURDIEU, 2007).

#### **2.1.4 O caráter arbitrário da avaliação escolar**

Toda educação, seja ela institucionalizada ou familiar, dá-se por meio de ações pedagógicas que consistem em dupla violência simbólica, pois não são mais que a inculcação, por um poder arbitrário, de um arbitrário cultural (BOURDIEU; PASSERON, 2010). Uma implicação dessa asserção é que as avaliações escolares, em geral, e o ENEM, em particular, são práticas igualmente arbitrárias.

O sentido de “arbitrário” na afirmação acima precisa ser observado com atenção. Arbitrário não é sinônimo de aleatório. Até mesmo porque, se as significações inculcadas por uma ação pedagógica pudessem ser, rigorosamente, qualquer coisa (uma combinação aleatória de ruídos e sinais) ela não seria compreendida. Na verdade, o objeto de toda ação pedagógica pode ser considerado arbitrário porque não deriva de um princípio universal. Essa condição de não poder ser derivada de princípios universais é satisfeita tanto pelas leis da Física (LIMA JUNIOR et al., 2014) quanto pelos costumes mais singulares de nossa cultura. Não

há, portanto, um princípio universal que segure firme as significações que ensinamos e as relações de poder que autorizam a ensinar. Quando algo poderia ser diferente (um sistema de ideias, de valores, de governo) dizemos que se trata de um *arbitrário*.

Os arbitrários culturais recebem esse nome porque foram arbitrados (como um lance numa partida de futebol) ainda que o árbitro nem sempre possa ser claramente identificado. Em outras palavras, arbitrários supõem uma decisão mais ou menos informada que nem sempre pode ser imputada a um ator individual. Currículos e avaliações escolares são, portanto, arbitrários. Eles poderiam ser de várias maneiras, mas acabam assumindo a forma mais conveniente aos grupos sociais que têm força para impor seus projetos educacionais. Ao ensinar a cultura dominante, aqueles que já possuem certa familiaridade com essa cultura (por estarem inseridos nela) terão uma vantagem sobre os demais. Tratando todos os alunos como iguais, deixando de lado as diferenças existentes, há um favorecimento dos favorecidos e um desfavorecimento dos desfavorecidos (BOURDIEU, 1998).

Sendo assim, o desempenho escolar tende a refletir a origem social dos estudantes. Enquanto os filhos de famílias com elevado capital terão uma relação mais amigável com a cultura escolar, outros terão mais dificuldade. Pensando nisso, podemos lançar um olhar mais crítico para o ENEM que, ao mesmo tempo em que indica os eleitos para o Ensino Superior, legitima o trabalho pedagógico das escolas mais capazes de garantir a eleição dos seus. Veem-se aqui argumentos sobre os quais se baseiam os discursos de qualidade escolar e qualidade do ensino de Ciências a partir de uma nota no ENEM ou em outros exames vestibulares. Nesses discursos, a qualidade escolar é medida e legitimada por desempenho em um certame, cujo conteúdo privilegia uma parcela da população, já que está mais próximo da cultura familiar dessas pessoas.

## 2.2 ESCOLAS QUE FAZEM A DIFERENÇA

O Relatório Coleman (COLEMAN, 1968) foi desenvolvido a partir de uma encomendação do Congresso norte-americano com o objetivo de investigar a igualdade de oportunidades educacionais entre os estudantes do país. O tipo de pesquisa desenvolvido por Coleman e colaboradores se enquadra como tipo *input-output* (entrada-saída) (BROOKE; SOARES, 2008). Esse tipo de análise considera a escola como uma unidade de produção que, recebendo insumos (como recursos materiais, financeiros e humanos), transformaria os indivíduos agregando-lhes um

valor (esse valor poderia ser, por exemplo, para a formação de capital humano para atuação no mercado de trabalho) (BRESSOUX, 2003). Contrariando o que se esperava do papel da escola, os resultados de Coleman (1968) indicaram que a diferença entre os investimentos financeiros, recursos e currículos das escolas parecia não explicar a variação no desempenho dos alunos (BROOKE; SOARES, 2008).

Ainda que, como apontado por Coleman (1968), os recursos das escolas não estejam relacionados ao sucesso escolar, isso não significa que não haja outros fatores que possuam essa relação (BRESSOUX, 2003). Outro ponto a ser levantado é que os resultados de Coleman e colaboradores, assim como de outras pesquisas feitas naquela época, associavam o sucesso escolar apenas ao desempenho em uma prova padronizada que aferia habilidades de leitura, escrita e matemática, conhecido como “três Rs”: *Reading, Writing e Arithmetics* (MORTIMORE et al., 1988). Todavia, como apontado por Mosteller e Moynihan (1972), o desempenho acadêmico é apenas um dos vários resultados que a escola deve proporcionar ao corpo discente. Essa crítica também foi apontada por Madaus, Airasian e Kellaghan (1980), onde os autores criticam a concentração de pesquisadores em resultados compostos por testes padronizados de desempenho, afirmando que os resultados escolares não são exclusivamente acadêmicos.

Esses resultados pessimistas e suas críticas desencadearam uma série de pesquisas sobre a eficácia escolar e o efeito-escola. O uso do termo “eficácia” no lugar de “eficiência” deve-se ao fato de que o último termo estaria mais relacionado ao pensamento econômico de custo e resultados, enquanto o primeiro termo diz respeito à qualidade das escolas. “Portanto, a eficácia não é uma relação entre o produto por unidade de investimento, mas sim a capacidade das escolas de produzirem efetivamente os resultados que a sociedade espera delas” (BROOKE; SOARES, 2008, p. 20). Essa capacidade se encontra na definição do termo “efeito-escola”. Seu entendimento é marcado por quanto “um dado estabelecimento escolar, pelas suas políticas e práticas internas, acrescenta ao aprendizado do aluno” (BROOKE; SOARES, 2008, p. 10). Sendo assim, as escolas passaram a ser avaliadas pelos seus resultados e deu-se início, também, a uma busca por fatores que contribuíssem para esses resultados.

A geração de trabalhos que veio em reação aos resultados pessimistas das pesquisas do tipo insumo-produto (*input-output*) estava preocupada em analisar os processos escolares, entendendo a escola como organização social e não como

unidade de produção (BRESSOUX, 2003). Esses trabalhos, portanto, foram além do olhar somente para os recursos e buscaram outros fatores que pudessem influenciar no sucesso escolar, com o objetivo de evidenciar os processos que podem desencadear em diferenças de eficácia entre as escolas (*ibid*).

Rutter et al. (1979) investigaram alguns fatores que poderiam estar relacionados a alguns resultados escolares. Os autores concluem que há diferenças estatisticamente significativas (mesmo quando controladas características dos estudantes) entre as escolas com relação às variáveis de frequência, comportamento dos alunos, delinquência e desempenho escolar. Afirmam que, no geral, bom comportamento e altos níveis de frequência costumam estar associados a altos níveis de sucesso nos exames (RUTTER et al., 1979). Ainda que os autores sinalizem que algumas escolas, situadas em uma região menos favorecida de Londres, sejam mais capazes de promover bom comportamento e desempenho, os seus resultados, sozinhos, não mostram que as variações encontradas são devidas exclusivamente às características das escolas (*ibid*). Além disso, Rutter et al. (1979) verificaram que tanto a forma como os estudantes eram tratados enquanto indivíduos quanto a cultura escolar influenciaram os alunos das escolas investigadas.

Mortimore et al. (1988), em seu livro intitulado “*School Matters*” (A escola tem importância), extrapolam o entendimento de que os resultados escolares estão associados à leitura, à escrita e à matemática, e lançam um olhar para os resultados cognitivos (desempenhos e progressos escolares) e não-cognitivos (taxa de frequência, comportamento, autoavaliação e atitudes) dos estudantes de escolas primárias. Como justificativa para seu olhar, os autores afirmam que dessa forma seria possível obter uma medida mais justa da eficácia geral da escola (MORTIMORE et al., 1988). Em suma, os autores encontram que a idade, classe social, sexo e raça têm um impacto tanto nos resultados cognitivos quanto nos resultados não-cognitivos. No entanto, os autores ressaltam que frequentar certa escola contribui para a explicação das variações nos resultados cognitivos e não-cognitivos dos estudantes e que a escola frequentada no ensino primário pode surtir um efeito positivo ou negativo não só no desempenho, mas também no progresso dos alunos (MORTIMORE et al., 1988). Mortimore et al. (1988) sintetizam seus resultados afirmando que provavelmente os efeitos da escola primária acompanharão o aluno na sua vida escolar e podem ter uma influência a longo prazo no sucesso educacional e nas perspectivas de emprego.

Edmonds (1979), em sua pesquisa, explicitou alguns fatores que se mostraram associados a bons desempenhos dos estudantes de escolas elementares localizadas em bairros desfavorecidos. São eles: (i) forte liderança administrativa; (ii) clima de expectativa onde nenhuma criança deve estar abaixo da linha eficaz de desempenho; (iii) atmosfera escolar ordenada sem ser rígida; (iv) prioridade na aquisição de habilidades escolares básicas pelos alunos; (v) se necessário, a energia e recursos de outras atividades podem ser utilizados para atingir os objetivos fundamentais; e (vi) monitoramento frequente do progresso dos alunos. Ainda que esses fatores não possam ser generalizados (ou seja, são sensíveis ao contexto social), eles foram amplamente utilizados em programas de melhorias das escolas (BRESSOUX, 2003). Pesquisas futuras observaram que, dos itens enunciados acima, apenas os itens referentes ao clima disciplinar e ao monitoramento frequente dos progressos dos alunos se mostraram não sensíveis ao contexto social da escola (BRESSOUX, 2003).

Ainda dentro das pesquisas sobre efeito-escola, houve um movimento de ruptura com a utilização de dados brutos e uma preocupação para o que é denominado como “valor agregado”. Essas técnicas, no fundo, pretendem determinar o quanto uma escola agregou ao progresso de um aluno, tomando como comparação o que seria o desempenho esperado desse aluno, segundo suas condições socioeconômicas, de acordo com um modelo estatístico. A essência desses métodos de valor agregado consiste no fato de focar no papel da escola no progresso do aluno e não apenas em um único resultado ao final do período escolar (SOARES; ALVES; OLIVEIRA, 2001). Esse tipo de análise foi possível com o avanço das técnicas estatísticas de análise de dados agrupados. Recomenda-se que sejam utilizados modelos multiníveis (que tratam os dados em agrupamentos hierarquizados – por exemplo, alunos – escola – estados) para uma investigação de valor agregado (GRAY; WILCOX, 1995).

Valerie Lee (2000) discute sobre a importância do uso de métodos multiníveis e apresenta algumas contribuições que a modelagem linear hierárquica (MLH) pode trazer para estudos de contextos sociais em pesquisas educacionais. De maneira muito didática, a autora comenta sobre cuidados e dá orientações para aqueles que desejam utilizar a MLH em suas pesquisas (LEE, 2000). Para exemplificar o poder dessa técnica para as pesquisas em Educação, ela traz um exemplo onde analisa o efeito do tamanho de uma escola de Ensino Médio no aprendizado de Matemática e

Leitura. Dada a relevância da pesquisa, os resultados desse trabalho ganharão mais destaque e estão apresentados nos parágrafos seguintes.

Como primeiro resultado, Lee (2000) observa que escolas cujo corpo discente varia de 600 a 900 estudantes se mostraram mais eficazes na aprendizagem de Matemática e Leitura do que outras faixas de corpo discente após o controle da condição social dos alunos, suas habilidades no momento de entrada no Ensino Médio e a composição e localização da escola. Esse resultado deve ser interpretado em relação às escolas com corpo discente de 1200 a 1500 alunos, que foram tomadas como referência para comparação. Além disso, as maiores escolas (com relação ao número de alunos matriculados) foram as que apresentaram os menores ganhos de aprendizagem nas duas disciplinas.

O segundo resultado interessante no trabalho de Lee (2000) diz respeito à distribuição mais equitativa do aprendizado entre o corpo discente das escolas. Em seu modelo, são medidos também os efeitos do tamanho da escola na relação entre o nível socioeconômico (NSE) dos alunos e seu aprendizado em Matemática e Leitura. Portanto, uma distribuição social mais igual de aprendizado é apresentada por uma relação menos extrema de NSE em relação ao desempenho nestas disciplinas (LEE, 2000). A autora comenta, então, que as escolas pequenas (com até 600 alunos) são mais equitativas, sendo o tamanho da escola mais influente na equidade de aprendizagem de Leitura. Resumindo os dois primeiros resultados, escolas entre 600 e 900 alunos são mais eficazes com relação ao aprendizado médio. Já escolas com até 600 alunos são mais equitativas.

Por fim, como terceiro resultado, Lee (2000) também investiga se os efeitos do tamanho da escola eram os mesmos em escolas com tipos de estudantes distintos. A autora traz três destaques: (i) os alunos aprendem mais Matemática quando estão em escolas de alto NSE durante o Ensino Médio; (ii) o tamanho ideal de uma escola é o mesmo, independente do NSE, e é de 600 a 900 estudantes; e (iii) o tamanho da escola faz mais diferença para aquelas cujo corpo discente possui NSE mais baixo. Portanto, a partir dos três resultados apresentados acima, vemos que escolas de Ensino Médio muito grandes podem ser ambientes desfavoráveis ao aprendizado de seus estudantes, sobretudo se o público dessa escola é majoritariamente desfavorecido socialmente. Para um panorama de outras pesquisas internacionais e nacionais sobre a eficácia escolar, confira Brooke e Soares (2008).

### 2.3 O QUE É FAZER A DIFERENÇA?

No Brasil, a partir da década de 1990, ganha força um movimento compulsivo de medição de qualidade escolar. Nesse movimento, que busca equiparar o funcionamento da escola ao de uma empresa, ocorre uma translação da ênfase na avaliação individual do aluno para a avaliação das instituições e redes de ensino (avaliações sistêmicas), onde opera a lógica sob a qual a qualidade escolar está atrelada ao desempenho em provas padronizadas (como as avaliações de larga escala tipo ENEM) (FREITAS, 2007). Segundo essa racionalidade, ao criar um *ranking* das escolas com base em seus desempenhos em um exame nacional, estaríamos contribuindo para que as escolas e os sistemas de ensino de maior qualidade ganhem visibilidade e sejam celebradas. Essas classificações poderiam ser usadas, por exemplo, como critério para a escolha do estabelecimento de ensino em função das ambições de cada família (NOGUEIRA; RESENDE; VIANA, 2015; RESENDE; NOGUEIRA; NOGUEIRA, 2011).

A reforma empresarial da educação busca incentivar uma melhoria na qualidade da educação a partir da criação de competitividade entre as instituições e redes de ensino, ao mesmo tempo em que transfere a outras esferas a responsabilidade pela qualidade do ensino ofertado (FREITAS, 2002). Vemos, com isso, como essa reforma está ancorada na racionalidade neoliberal, que tem por objetivo, entre outros, aplicar a lógica competitiva e empresarial em outros setores da sociedade (DARDOT; LAVAL, 2016). São procedimentos da reforma empresarial da educação: padronização, testes e responsabilização verticalizada (FREITAS, 2018). “A lógica esperada é que, definindo o que se deve ensinar, a escola saberá o que ensinar, os testes verificarão se ela ensinou ou não, e a responsabilização premiará quem ensinou e punirá quem não ensinou” (FREITAS, 2018, p. 78).

Aplicando o mesmo currículo às escolas (padronização do currículo) e avaliando a eficiência de sua implementação pelo mesmo instrumento (por exemplo, o ENEM), criamos um ambiente favorável à seguinte argumentação de (des)legitimação de algumas escolas e redes de ensino: as diferenças entre os desempenhos obtidos na prova de Ciências da Natureza e no ENEM seriam o produto da relação entre dois fatores – o trabalho pedagógico e a gestão. Sendo assim, as escolas e estados que alcançassem um desempenho abaixo, que "se saíssem mal" na avaliação, seriam aqueles onde estariam os piores professores e gestores.

Analogamente, os melhores professores, aqueles que mereceriam ser premiados pelos seus trabalhos, estariam nas melhores escolas segundo o *ranking*. Diante do que foi discutido até aqui, queremos destacar dois efeitos que as classificações das escolas exercem e suas implicações para a definição do que é fazer a diferença.

Em primeiro lugar, podemos observar que há diferentes pontos de vista sobre qualidade escolar e, mais especificamente, o que deve ser levado em conta para considerar uma escola eficaz. Willms (1992), em seu livro "*Monitoring School Performance: a Guide for Educators*" (Monitoramento do desempenho de escolas: um guia para educadores) comenta sobre as diferentes concepções do termo "efeito da escola" em pesquisas educacionais. Podemos enumerá-las da seguinte maneira: (1) atrelada aos efeitos de políticas, práticas ou intervenções sobre o desempenho; (2) atrelada à diferença entre o nível médio de desempenho de uma escola e um determinado padrão após o controle de variáveis de características dos alunos ao ingressarem na escola (aqui, há um efeito próprio associado à frequência de uma escola em específico); (3) atrelada à indicação do quanto as escolas variam em relação aos resultados de seus alunos (aqui entram estudos cujos resultados apontam para o percentual de diferença entre os desempenhos dos estudantes que é devido à diferença entre as escolas) (WILLMS, 1992).

As decisões que os pesquisadores tomam para investigar as escolas eficazes não são neutras, mas estão relacionadas à visão que aquelas pessoas têm da educação e à sua posição na estrutura social. Ou seja, há um interesse por trás de cada forma de classificar as escolas. Por exemplo, Willms (1992) comenta que a depender de como uma pesquisa é conduzida, os resultados podem ser mais interessantes para os pais, que podem usá-los para escolher onde matricular seus filhos, ou para gestores e professores, que podem usá-los para avaliar suas práticas educacionais. Ademais, os diferentes modelos, incluindo ou não certas variáveis, produzem classificações diferentes (SOARES; ALVES; OLIVEIRA, 2001), podendo até mesmo prejudicar certos tipos de escolas (FREITAS, 2018). Além disso, vemos que as decisões dos pesquisadores também se explicitam no grupo de variáveis que eles consideram como resultados de eficácia escolar, onde, muitas vezes, vemos a presença apenas de fatores relacionados ao desempenho em uma prova padronizada de Leitura e Matemática (BRESSOUX, 2003).

Portanto, vemos que não há uma maneira única e universal de aferir qualidade. A maneira dominante de classificação é fruto das lutas e disputas entre os

agentes sociais que estão envolvidos no campo educacional e, conseqüentemente, refletem, também, seus interesses. No entanto, o sucesso de uma maneira de classificar depende, de certa forma, da dissimulação do interesse por trás dessa forma. Quanto mais oculta e dissimulada forem as intenções por trás de uma determinada técnica de ranqueamento, mais provável é que tenha sucesso, ou seja, que seja aceita como medida legítima de diferenças escolares objetivas. Como comentado anteriormente, ao reconhecermos alternativas ao modelo hegemônico empregado, vemos sua arbitrariedade. Note que as alternativas não são quaisquer, mas fazem sentido de acordo com o que se pretende medir. De fato, afirmar que um *ranking* é arbitrário não quer dizer que ele seja aleatório, mas que sua maneira de classificar não é única e pode ser relacionada a um conjunto de interesses e posições sociais.

Em segundo lugar, precisamos refletir sobre as classificações que fazemos e suas conseqüências. Podemos nos perguntar se todas as classificações que fazemos, mesmo as mais corriqueiras, estão no mesmo patamar. Sem dúvida, estamos a todo momento classificando objetos e/ou pessoas. Um exemplo diz respeito à matéria do gosto. Dizer que não gosto de algo que uma outra pessoa gosta vai além de uma simples repulsa ao que foi julgado. No fundo, pode ser interpretado como “não gosto do seu gosto”, onde, ao mesmo tempo em que me diferencio, classifico-me em relação a essa pessoa (BOURDIEU, 2020). Importante frisar que uma classificação é sempre feita em relação a algo, comparando os objetos dessa classificação, muitas vezes tomando como referência padrões idealizados (LIMA JUNIOR et al., 2020a).

Algumas classificações que atuam no mundo social têm efeitos sociais muito importantes. Os critérios que utilizamos para classificar são mais do que recortes da qualidade das coisas (ou mais especificamente, das escolas), já que as qualidades dos objetos não pertencem a eles mesmos, mas às relações que temos com eles. Dessa forma, alguns critérios empregados nas pesquisas são constitutivos da realidade, recortando-a de acordo com algum interesse e, ao mesmo tempo, produzindo, de certa maneira, as características do que foi nomeado (BOURDIEU, 2020). Essas classificações podem ser pensadas, então, dentro dos chamados *atos de instituição* (BOURDIEU, 2020). São *atos performativos* onde significamos algo para alguém no sentido, também, de “‘eu te convoco’, ‘eu te mando’, ‘eu te ordeno’ a ser aquilo que eu digo que você é” (BOURDIEU, 2020, p. 33).

No entanto, para que um ato performativo, ou ato de instituição, como os *rankings*, sejam bem-sucedidos, é necessário que essa classificação tenha chances de ser verificada na sociedade (BOURDIEU, 2020). Por exemplo, ao ranquear uma escola como “boa”, segundo algum critério arbitrário de quem classifica, surge um efeito onde as pessoas e instituições que estão dentro desse contexto classificatório tenderão a agir de forma a manter sua posição privilegiada (NOGUEIRA; LACERDA, 2014), validando, portanto, o *ranking* que a classificou. Essa é a característica de um ato performativo, uma ação “mágica” que visa agir sem agir, agir de maneira simbólica e que, além disso, detém uma força (BOURDIEU, 2020). Podemos ver esses casos, por exemplo, quando professores, por meio de falas que pretendem ser livres de intenções, interpelam seus alunos a agirem de forma a alcançar um padrão idealizado de desempenho acadêmico (LIMA JUNIOR et al., 2020a).

Afirmar, com base em um ranqueamento, que uma escola não é “boa” vai além de simplesmente constatar. Sob as aparências de um enunciado constativo, que pode ser julgado como verdadeiro ou falso, essa afirmação dissimula que o simples fato de usar um *ranking* para classificar uma escola evidencia seu caráter criativo e participativo da qualidade escolar. É justamente porque, ao utilizarmos um critério para ranquear a qualidade das escolas, por exemplo, os dados brutos de desempenho no ENEM, criamos uma realidade (BOURDIEU, 2020), que podemos perceber que o critério utilizado, e sua posterior classificação, participa e constrói a qualidade escolar medida. Ou seja, os *rankings* não descrevem a realidade das escolas, mas participam na e da construção da qualidade que pretendem “medir”.

Diante do discutido até aqui, vemos como a qualidade escolar não é um dado de fato, mas resulta de construções arbitrárias das pessoas envolvidas no campo educacional, sendo, portanto, objeto de luta e disputa, e que os meios institucionais que foram instituídos para medi-la tendem a exercer um papel que vai além de constatação, verificação de uma característica, mas tendem a participar da construção daquilo que pretendem medir. Como consequência, estar bem classificado nos *rankings* suporia um bom trabalho pedagógico e uma boa gestão. Dentro dessa lógica, esses ranqueamentos apenas estariam tornando público o mérito das instituições escolares e das Unidades Federativas, enquanto dissimulam que a qualidade divulgada é criada, também, pelo próprio indicador de qualidade empregado.

Sabendo que os desempenhos no ENEM são influenciados pela origem social dos candidatos (LIMA JUNIOR; FRAGA JUNIOR, 2021) e que os recursos não estão

igualmente distribuídos no sistema de ensino, acumulando-se nas regiões mais centrais e abastadas (GARCIA DA SILVA; LIMA JUNIOR, 2020), segue, então, que as instituições ou Unidades da Federação cuja posição social média é mais elevada terão uma vantagem, aumentando suas chances de ocupar as primeiras colocações nos *rankings* de “qualidade” educacional e “qualidade” de ensino de Ciências. Com isso, algumas escolas e territórios são prejudicados pela forma como se dão os ranqueamentos (a partir de dados brutos), pois os *rankings* são obtidos dissimulando o efeito da origem social. Dessa forma, opera-se uma lógica que atribui às vítimas desse efeito a responsabilização por não alcançarem certos desempenhos (FREITAS, 2018), ao mesmo tempo em que são premiadas e celebradas as “conquistas” daqueles que são beneficiados pelo mesmo efeito. No entanto, esse é um resultado fruto do discurso neoliberal: fala-se de igualdade de oportunidades, não de resultados (FREITAS, 2007).

Vemos, portanto, como os ranqueamentos fazem parte de um projeto que atrela qualidade ao desempenho em uma avaliação e que, no fundo, dissimulam um desmonte da educação pública brasileira. A educação em Ciências não escapa à essa racionalidade. O ensino de Ciências realizado nas instituições escolares será avaliado e premiado, ou não, com base na posição nessas classificações e no resultado de indicadores arbitrariamente construídos. Ou seja, o ensino de Ciências das escolas, que deve atender ao desenvolvimento de competências e habilidades descritos nos documentos normativos da educação brasileira, tem sua qualidade verificada em avaliações padronizadas.

Ora, como a excelência escolar é arbitrária, ou seja, não está dada na natureza das escolas, há alternativas para medi-la, utilizando variáveis como nível de formação docente, estrutura física da escola, desempenho em provas de Língua Portuguesa e Matemática etc. Note que, ao existir alternativas ao modelo empregado, vemos sua arbitrariedade. Ao escolhermos uma maneira de averiguar a excelência escolar, escolhemos, também, o que será levado em conta nessa medição, isto é, o que acreditamos ser excelência escolar.

### 3 REVISÃO DA LITERATURA

Baseados na metodologia empregada por Schnorr e Pietrocola (2022), realizamos uma revisão sistemática dos artigos publicados até o momento que contivessem, em seu título, as expressões “Exame Nacional do Ensino Médio” ou “ENEM” e, em qualquer campo, as expressões “Educação” ou “Ensino”. A busca foi realizada na plataforma Portal Periódicos e resultou, inicialmente, em 518 trabalhos. Após a retirada dos trabalhos homônimos e daqueles que não possuíam título e resumo em português também, restaram 364 artigos. Seus metadados foram exportados em uma tabela e trabalhados no *software* de programação *RStudio*, com o auxílio das bibliotecas “*tidytext*”, “*dplyr*”, “*wordcloud*”, “*widyr*”, “*igraph*” e “*ggraph*”.

O objetivo dessa revisão é responder às seguintes perguntas:

- 1- *O que se discute sobre o Exame Nacional do Ensino Médio (ENEM) na literatura?*
- 2- *Quais os resultados das pesquisas sobre o ENEM que tratam de questões socioeconômicas e desempenho?*

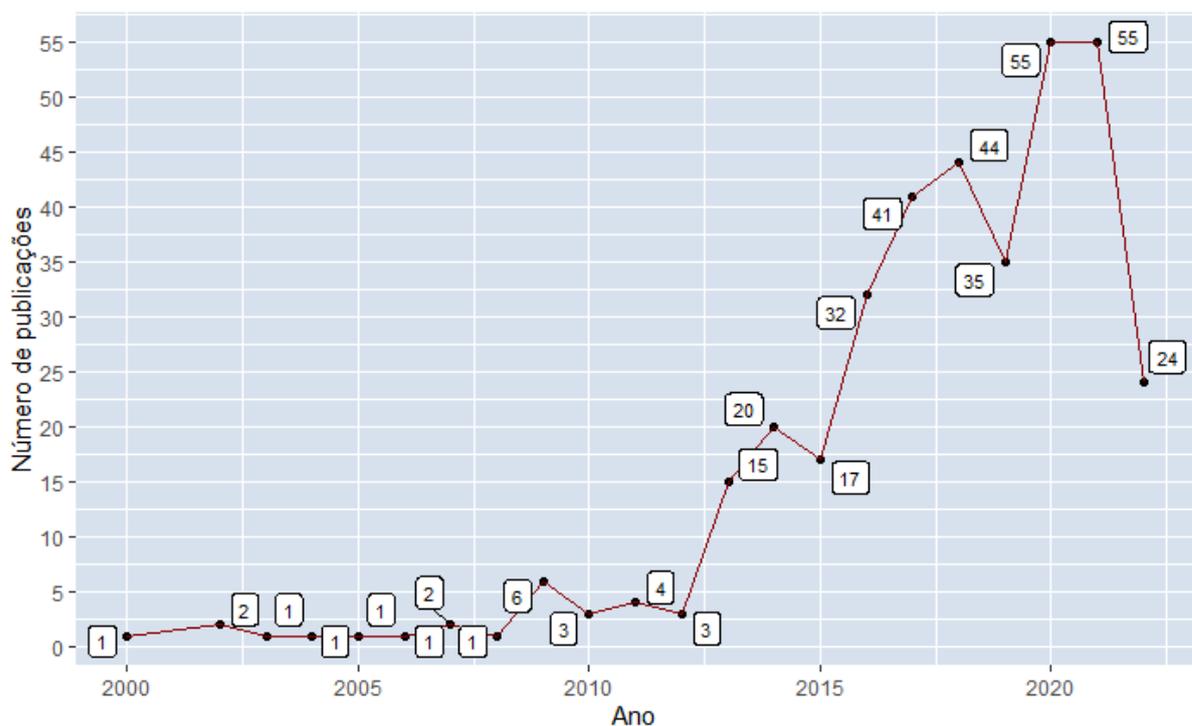
Para responder à pergunta 1, baseamo-nos na mineração de textos por meio da linguagem de programação *R* (SILGE; ROBINSON, 2017) e utilizamos recursos como nuvens de palavras e redes de conexões, olhando para quais palavras mais aparecem e mais se relacionam nos títulos e resumos dos 364 artigos selecionados. Dentre esses 364 artigos, realizamos um segundo filtro separando aqueles artigos que tratavam de questões socioeconômicas e desempenho no ENEM, dos quais fizemos a leitura completa e classificamos os resultados desses trabalhos em temas.

A mineração de texto aqui empregada funciona a partir da quebra dos títulos e resumos dos artigos em *tokens*, através do processo de *tokenization* (SILGE; ROBINSON, 2017). Neste exame, decidimos utilizar as palavras como unidades dos *tokens*. Ou seja, cada palavra gerará um *token* diferente. Ao fazer isso, métodos estatísticos podem ser utilizados com mais facilidade, como a contagem de frequência e a associação entre as palavras. De posse desses números, é possível confeccionar gráficos que podem revelar informações interessantes sobre os textos analisados.

### 3.1 DISTRIBUIÇÃO DOS ARTIGOS AO LONGO DOS ANOS ANALISADOS

No Gráfico 1, é possível ver como está a distribuição desses 364 artigos ao longo dos anos. Como vemos, o número de publicações ao longo dos anos 2000 é baixíssimo, tendo uma média de 1.6 artigos publicados nesse decênio. Merece destaque o ano de 2009, com um número de publicações igual 6. Talvez esse aumento na atenção dada ao ENEM seja decorrente da mudança no perfil do certame, que passou a valer como porta de ingresso para o Ensino Superior público brasileiro.

Gráfico 1 – Distribuição das 364 publicações de 2000 a 2022



Fonte: elaborado pelos autores

Ainda nesse gráfico, vemos que nos anos de 2013 e 2016 os artigos sobre o ENEM destoam bastante dos anos anteriores a eles, alcançando um número de publicações igual a 15 e 32, respectivamente. Mais recentemente, vemos que os anos de 2020 e 2021 são os que possuem mais publicações, totalizando 55 artigos cada. Com isso, até o momento, aproximadamente 37% ( $n = 134$ ) dos artigos analisados nesta revisão sistemática estão contidos nos 3 últimos anos analisados (2020-2022) contra 59% ( $n = 214$ ) distribuídos de 2010 a 2019. Da tendência observada no Gráfico 1, é esperado que as publicações de 2020 a 2029 sejam, em número, ainda maiores que os anos anteriores presentes nessa análise.

## 3.2 FREQUÊNCIA E ASSOCIAÇÃO ENTRE AS PALAVRAS

### 3.2.1 Títulos dos artigos

Ao dividir os títulos das publicações selecionadas em *tokens*, onde cada *token* representará uma palavra diferente, podemos verificar como se dá a frequência e associação entre as palavras dos artigos selecionados. Na Figura 2, é possível ver uma nuvem das 50 palavras mais frequentes nos 364 artigos no período analisado. Nessa figura, o tamanho das palavras é proporcional ao seu número de ocorrências. Portanto, quanto maior for a palavra dentro dessa nuvem, mais frequente ela foi nos títulos dos artigos.

Figura 2 – Nuvem das 50 palavras mais frequentes nos títulos dos artigos no período analisado



Fonte: elaborado pelos autores

Como é possível observar da Figura 2, as palavras que mais se destacam são: “análise” (n = 68), “questões” (n = 44), “física” (n = 35), “redação” (n = 32) e “desempenho” (n = 28). Podemos depreender, a partir das informações sobre os títulos dos artigos, que o foco das pesquisas está mais alinhado às questões e ao conteúdo das provas do ENEM do que a outros papéis que o exame pode assumir, como avaliador da qualidade da educação no Brasil ou como reproduzidor de desigualdades sociais, por exemplo. Ainda que em menor número, existem pesquisas preocupadas com o significado político do ENEM ou com as desigualdades sociais presentes em suas provas.

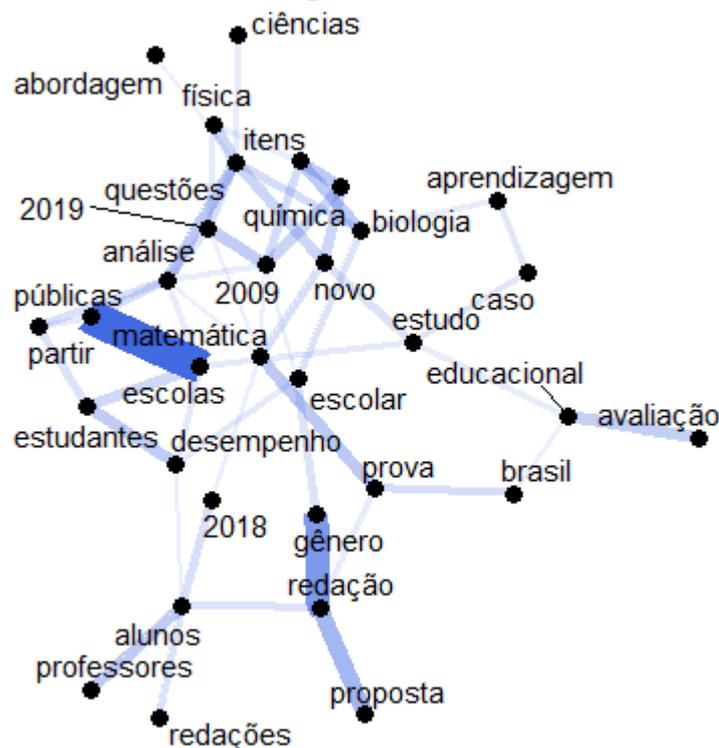


“redação” e “proposta” ( $p = 0.28$ ); “análise” e “questões” ( $p = 0.22$ ); e “itens” e “química” ( $p = 0.22$ ).

Dos resultados presentes na Figura 4, podemos depreender que as pesquisas cujo título contém a palavra “escolas” estão mais preocupadas em investigar aquelas que são “públicas” em detrimento das escolas privadas, por exemplo. Além disso, vemos que a palavra “gênero” está mais associada às investigações referentes à prova de Redação do ENEM do que questões de gênero relacionadas ao desempenho no exame.

Também corroborando as discussões a partir das Figuras 2 e 3, vemos que as palavras “questões”, “itens” e “análise” estão mais correlacionadas às palavras que se referem à área de Ciências da Natureza (“física”, “química” e “biologia”), mostrando que, em sua maioria, os artigos aqui analisados estão mais preocupados com o conteúdo e as questões da prova do ENEM do que com outros temas de igual importância.

Figura 4 – Rede de conexões das correlações entre as palavras presentes nos títulos dos artigos analisados



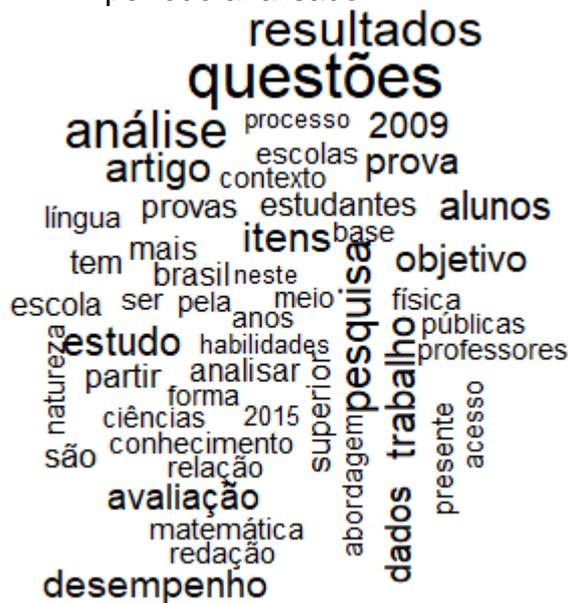
Fonte: elaborado pelos autores

### 3.2.2 Resumo dos artigos

Repetindo os procedimentos de mineração de textos aplicados aos títulos nos resumos dos artigos, obtemos as Figuras 5, 6 e 7 que representam a nuvem das 50

palavras mais frequentes, os pares de palavras que ocorrem mais vezes e as palavras mais correlacionadas nos resumos dos trabalhos analisados, respectivamente. Na Figura 5, temos que as palavras mais frequentes são: “questões” (n = 292), “análise” (n = 218), “resultados” (n = 212), “pesquisa” (n = 176) e “artigo” (n = 159). Considerando que é nessa seção onde os autores enunciam os pontos-chave de sua pesquisa, não surpreende a aparição das palavras “artigo”, “pesquisa” e “resultados”. No entanto, a mesma interpretação não se aplica às palavras “questões” e “análise”, já que também são as que mais aparecem nos títulos dos artigos (vide Figura 2).

Figura 5 – Nuvem das 50 palavras mais frequentes nos resumos dos artigos no período analisado



Fonte: elaborado pelos autores

Na Figura 6, os pares de palavras que aparecem mais vezes são: “análise-resultados” (n = 81), “resultados-pesquisa” (n = 74), “artigo-análise” (n = 73), “análise-pesquisa” (n = 69) e “artigo-resultados” (n = 67). Mais uma vez, por se tratar de uma seção onde todas as informações relevantes da pesquisa encontram-se condensadas, dando um panorama geral da investigação realizada, não causa surpresa os pares de palavras aqui elencados.

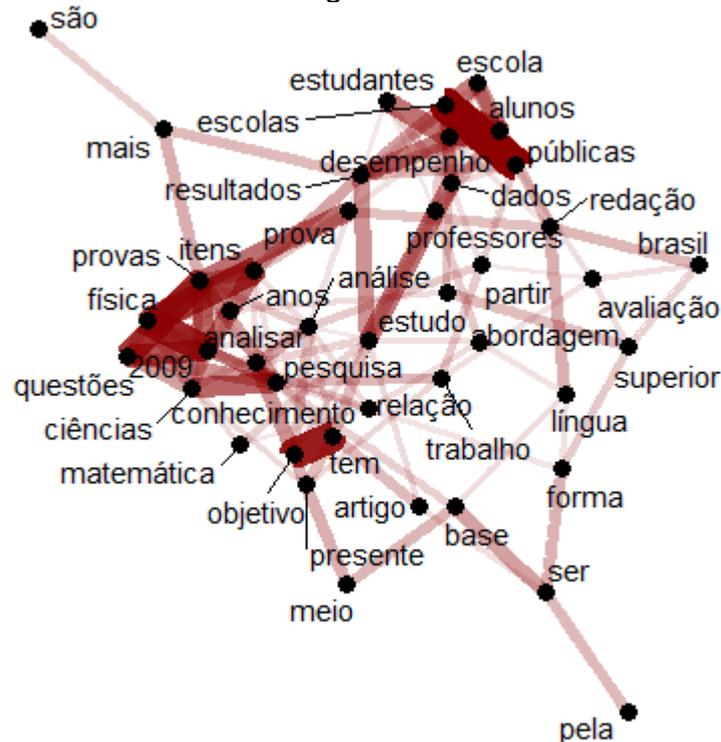
Figura 6 – Rede de conexões entre os pares de palavras presentes nos resumos dos artigos analisados



Fonte: elaborado pelos autores

Na Figura 7, os pares de palavras que possuem maior correlação são: “públicas-escola” ( $p = 0.36$ ), “tem-objetivo” ( $p = 0.33$ ), “física-2009” ( $p = 0.31$ ), “questões-prova” ( $p = 0.30$ ) e “física-itens” ( $p = 0.29$ ). Semelhantemente ao resultado da Figura 4, as palavras “escola” e “públicas” aparecem mais juntas do que com outras palavras. Além disso, a partir das correlações entre a palavra “física” e as palavras “2009” e “itens”, podemos inferir que os artigos aqui analisados cuja investigação toca algum tema relacionado à Física, tomam por base a edição do exame de 2009 e estão mais focados nos itens da prova. Mais uma vez aparece a correlação entre as palavras “questões” e “prova”, contribuindo para o argumento do enfoque que é dado ao ENEM dentre as publicações selecionadas.

Figura 7 – Rede de conexões das correlações entre as palavras presentes nos resumos dos artigos analisados



Fonte: elaborado pelos autores

### 3.3 ENEM E QUESTÕES SOCIOECONÔMICAS

Complementando a análise até agora apresentada, realizamos a leitura de artigos que tratassem de questões socioeconômicas em suas investigações. Dessa forma, dentre os 364 artigos, realizamos um filtro cuja função era selecionar trabalhos que contivessem os termos “classe\*”, “reprod\*”, “desiguald\*” e “socio\*” (o asterisco indica a possibilidade de inserção de termos com terminações variadas). Nessa primeira filtragem, 21 artigos foram encontrados. Realizamos, então, uma segunda filtragem, considerando o índice h5 do periódico onde o artigo fora publicado (segundo o Google Acadêmico). Desses, 14 trabalhos estavam publicados em revistas cujo índice h5 era igual ou maior que 10. Portanto, nessa segunda parte da revisão da literatura, olhamos para os objetivos, métodos e resultados dos 14 artigos que tratam sobre questões socioeconômicas posicionados nos primeiros estratos de avaliação.

Na Tabela 1, é possível verificar a autoria, o contexto, os objetivos e o tema do resultado dos artigos aqui analisados. Como é possível perceber, 50% (n=7) dos artigos foram produzidos na região Sudeste, com destaque para o estado de São Paulo, com 6 trabalhos. Os outros 7 trabalhos dividem-se entre: Nordeste (n=4), Sul (n=2) e Centro-Oeste (n=1).

A categoria “Tema do resultado” foi elaborada a partir do caráter dos resultados apresentados pelos pesquisadores. São eles: (i) *desempenho individual* (n=3) – os resultados estão a nível dos indivíduos (i.e., dos candidatos do ENEM); (ii) *desempenho agrupado* (n=3) – os resultados agrupam os desempenhos de alguma forma, seja por escola, seja por municípios/estados/regiões; (iii) *desempenho individual e agrupado* (n=5)– os resultados aparecem tanto a nível dos candidatos quanto sob algum tipo de agrupamento; e (iv) *itens/conteúdo* (n=3) – os resultados estão preocupados com a relação entre o conteúdo da prova com questões socioeconômicas. Apresentaremos os principais resultados dos artigos por tema nas subseções a seguir.

Tabela 1 – Autoria, contexto e objetivos dos 14 artigos analisados que tratam de questões socioeconômicas no ENEM

<b>Autoria</b>	<b>Contexto</b>	<b>Objetivos</b>	<b>Tema do resultado</b>
Almeida et al. (2017)	Alagoas	Comparar resultados das escolas de AL com as demais do Nordeste.	Desempenho agrupado
Dutra et al. (2019)	Ceará	Investigar o poder dos contextos no desempenho dos Institutos Federais.	Desempenho agrupado
Figueirêdo, Nogueira e Santana (2014)	Paraíba	Investigar o quanto a origem individual influencia no desempenho dos alunos.	Desempenho individual
Justiniano e Queiroz (2021)	São Paulo	Analisar dados de alunos de escolas públicas, agrupando por escola e por zona, e representando em diferentes unidades administrativas.	Desempenho individual e agrupado
Kleinke (2017)	São Paulo	Compreender como fatores econômicos podem afetar o desempenho em Física no Enem.	Itens/conteúdo
Lima Junior e Fraga Junior (2021)	Distrito Federal	Investigar o efeito da desigualdade social no desempenho dos estudantes brasileiros em Ciências da Natureza.	Desempenho individual
Lucena e Santos (2020)	Rio Grande do Norte	Analisar como as características socioeconômicas dos estudantes brasileiros podem influenciar a nota geral dos candidatos.	Desempenho individual
Marcom e Kleinke (2021)	São Paulo	Utilizar os dados do ENEM para redefinir seu papel de processo seletivo.	Itens/conteúdo
Melo et al. (2021)	São Paulo	Identificar variáveis com maior impacto no desempenho dos municípios tanto na prova objetiva quanto na redação.	Desempenho individual e agrupado
Moraes et al. (2022)	Rio de Janeiro	Debater equidade e desigualdade de desempenho para alunos de escolas públicas.	Desempenho individual e agrupado
Nascimento, Cavalcanti e Ostermann (2018)	Rio Grande do Sul	Analisar as questões de Física do ENEM e encontrar exemplos de itens menos associados com o nível socioeconômico.	Itens/conteúdo
Navarro et al. (2021)	São Paulo	Analisar o desempenho de estudantes em Ciências da Natureza comparativamente a outras áreas do conhecimento.	Desempenho individual e agrupado
Torres et al. (2020)	Rio Grande do Sul	Averiguar a ocorrência de diferenças entre os ensinos público e privado no RS a partir de dados do ENEM	Desempenho agrupado
Travitzki, Ferrão e Couto (2016)	São Paulo	Examinar a relação entre desempenho educacional e atributos sociodemográficos.	Desempenho individual e agrupado

Fonte: elaborado pelos autores

### 3.3.1 Itens/conteúdo

Como visto nas Figuras 2 e 5, as palavras “análise”, “itens”, “questões”, “física” e “desempenho” foram bastante utilizadas nos títulos e resumos dos 364 artigos analisados via mineração de texto. Na Tabela 1, encontram-se três artigos que estão preocupados em analisar itens da prova do ENEM com uma abordagem de questões

socioeconômicas (KLEINKE, 2017; MARCOM; KLEINKE, 2021; NASCIMENTO; CAVALCANTI; OSTERMANN, 2018).

Segundo Kleinke (2017), ao analisar os itens de Física das edições de 2011, 2012 e 2014 do ENEM, a média de desempenho nesses itens cresce em função do nível socioeconômico (NSE) do candidato. Além disso, o autor afirma que, por um lado, as questões de Física que envolvem cálculos matemáticos apresentaram maior reprodução das desigualdades sociais. Por outro lado, as questões contextualizadas apresentaram menos sensibilidade à origem social do candidato.

Em outro trabalho, Marcom e Kleinke (2021) confirmaram o resultado da relação entre desempenho em Física e NSE e, além disso, observaram que aproximadamente 70% dos itens de Física são considerados difíceis, a partir dos desempenhos dos estudantes. Complementando os resultados anteriormente expostos, Nascimento, Cavalcanti e Ostermann (2018) verificaram que itens qualitativos tendem a ser menos associados ao NSE dos candidatos, diferentemente dos itens quantitativos. Outro resultado de sua investigação diz que os fatores mais relacionados ao desempenho no ENEM foram o capital econômico (renda familiar) e o capital cultural institucionalizado (diplomas, títulos escolares e acadêmicos).

Os 3 artigos aqui expostos nos mostram a possibilidade de olhar para o conteúdo ou para os itens do ENEM com uma preocupação mais alinhada ao papel político do exame, investigando questões que possam ser menos reprodutoras de privilégios sociais. Coincidentemente, todos os artigos dessa subseção tratam de itens da prova de Ciências da Natureza, mais especificamente da Física. Pesquisas nesse sentido que abordem as outras áreas do conhecimento são importantes para que possamos pensar um ENEM menos sensível à classe social de seus candidatos e que democratize, de fato, o acesso ao Ensino Superior público brasileiro.

### **3.3.2 Desempenho individual**

A associação entre o desempenho escolar e a origem social dos estudantes não é novidade no campo educacional. Todos os artigos aqui classificados preocupam-se em investigar o poder da desigualdade social no desempenho no ENEM. Porém, um trabalho analisa também a evolução temporal dessa desigualdade (LIMA JUNIOR; FRAGA JUNIOR, 2021), enquanto outros dois trazem contribuições olhando para os fatores mais determinantes na diferença entre os desempenhos (FIGUEIRÊDO; NOGUEIRA; SANTANA, 2014; LUCENA; DOS SANTOS, 2020).

Figueirêdo, Nogueira e Santana (2014) apresentam resultados que estão baseados em modelos econômicos de igualdade de oportunidades. Os autores apontam que, independentemente da região onde o candidato reside, os alunos de baixo NSE estão aquém dos demais. Suas chances de alcançar uma nota no ENEM acima da média é quatro vezes menor que de um candidato de alto NSE, o que leva a uma probabilidade de 0,27% contra 35% de chances de se figurar entre as 5% melhores notas, respectivamente. Com relação aos fatores mais incisivos na desigualdade educacional, os autores destacam que a renda, seguida da escolaridade da mãe e do pai, são as três variáveis mais incisivas (o tipo de escola, a raça, a localização da moradia e o sexo também são variáveis incisivas). Por fim, observam que as pessoas do sexo feminino e não-negras apresentam vantagem no desempenho.

Essa associação entre renda familiar e alto desempenho no ENEM também foi reforçado pelo trabalho de Lucena e Santos (2020). Seu modelo, que considera informações socioeconômicas de renda e escolaridade dos pais, além de informações sobre raça e tipo de escola, apresentou um poder explicativo de 30% sobre o desempenho geral no ENEM. Olhando para cada tipo de variável especificamente, os candidatos que se sobressaíram aos demais foram aqueles: (i) cujos pais são graduados ou pós-graduados; (ii) que nunca trabalharam; (iii) que se autodeclararam brancos; (iv) que estudaram a maior parte do Ensino Médio em escolas particulares.

Ainda sobre poder explicativo de modelos estatísticos para as notas no ENEM, Lima Junior e Fraga Junior (2021) encontraram que aproximadamente 35% da nota individual na prova de Ciências da Natureza nos anos de 2012 a 2019 pode ser predita conhecendo a renda familiar e a escolaridade dos pais dos candidatos. Além disso, olhando para o contexto escolar, os autores encontram que o efeito social da escola (que leva em consideração a média da posição social dos alunos da mesma escola) apresentou-se quatro vezes maior que o efeito social a nível individual. Ou seja, a posição social média dos discentes de uma escola pareceu ter mais peso sobre o desempenho em Ciências da Natureza que as posições sociais individuais. No entanto, ao olhar para a evolução temporal das influências da origem social (tanto a nível individual quanto a nível escolar), os autores observaram que há uma tendência de queda no período analisado (2012 a 2019), alcançando o seu menor valor no ano de 2017.

As contribuições dos 3 artigos aqui elencados mostram como a prova do ENEM ainda continua sob forte influência da origem social dos candidatos. Além disso, mais do que a origem social individual, a origem social dos colegas de classe também possui um poder relevante para o desempenho no exame. Os privilégios sinalizados por essas pesquisas mostram o quanto as vias de acesso ao Ensino Superior público brasileiro estão mais facilitadas a certa parcela da população, enquanto para outra, parece algo distante e quase impossível. Além disso, os resultados expostos aqui reforçam a necessidade da luta pela criação e manutenção de políticas públicas de ações afirmativas de acesso à educação.

### **3.3.3 Desempenho agrupado**

Da mesma forma como os artigos citados na subseção anterior, os artigos dessa subseção estão preocupados em analisar alguns fatores e diferenças no desempenho escolar, porém, agora os candidatos estão sob certo tipo de agrupamento, seja a nível escolar, seja a nível municipal/estadual/regional. Na Tabela 1, é possível observar que há 3 trabalhos que foram classificados nesse tema (ALMEIDA et al., 2017; DUTRA et al., 2019; TORRES et al., 2020).

Almeida et al. (2017) estão preocupados em comparar os desempenhos das escolas de Alagoas (AL) com as demais escolas do Nordeste, sobretudo dos estados da Bahia (BA) e Pernambuco (PE) – estados com os melhores desempenhos em sua análise. Eles observaram que AL foi o estado com a maior queda de desempenho no período de 2013 a 2015 e, dentre as escolas públicas, no ano de 2015, foram as escolas públicas alagoanas que alcançaram os piores desempenhos na região. Observaram, também, que os investimentos em educação sofreram redução no mesmo período, enquanto o investimento em Sergipe permaneceu praticamente inalterado. Quando os autores lançam um olhar para as melhores escolas da BA e PE, percebem que as maiores diferenças no desempenho entre essas escolas e aquelas localizadas em AL estão nas provas de Redação e Matemática.

Além do tipo de escola, a região onde uma escola está localizada também possui influência no desempenho do estudante. Dutra et al. (2019), investigando o desempenho no ENEM dos Institutos Federais (IFs) do Brasil, observaram que os IFs localizados nas regiões Sul e Sudeste apresentaram as maiores médias de desempenho em relação aos outros institutos. Além disso, há diferença também se o IF está localizado em região urbana ou rural, com um maior desempenho para aqueles

localizados em área urbana. O NSE dos IFs se mostrou associado à diferença de desempenho alcançado entre eles e o porte escolar (número de alunos por escola) se mostrou estatisticamente significativo apenas para a prova de Matemática. Ademais, em todas as regiões do Brasil, as maiores médias de desempenho foram nas provas de Ciências Humanas e de Redação. Os autores afirmam que, ainda que haja uma diferença com relação à localização do IF (seja com relação ao estado, seja com relação à área), essa diferença é pequena e suspeitam que seja um resultado das políticas educacionais desse tipo de instituição escolar.

Seguindo essa linha de investigação cujo objetivo é comparar desempenho de diferentes instituições, Torres et al. (2020) comparam as escolas do Rio Grande do Sul (RS) e separam-nas em públicas e privadas. A partir desse recorte, encontram que a diferença na pontuação média desses alunos é de quase 60 pontos e que as notas mais dispersas e heterogêneas estão entre os estudantes oriundos de escolas privadas. Os autores lançaram um olhar para características socioeconômicas que também podem influenciar no desempenho e observaram que candidatos do sexo masculino e autodeclarados brancos alcançaram os melhores desempenhos. Em contrapartida, os candidatos do sexo feminino e/ou autodeclarados negros alcançaram os piores desempenhos. A escolaridade da mãe, a renda e o número de integrantes na família também se mostraram influentes no desempenho desses estudantes.

A partir dos resultados das investigações apresentados aqui, podemos observar como variáveis relacionadas ao contexto escolar e/ou regional influenciam também no desempenho, além da própria origem social do aluno. As diferenças podem ser consequências de políticas de investimento na educação, condições de trabalho, formação docente, bem como quantitativo de alunos por escola. Vale destacar, como foi possível observar, que o desempenho escolar das instituições se diferencia não só entre escolas de redes de ensino distintas, mas também entre instituições da mesma rede, o que nos leva a perceber que os privilégios educacionais ocorrem, também, a nível territorial.

#### **3.3.4 Desempenho individual e agrupado**

Por fim, nesta subseção, trataremos de algumas pesquisas cujas análises estão tanto no nível individual quanto no nível agrupado. Dos 14 artigos analisados, 5 foram classificados nesse tema (JUSTINIANO; QUEIROZ, 2021; MELO et al., 2021;

MORAES et al., 2022; NAVARRO et al., 2021; TRAVITZKI; FERRÃO; COUTO, 2016). Desses, apenas um trabalho está preocupado em comparar o desempenho em Ciências da Natureza com as outras provas do ENEM (por faixa de pontuação e por tipo de escola). Os demais trabalhos estão preocupados em relacionar características socioeconômicas e territoriais/demográficas ao desempenho no certame.

Justiniano e Queiroz (2021) analisaram dados de desempenho de alunos de escolas públicas do país que realizaram o ENEM no período que vai de 2012 a 2018, agrupando-os por escolas e zonas, e testando sua correlação com informações socioeconômicas. No período analisado, o número de alunos que não alcançaram ou alcançaram 450 pontos pelo menos em uma prova do ENEM diminuiu. No entanto, o número de candidatos que alcançaram 450 pontos em todas as provas (500 pontos na prova de Redação) não aumentou. Tal como em outros trabalhos, a renda familiar e a escolaridade dos pais mostraram correlação positiva com a pontuação no ENEM. Além disso, a proficiência nas provas do ENEM e a renda *per capita* se mostraram bem correlacionadas quando agrupadas por escolas ou em diferentes níveis de agregação territorial (município, estado, microrregião, região).

Um olhar para características da população de um município e sua relação com a nota no ENEM foi objeto de estudo de Melo et al. (2021). Além de verificar que os melhores desempenhos estão no eixo Sul-Sudeste e os piores desempenhos, no eixo Norte-Nordeste, os autores levantaram alguns fatores importantes para a nota no ENEM de um município. São eles: (i) percentual de alunos bolsistas em escola privada; (ii) renda das famílias residentes; (iii) alta escolaridade das mães dos candidatos; e (iv) maior proporção de estudantes autodeclarados brancos. O desempenho na prova de Redação se mostrou positivamente correlacionado com os fatores (i), (ii) e (iii) citados anteriormente. Além disso, os autores afirmam também, a partir de testes estatísticos, que alguns fatores regionais, não-socioeconômicos, influenciam no desempenho e na dispersão das notas dos municípios (MELO et al., 2021).

Seguindo nesse olhar para o desempenho de municípios, Moraes et al. (2022) olharam para o desempenho em Matemática de alunos de escolas públicas residentes em municípios de 50 a 500 mil habitantes e que prestaram o ENEM no ano de 2017. Quando comparadas as diferenças de raça e sexo, a diferença da variável correspondente ao sexo do candidato se mostrou maior que a de raça. De um lado, os candidatos autodeclarados brancos se sobressaíram, enquanto as candidatas

autodeclaradas pretas alcançaram os piores desempenhos, independentemente da faixa de renda. Entre os alunos de renda alta, os estudantes brancos alcançam notas maiores que estudantes negros, independentemente do sexo. Em seu modelo estatístico, os autores encontraram que a maior parte do impacto nas notas se dá por fatores externos ao controle de escolas e municípios. Ainda dentro desse modelo, os autores olharam para os 30 municípios mais iníquos com relação à raça e sexo e observaram que, em sua maioria, seus desempenhos estão acima da média. Já dentre os 30 municípios mais equitativos, apenas 2, tanto em relação à raça quanto em relação ao sexo, alcançaram um desempenho acima da média.

Passando de um olhar para os municípios e caminhando para um olhar a nível estadual, Travitzki, Ferrão e Couto (2016) focalizaram o desempenho em Linguagens e Códigos e em Matemática de quatro Unidades Federativas (UFs) (Alagoas, Maranhão, Rio Grande do Sul e São Paulo). Em seu modelo estatístico, observam que o desempenho das UFs acompanha a renda nominal *per capita*, ou seja, UFs de maior renda alcançam maiores médias. Com relação ao NSE das UFs, ele se mostrou correlacionado e significativo. A cor/raça autodeclarada branca ou amarela se mostrou como um fator de diferenciação nas notas, sobretudo nos estados de SP e RS. No que tange à variabilidade da nota, os autores encontram que entre 13% e 38% dessa variabilidade pode ser associado ao agrupamento ao nível de escola ou município, com valores mais altos para a nota em Matemática. Ou seja, a variabilidade da nota pode chegar a 38% em decorrência das diferenças entre as escolas e municípios.

Finalmente, Navarro et al. (2021) lançaram um olhar comparativo do desempenho dos estudantes de diversas redes de ensino em Ciências da Natureza (CN) e outras áreas do conhecimento que estão presentes na prova do ENEM. Ao olhar para distribuição de notas dentre os tipos de escolas, os autores observaram que a proporção de estudantes oriundos da rede pública decresce, enquanto essa proporção aumenta entre os alunos da rede privada de ensino quando caminhamos para as maiores notas. Essa informação está relacionada ao dado de que cerca de 60% dos alunos da rede estadual não alcançaram 600 pontos em CN, enquanto esse mesmo número é de apenas aproximadamente 20% entre os alunos da rede privada. Com relação às outras áreas de conhecimento, o desempenho em CN se sobressaiu ao de Matemática dentre os alunos das redes estadual e municipal de ensino. Se por um lado, na faixa de pontuação de 700 a 1000 pontos, a nota em CN foi maior que a nota em Ciências Humanas e Linguagens e Códigos, independentemente do tipo de

escola do candidato, por outro lado, na faixa de pontuação abaixo de 500 pontos, a média em CN foi a pior dentre todas as provas.

### 3.4 SÍNTESE DA REVISÃO

Para responder à questão de pesquisa 1 desta revisão da literatura, realizamos uma busca no Portal Periódicos por artigos cujo título contivesse as palavras ENEM ou “Exame Nacional do Ensino Médio” e que possuíssem, em qualquer campo, as palavras “Educação” e “Ensino”. Dessa busca, encontramos 364 artigos que foram submetidos à mineração de texto com auxílio da linguagem de programação *R* (SILGE; ROBINSON, 2017). Observamos que grande parte dos artigos estão mais preocupados com questões concernentes ao conteúdo, às questões e às provas do ENEM (seja objetiva, seja de Redação), do que ao papel político do certame ou a outras questões associadas ao exame.

Dentre os trabalhos que investigam as áreas componentes das Ciências da Natureza (Biologia, Física e Química), observamos que as análises tratam tanto de verificar como determinado assunto é abordado na prova quanto as qualidades psicométricas dos itens dessa prova (CORDEIRO et al., 2020; FERNANDES; CAMPOS, 2016; VIZZOTTO, 2022a, 2022b, 2022c; VIZZOTTO; MACKEDANZ, 2020). Alguns poucos artigos investigam as questões de Ciências da Natureza com um enfoque para as desigualdades socioeconômicas (KLEINKE, 2017; MARCOM; KLEINKE, 2021; NASCIMENTO; CAVALCANTI; OSTERMANN, 2018).

Com o objetivo de complementar a investigação aqui apresentada, e respondendo à questão de pesquisa 2, realizamos um segundo filtro com os 364 artigos dessa revisão, buscando aqueles que contivessem os termos “classe\*”, “reprod\*”, “desiguald\*” e “socio\*” e cujo periódico onde fora publicado tivesse índice h5 igual a 10, no mínimo. Chegamos a um número de 14 artigos, todos cujas análises foram realizadas a partir de técnicas estatísticas (tanto descritivas quanto inferenciais). Seus resultados foram classificados em 4 grandes temas: (i) *desempenho a nível individual*; (ii) *desempenho a nível agrupado*; (iii) *desempenho a nível individual e agrupado*; e (iv) *itens/conteúdo*.

De maneira geral, apesar do uso de técnicas de análise distintas e o emprego de modelos estatísticos diferentes, os trabalhos aqui analisados apontam para a perpetuação de privilégios sociais no ENEM. Seja a nível individual, seja sob alguma forma de agrupamento (escola, município, estado, região), o nível socioeconômico

ainda é um poderoso fator na nota alcançada por um candidato. No entanto, não apenas a origem social individual influencia a nota, mas também a origem social das pessoas que cercam o candidato (LIMA JUNIOR; FRAGA JUNIOR, 2021). Isso, entretanto, não retira a importância e o peso que o trabalho pedagógico exercido nas instituições escolares possui sobre o desempenho no certame, podendo até mesmo superar expectativas dada a posição social do corpo discente.

Ainda que as associações entre origem social e desempenho escolar estejam bem configuradas há muito tempo na literatura (desde a década de 1960), investigações empíricas que tratam de revisitar e refinar esse tipo de conhecimento continuam sendo publicadas nos periódicos mais citados das áreas de ensino e educação. Isso mostra que, para uma fração expressiva dos pesquisadores brasileiros, esse tipo de teste ainda pode ser considerado conhecimento novo – possivelmente pelos enquadramentos específicos que dão ao tema; ou porque, no mundo da vida comum, a ilusão meritocrática tem sido reavivada e reafirmada por movimentos conservadores e negacionistas a despeito da existência de evidência empírica contrária.

## 4 MÉTODOS

As informações necessárias para esta investigação foram obtidas dos microdados do ENEM e estão disponibilizados eletronicamente no site do INEP (2022). Toda análise presente nesta pesquisa foi desenvolvida no software de programação *RStudio* com o auxílio dos pacotes “*ca*”, “*plyr*”, “*dplyr*”, “*geobr*”, “*ggplot2*”, “*Hmisc*” e “*lme4*”.

Para dar início à análise estatística, reunimos os dados das cinco edições e realizamos dois processos de filtragem. Primeiramente, retivemos apenas aqueles indivíduos com nota nas cinco provas (Ciências da Natureza, Ciências Humanas, Linguagens, Matemática e Redação) e que continham informações sobre a escolaridade e ocupação dos responsáveis, a renda familiar e os dados da escola onde cursou o Ensino Médio, como o código da escola, o tipo de dependência administrativa (Federal, Estadual, Municipal e Privada) e a Unidade da Federação onde está localizada. Depois, selecionamos apenas aquelas escolas que tiveram participação consistente no ENEM, ou seja, possuíam pelo menos 30 estudantes concorrendo na prova. Após essas etapas, obtivemos uma amostra de aproximadamente 3,8 milhões de participantes ( $n = 3.775.159$ ), distribuídos em cerca de 15 mil escolas ( $n = 15.548$ ).

De posse das informações socioeconômicas dos indivíduos (i.e., escolaridade e ocupação dos responsáveis e renda familiar), realizamos uma Análise de Correspondência Múltipla (ACM) Ajustada (GREENACRE, 2007). A saber, esse é um método bastante empregado em pesquisas no meio acadêmico para análises de dados categóricos (dados tabulados em níveis, como as variáveis que dizem respeito à origem socioeconômica - cor, sexo, raça, faixa de renda, entre outros). Do mapa gerado pela ACM Ajustada, pudemos mensurar a posição social dos indivíduos na estrutura social dos candidatos do certame. A partir disso, calculamos a posição social da escola como a média aritmética simples das posições sociais do corpo discente de cada escola.

Para testar o efeito-escola nas provas do ENEM, bem como encontrar as escolas e Unidades da Federação que fazem a diferença, utilizamos um modelo de Regressão Linear Multinível (HOX, 2010). Isto é, testamos qual a porcentagem da nota na avaliação pode ser explicada simplesmente pelas variáveis contextuais do modelo e qual percentual da variabilidade pode ser imputada às diferenças entre

escolas. Já é sabido que, na prova de Ciências da Natureza, cerca de 35% do desempenho individual deve-se à origem social do indivíduo (LIMA JUNIOR; FRAGA JUNIOR, 2021). Mas qual é o poder explicativo da origem social sobre o efeito da escola como um todo?

Para encontrar as escolas que fazem a diferença, lançamos um olhar para os interceptos das escolas presentes no modelo, comparando-os com a nota média geral da amostra. Esse mesmo procedimento foi aplicado às Unidades Federativas, buscando visualizar quais são aquelas que estão no topo dos *rankings* usualmente divulgados e quais são aquelas que despontam quando controlamos a influência da origem social sobre a nota. Vamos agora a uma explicação da análise multivariada utilizada nesse exame: a análise de correspondência.

#### 4.1 ANÁLISE DE CORRESPONDÊNCIA

Em todas as suas principais versões (simples, múltipla e conjunta), a Análise de Correspondência tem como ponto de partida uma **tabela de contingência**, que apresenta o número de ocorrências da tabulação cruzada dos níveis de duas ou mais variáveis categóricas. A saber, todas as variáveis codificadas em níveis (ordenados ou não) são consideradas categóricas (cor, sexo, raça, faixa de renda, entre outros). A Tabela 2 apresenta o resultado da tabulação cruzada entre as variáveis de escolaridade do pai dos participantes e seu respectivo desempenho em Ciências da Natureza. Para apresentação em uma tabela, a escala de desempenho foi codificada em três faixas: N1 (até 450 pontos), N2 (de 450 a 520 pontos) e N3 (acima de 520 pontos).

As células da tabela de contingência exibem a quantidade de casos que se encaixam simultaneamente nos níveis descritos em cada linha e coluna (i.e., 83.140 participantes alcançaram até 450 pontos e, ao mesmo tempo, possuem pai que nunca estudou). Observe que as tabelas de contingência também informam ocorrências totais por linha, coluna e geral. Essas informações serão importantes para todas as inferências posteriores. A propósito dessas inferências, geralmente empregamos o **teste chi-quadrado** para testar a associação entre as variáveis em uma tabela de contingência (LIMA JUNIOR, 2023). Ele estima a probabilidade de ser verdadeira a hipótese de independência entre as variáveis da tabela (nesse caso, a probabilidade de a escolaridade do pai não guardar nenhuma relação com o desempenho na prova de Ciências da Natureza).

Tabela 2 – Tabulação cruzada da escolaridade do pai (linhas) contra as faixas de desempenho em Ciências da Natureza (colunas) no ENEM no período analisado

	<b>Até 450</b>	<b>De 450 a 520</b>	<b>Acima de 520</b>	Total da linha
Nunca estudou	83.140	49.526	19.705	152.371 (04,0%)
Alfabetização	334.252	242.788	127.775	704.815 (18,7%)
Fundamental 1	252.881	214.207	149.304	616.392 (16,3%)
Fundamental 2	214.181	196.697	160.917	571.795 (15,1%)
Médio	347.973	380.055	440.682	1.168.719 (31,0%)
Graduação	53.995	79.597	212.596	346.188 (09,2%)
Pós-Graduação	24.080	37.830	152.978	214.888 (05,7%)
Total da coluna	1.310.502 (34,7%)	1.200.700 (31,8%)	1.263.957 (33,5%)	3.775.159

Fonte: elaborado pelos autores

As **frequências relativas** são proporções obtidas ao dividir as frequências observadas por seus totais. Como cada célula tem um total de linha, um total de coluna e o total geral, há várias frequências relativas. As mais importantes já foram publicadas entre parênteses na Tabela 2 e resultam da razão entre os totais de linha e de coluna pelo total geral. Elas permitem avaliar as **frequências esperadas** em cada uma das células sob a hipótese de que as variáveis categóricas não estejam associadas.

Se as variáveis categóricas da tabela não estiverem estatisticamente associadas, seus níveis serão eventos independentes e a probabilidade de pertencimento a cada célula da tabela será igual ao produto das suas respectivas frequências relativas. Portanto, podemos obter as **frequências esperadas** em cada célula ao multiplicar as frequências relativas dos seus totais de linha e coluna pelo total de observações na amostra. Por exemplo, a frequência esperada para a primeira célula será igual ao produto das suas frequências relativas (4,0% e 34,7%) pelo tamanho total da amostra (3.775.159), o que resulta em 52.399,20. O mesmo cálculo pode ser feito para cada célula.

Enfim, é possível avaliar a probabilidade de não-associação das variáveis da tabela examinando o quanto as frequências observadas desviam-se das frequências esperadas sob a hipótese nula. Para isso, é conveniente definir uma quantidade que acumule os desvios quadráticos entre as frequências observadas e esperadas (GREENACRE, 2007):

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Nessa expressão,  $O_{ij}$  é a frequência observada e  $E_{ij}$  é a frequência esperada na célula  $ij$ , onde  $i$  e  $j$  representam a linha e a coluna da tabela, respectivamente. A frequência esperada  $E_{ij}$ , por sua vez, deve ser obtida pelo produto das respectivas

frequências relativas e o total de elementos da amostra ( $E_{ij} = f_i f_j N$ ). Quanto maiores forem os desvios entre as frequências observadas e esperadas, maior será o resultado do somatório e menos provável será a hipótese de independência entre as variáveis da tabela.

#### 4.1.1 Definindo alguns conceitos fundamentais

**Perfil** é o nome dado a um conjunto de frequências relativas dos dados da tabela de contingência. Podemos calculá-lo tanto para as linhas quanto para as colunas, obtendo os perfis-linha e os perfis-coluna. Para calcular o perfil, devemos dividir as frequências observadas em cada linha (ou coluna) pelo total dessa linha (ou coluna). Os perfis-linha da Tabela 2 foram calculados e publicados na Tabela 3. As linhas dessa tabela podem ser interpretadas como coordenadas de vetores no espaço-perfil, também chamado *simplex*. Ou seja, os perfis são representados por uma **posição** no espaço. Observe que, na Tabelas 3, há um perfil-linha médio que localiza, no *simplex*, a média ponderada das posições marcadas por todos os demais perfis-linha. Por ser uma média ponderada de posições, o perfil médio pode ser interpretado como um centroide (ou centro de massa).

Além de uma posição (definida por suas coordenadas no *simplex*), dizemos que cada perfil tem uma **massa**, que está relacionada ao grau relativo de ocupação desse perfil. Perfis muito “massivos” carregam mais casos que os demais. Para obtermos a massa dos perfis-linha a partir da Tabela 2, devemos efetuar a divisão do total de cada linha pelo total geral da amostra. Analogamente, obteríamos as massas dos perfis-coluna dividindo o total de cada coluna na Tabela 2 pelo total geral.

Tabela 3 – Perfis-linha da escolaridade do pai (linhas) contra as faixas de desempenho em Ciências da Natureza (colunas) – vide Tabela 2

	Até 450	De 450 a 520	Acima de 520	Massa
Nunca estudou	0,546	0,325	0,129	0,040
Alfabetização	0,474	0,344	0,181	0,187
Fundamental 1	0,410	0,348	0,242	0,163
Fundamental 2	0,375	0,344	0,281	0,151
Médio	0,298	0,325	0,377	0,310
Graduação	0,156	0,230	0,614	0,092
Pós-Graduação	0,112	0,176	0,712	0,057
Perfil-linha médio	0,347	0,318	0,335	

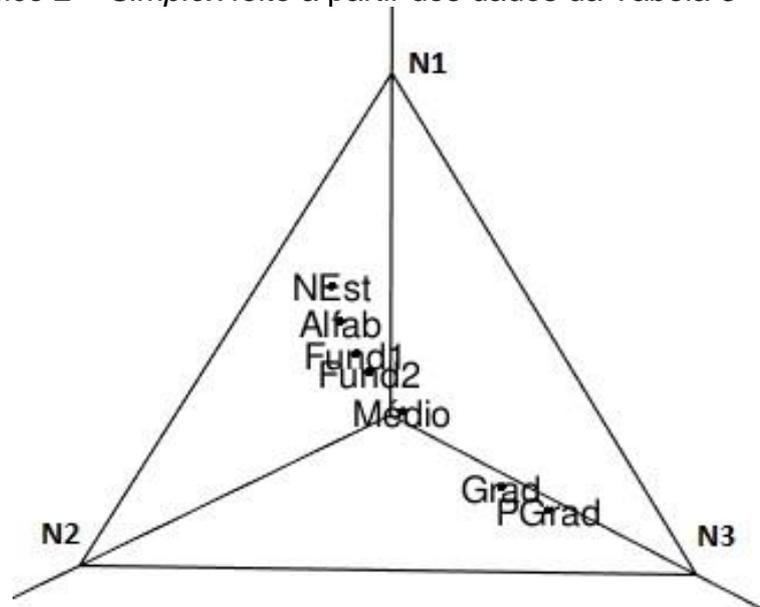
Fonte: elaborado pelos autores

Como os vetores-linha da Tabela 3 têm três coordenadas, sua representação gráfica é relativamente simples. Essa interpretação vetorial dos perfis implica que categorias semelhantes em termos das suas frequências relativas serão

representadas por pontos próximos no espaço-perfil. Observe também que, por construção, as coordenadas de todos esses vetores-perfis são tais que, somadas, produzem uma unidade. Portanto, os vetores-linha estão restritos a uma região triangular com 3 dimensões (*simplex*) e representada no Gráfico 2. O *simplex*, ou espaço-perfil, tridimensional pode ser representado em duas dimensões sem que haja perda de informação. No nosso caso, sua representação bidimensional terá forma de um triângulo, onde cada vértice representaria uma coluna da Tabela 3. Quando os perfis têm mais de três dimensões, a redução do *simplex* a um mapa bidimensional não é perfeita, mas aproximada.

No Gráfico 2, os vértices N1, N2 e N3 representam os intervalos de notas: N1, de até 450; N2, de 450 a 520; N3, a partir de 520. Como podemos observar, as categorias de escolaridade do pai foram plotadas e não os indivíduos em si. Ou seja, ao invés de representarmos todos os participantes no *simplex* (o que seria possível, mas dificultaria a visualização), plotamos apenas as categorias onde os participantes se encaixam, utilizando os perfis-linha como coordenadas. Porém, é importante termos em mente que as posições das categorias correspondem à média da posição dos indivíduos e que esses indivíduos estão geralmente dispostos de forma muito dispersa em torno das suas respectivas categorias.

Gráfico 2 – *Simplex* feito a partir dos dados da Tabela 3



Fonte: elaborado pelos autores

Como os três níveis de proficiência em ciências da natureza ficaram representados pelos vértices, e como parece haver uma relação mais ou menos linear entre escolaridade do pai e desempenho escolar do filho, os perfis-linha desenharam

uma curva no *simplex*. Em alguns casos, a curvatura é bastante acentuada, como se fosse uma ferradura. A saber, esse efeito está intimamente relacionado à geometria do *simplex* e é geralmente por ferraduras que variáveis linearmente associadas são representadas.

A analogia mecânica da análise de correspondência pode ser aprofundada. Até agora, já sabemos que cada perfil tem posição e massa. Portanto, à semelhança da noção de variância, podemos definir a dispersão dos dados como o somatório do produto das massas dos perfis pelo quadrado das suas respectivas distâncias ao centro de massa. Essa quantidade será chamada **inércia**:

$$\text{Inércia} = \sum_i m_i d_i^2$$

Nessa equação,  $m_i$  é a massa do  $i$ -ésimo perfil-linha,  $d_i^2$  é o quadrado da sua distância ao centroide e o somatório é feito sobre todas as linhas.

#### 4.1.2 Redução da dimensionalidade

No exemplo anterior, percebemos que o *simplex* pode ser representado por um triângulo de 2 dimensões quando temos uma tabela com três colunas. Nesse caso, os vértices do triângulo serão as categorias das colunas da tabela. Como nosso limite visual é tridimensional, representar quatro ou mais dimensões em um gráfico será desafiador. Portanto, tabelas com mais de três linhas ou colunas serão mais difíceis de representar graficamente a menos que empreguemos um método capaz de **reduzir a dimensionalidade do espaço-perfil**. Há, no entanto, um preço a pagar, pois a redução de dimensionalidade acarreta perda de informação. Sendo assim, um dos principais desafios da análise de correspondência é encontrar a menor quantidade de dimensões capazes de representar os dados da maneira mais verossímil possível. Esse critério de otimização corresponde à **maximização da inércia explicada**.

A Tabela 4 apresenta a tabulação cruzada entre escolaridade do pai (linhas) contra a escolaridade da mãe (colunas). A dimensão do espaço-perfil é determinada pelo número de categorias menos um. Ou seja, se fôssemos representar o espaço-perfil dessa tabela (com 7 níveis), seriam necessárias 6 dimensões! Portanto, para que possamos visualizar as distâncias e a dispersão dos perfis no espaço, é preciso fazer uma redução de dimensionalidade. Para decidirmos quantas dimensões reter, usualmente utilizamos um ou os dois dos critérios a seguir: máximo percentual explicado (critério de Catell) e critério de interpretabilidade (LIMA JUNIOR, 2023), ou

seja, escolhemos aquelas dimensões que mais explicam a inércia dos dados e/ou que possam ser interpretados.

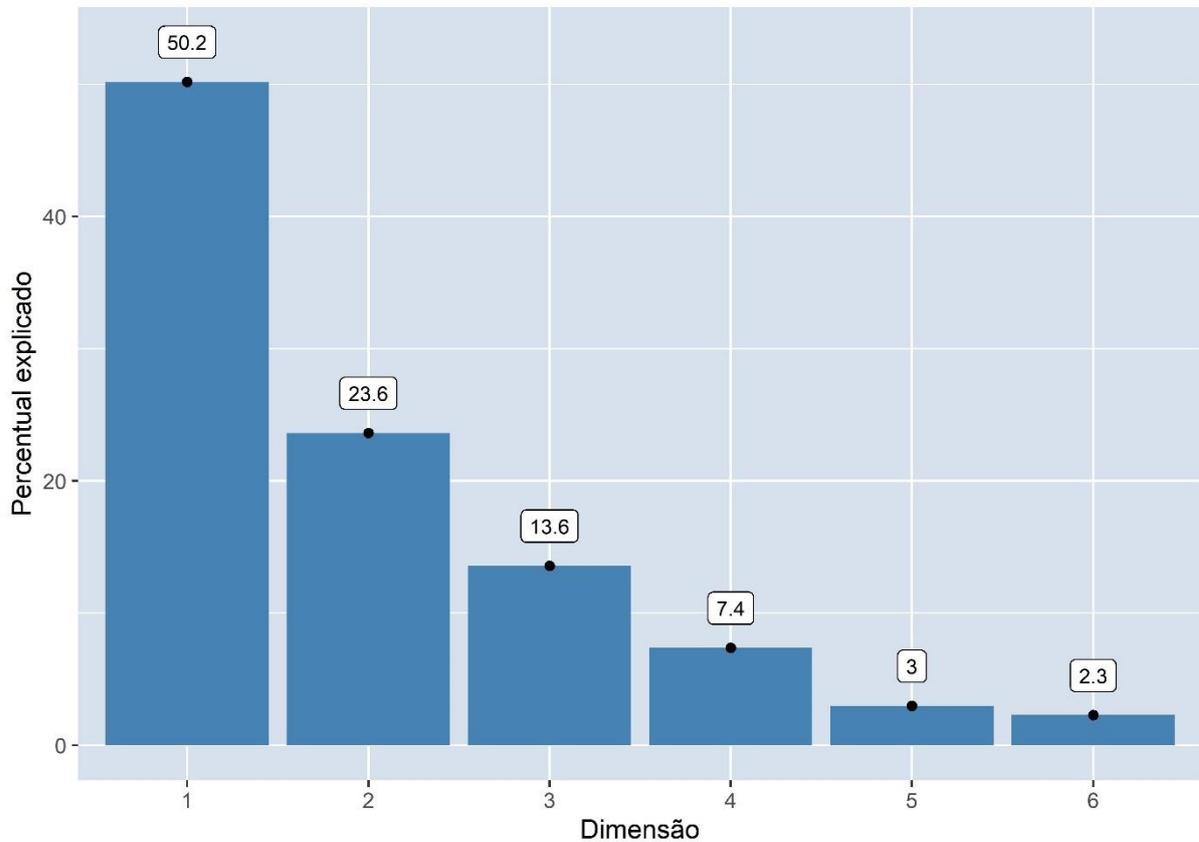
Tabela 4 – Tabulação cruzada da escolaridade do pai (linhas) contra a escolaridade da mãe (colunas) entre os participantes do ENEM no período analisado

	M.NEst	M.Alfab	M.Fund1	M.Fund2	M.Médio	M.Grad	M.PGrad	Total
P.NEst	53.566	47.767	20.426	12.349	14.687	2.177	1.399	152.371
P.Alfab	21.732	292.033	123.002	97.157	134.234	21.593	15.064	704.815
P.Fund1	8.113	61.080	198.649	119.360	173.874	32.003	23.313	616.392
P.Fund2	3.937	37.273	82.048	174.652	200.576	43.846	29.463	571.795
P.Médio	3.342	38.413	82.289	145.425	635.518	160.970	102.753	1.168.710
P.Grad	479	3.900	8.692	18.512	114.762	130.263	69.580	346.188
P.PGrad	231	1.622	2.979	6.379	42.632	55.025	106.020	214.888
Total	91.400	482.088	518.085	573.834	1.316.283	445.877	347.592	3.775.159

Fonte: elaborado pelos autores

O gráfico para essa tomada de decisão é conhecido como *scree plot* e é utilizado em outras análises multivariadas, como análise fatorial e análise de componentes principais (método pertencente à mesma família da análise de correspondência). O Gráfico 3 representa o *scree plot* do percentual explicado por cada dimensão da análise de correspondência simples das escolaridades dos pais e das mães dos estudantes. Como é possível observar, as duas primeiras dimensões são as que possuem maior poder explicativo (cerca de 74%). Portanto, seguindo o critério de Catell, reteremos apenas as dimensões 1 e 2. Note que, com essa escolha, garantimos o máximo de inércia explicada com um número mínimo e interpretável de dimensões, sacrificando apenas 26% da informação que está distribuída ao longo das outras 4 dimensões.

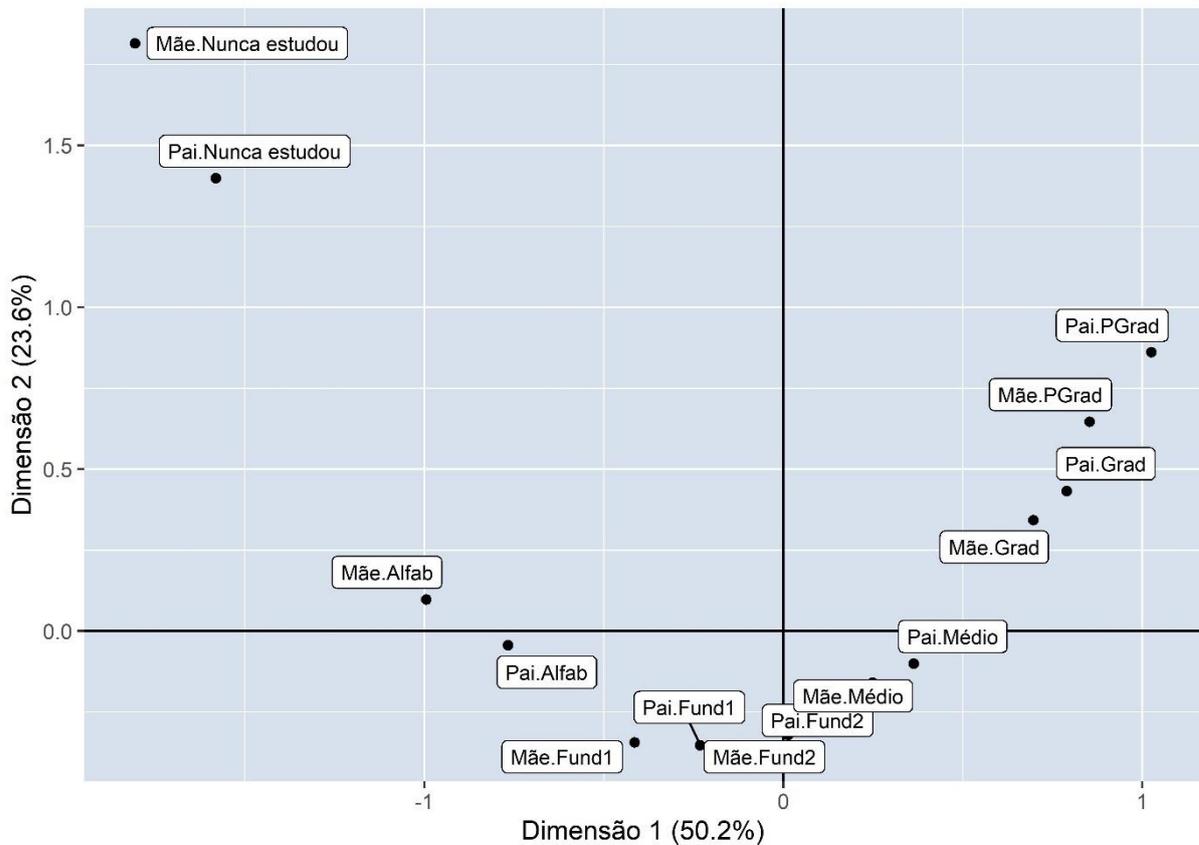
Gráfico 3 – Percentual explicado pelas dimensões da análise de correspondência simples das escolaridades dos responsáveis



Fonte: elaborado pelos autores

As duas dimensões principais do espaço reduzido encontram-se representadas no Gráfico 4. Note que a qualidade do mapa gerado é dada pela fração da inércia explicada por cada dimensão. Como é possível perceber, esse mapa bidimensional explica aproximadamente 74% da inércia dos dados, a máxima inércia explicada para uma representação bidimensional. As variáveis da escolaridade dos pais e das mães estão representadas no mapa gerado. Temos, portanto, um fenômeno interessante sendo representado aqui. Observe que os pontos que se referem à mesma escolaridade estão localizados próximos uns aos outros. Isso nos indica que os pares pai-e-mãe são geralmente compostos por pessoas de escolaridade semelhante. Esse fenômeno é chamado de endogamia das classes (BOURDIEU, 1998).

Gráfico 4 – Mapa simétrico obtido a partir da análise de correspondência simples dos dados da Tabela 4



Fonte: elaborado pelos autores

Outra observação importante que podemos fazer a partir do Gráfico 4 é a disposição das escolaridades ao longo da dimensão 1 (horizontal). Se olharmos da esquerda para a direita, veremos que há um crescimento no nível de escolaridade dos pais, indo desde “Nunca estudou” (ponto mais extremo à esquerda) até “Pós-Graduação” (ponto mais extremo à direita). Já em relação à dimensão 2 (vertical), vemos que, na parte de cima, estão reunidos os pontos mais extremos de escolaridade, enquanto os casos de escolaridade intermediária estão acumulados abaixo (Fundamental/Médio). Ter clareza sobre a interpretação dos eixos da análise de correspondência é fundamental quando desejamos testar como esses eixos se correlacionam com outras variáveis.

#### 4.1.3 Análise de Correspondência Múltipla

Como vimos até aqui, a análise de correspondência simples diz respeito à associação entre duas variáveis categóricas. Quando essas variáveis possuírem mais de três níveis, sua representação bidimensional demandará uma redução de dimensionalidade. Nesse caso, é preciso extrapolar a análise de correspondência

simples. A análise de correspondência múltipla (ACM) toma como ponto de partida uma tabela que representa, ao mesmo tempo, as relações entre todas as variáveis que desejamos analisar. A saber, há duas tabelas que cumprem essa função (1) a matriz indicadora e (2) a matriz de Burt. Para alguns propósitos, a análise dessas matrizes produz resultados equivalentes. Nesta pesquisa, escolhemos dar atenção à **matriz de Burt**, que pode ser definida como um conjunto de tabelas de contingência empilhadas.

Tabela 5 – Matriz de Burt dos dados de notas em Ciências da Natureza e Matemática

	CN até 450	CN de 450 a 520	CN acima de 520	MT até 450	MT de 450 a 550	MT acima de 550
CN até 450	1.310.502	0	0	703.737	468.813	137.952
CNT de 450 a 520	0	1.200.700	0	453.239	476.225	271.236
CN acima de 520	0	0	1.263.957	149.427	310.636	803.894
MT até 450	703.737	453.239	149.427	1.306.403	0	0
MT de 450 a 550	468.813	476.225	310.636	0	1.255.674	0
MT acima de 550	137.952	271.236	803.894	0	0	1.213.082

Fonte: elaborado pelos autores

Para gerar a matriz de Burt, devemos utilizar todas as variáveis categóricas na mesma tabulação cruzada. A Tabela 5 representa uma matriz desse tipo. Nela, tomamos somente duas variáveis – faixas de desempenho em Ciências da Natureza (CN) e Matemática (MT) –, mas deve ser fácil perceber que a matriz de Burt resulta do empilhamento de matrizes: na sua diagonal principal, encontram-se as tabulações cruzadas das variáveis contra elas mesmas; fora da diagonal principal, as tabulações cruzadas das variáveis entre si. A análise de correspondência múltipla pode ser definida como a análise de correspondência simples aplicada à matriz de Burt (GREENACRE, 2007). Nesse exame, utilizamos a análise de correspondência múltipla ajustada, que é um caso específico da ACM. Ela se baseia na matriz de Burt, porém foca nas submatrizes que estão fora da diagonal principal, melhorando, com isso, os percentuais explicados da inércia dos dados.

#### 4.1.4 A estrutura do espaço

A possibilidade de reduzir a dimensionalidade em uma análise de correspondência (simples ou múltipla) está intimamente relacionada às associações entre as variáveis. Quando pouco associadas, as variáveis não são bem

representadas por um conjunto reduzido de dimensões. Quando muito associadas, podemos imaginar que há uma estrutura (i.e., um sistema de relações) que determina os valores assumidos pelas variáveis. Podemos, por exemplo, pensar que a quantidade de itens da casa (geladeira, carro, freezer, computador...) são flexivelmente determinadas pelo “poder de compra” da família e, portanto, o poder de compra pode ser inferido pela análise de correspondência dessas variáveis. Em outras palavras, a análise de correspondência permite revelar a estrutura das relações determinantes (LIMA JUNIOR, 2023).

Na versão múltipla da análise de correspondência, o analista precisa separar suas variáveis categóricas em dois grupos: ativas e suplementares. Variáveis categóricas ativas participam da determinação da estrutura de relações que pretendemos mapear. Suas associações influenciarão a redução dimensional. Variáveis suplementares não determinam a estrutura das relações, mas são representadas no mapa. No algoritmo, a distinção das variáveis ativas é feita zerando a massa das categorias suplementares. Com massa nula, as variáveis suplementares não contribuem para a inércia dos dados. Uma vez que o critério de redução de dimensionalidade corresponde à maximização da inércia explicada pelas dimensões, a inclusão ou exclusão de variáveis suplementares não altera o mapa. Porém, ainda que não sejam estruturais, as variáveis suplementares nos ajudam na interpretação dos resultados.

#### **4.1.5 Interpretando o mapa simétrico: a posição social da escola**

Há vários *outputs* gráficos da análise de correspondência. De todas opções disponíveis, utilizamos os chamados “mapas simétricos”. Para interpretá-los, devemos projetar os pontos do gráfico (que representam níveis das variáveis categóricas) sobre os eixos. Essas projeções permitem dar sentido aos eixos. Para mensurar a posição social das escolas em nossa análise, submetemos as informações socioeconômicas dos participantes (escolaridade do pai e da mãe e renda familiar) à Análise de Correspondência Múltipla (ACM) Ajustada (GREENACRE, 2007), gerando o mapa presente no Gráfico 5.

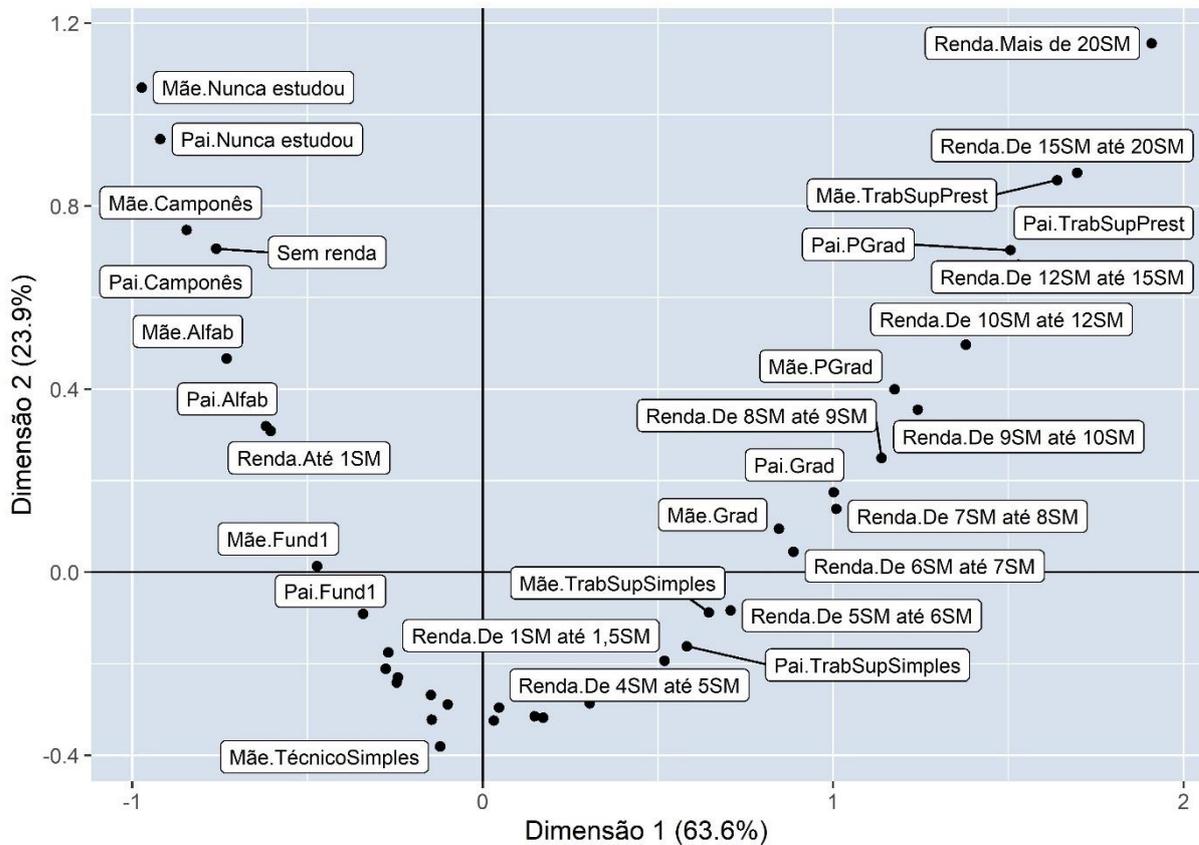
Ao olhar para o Gráfico 5, percebemos que todas as categorias estão organizadas em torno de uma ferradura, sugerindo uma direção de leitura do mapa. Além disso, o mapa é bidimensional, onde cada dimensão (vertical e horizontal) possui um percentual explicativo da dispersão dos dados analisados (inércia). Juntas, as

dimensões 1 e 2 explicam 87,5% dessa dispersão. Ou seja, na amostra de candidatos que analisamos, as variáveis de escolaridade e ocupação dos responsáveis, juntamente com a informação da renda familiar, são capazes de explicar quase toda a estrutura social. De fato, não é necessário utilizarmos todas as informações dos agentes sociais para montar um indicador socioeconômico. Basta selecionarmos aquelas variáveis que se correlacionam bem e que, de certa forma, as demais são correlacionadas a elas (como os bens possuídos e o “poder de compra”, por exemplo).

Analisando a dimensão 1 (horizontal), podemos observar algo muito interessante com relação a essa estrutura em forma de ferradura. Os últimos pontos do lado esquerdo dizem respeito às famílias com escolaridade mais baixa, pouca ou nenhuma renda e que possuem ocupações que podem ser exercidas com uma vida escolar mais breve. À medida que caminhamos para o centro, vemos que estão localizadas as famílias com uma trajetória escolar um pouco mais longa, alcançando títulos de nível Médio e Superior, renda por volta de 5 a 8 salários mínimos e trabalhos de nível técnico ou superior com pouco prestígio social. Já no lado direito, encontram-se as famílias com renda, escolaridade e ocupação de alto nível (mais de 10 salários mínimos, Pós-Graduação e ocupações prestigiadas).

Tendo isso em mente, podemos concluir que estamos diante de um parâmetro o qual diz respeito à posição social das famílias dos indivíduos. Ainda que pareça tentador classificar as regiões ao longo da dimensão 1 como classe popular, classe média e classe dominante, precisamos ter o cuidado e refletir se, de fato, os filhos das classes dominantes na sociedade brasileira prestam o ENEM. Todas as famílias representadas no Gráfico 5 são trabalhadoras, ocupando cargos que se diferenciam pelo tempo de vida escolar que requerem. No entanto, há diferenças que distinguem essas famílias. Talvez, aqui, o mais correto seja olhar para a linha horizontal como uma escala onde temos as classes e frações de classe popular e média. Diante disto, o valor de cada pessoa nessa linha horizontal nos serve como uma indicação de sua classe ou fração de classe, seu perfil socioeconômico.

Gráfico 5 – Mapa simétrico da ACM Ajustada das variáveis socioeconômicas do ENEM 2015-2019 (renda familiar, escolaridade e ocupação dos pais)



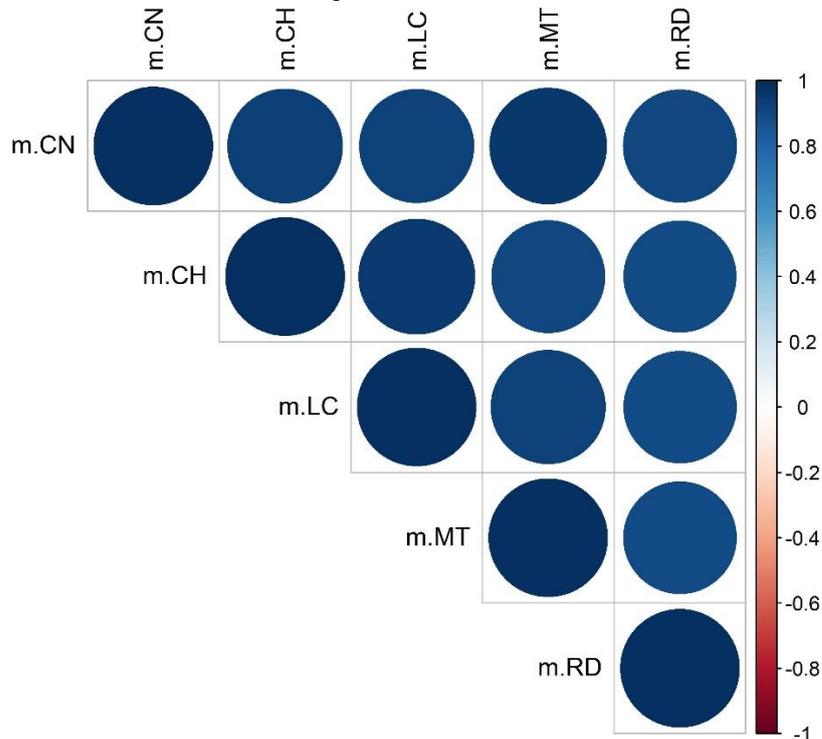
Ao agrupar os estudantes por escola e realizar uma média aritmética simples de suas origens sociais, encontramos um valor que indica a *posição social da escola*. Ou seja, somando os valores na dimensão 1 correspondentes a cada aluno de uma escola e dividindo pela quantidade de alunos dessa escola, chegamos a um valor na escala da posição social que representa, em média, o perfil socioeconômico do corpo discente. Essa informação é importante para que possamos avaliar em que medida a posição social da escola na estrutura das relações de classe determina seu desempenho médio nas provas do ENEM.

## 5 RESULTADOS

### 5.1 DESEMPENHO MÉDIO POR ESCOLA E POR UNIDADE DA FEDERAÇÃO NO ENEM

Para calcular o desempenho médio por escola, agrupamos os alunos e calculamos a média aritmética simples de suas notas nas 5 provas: Ciências da Natureza, Ciências Humanas, Linguagens, Matemática e Redação. Com isso, cada escola teve 5 notas representando o desempenho médio de seu corpo discente nas provas acima. Ao investigarmos as correlações entre os desempenhos nas áreas do ENEM, verificamos que as notas têm correlação muito próximas a 1 unidade (com  $p < 0.05$ ). Ou seja, a despeito da ideia vulgar de que, na escola, uns são bons nas disciplinas de Linguagens e Humanidades enquanto outros são bons nas disciplinas de Matemática e Ciências da Natureza, ter um bom desempenho em um desses componentes geralmente está associado a um bom desempenho nos demais. O Gráfico 6 apresenta a matriz de correlação entre as notas nas 5 áreas do ENEM. Nela, os círculos de cor azul escura indicam elevada correlação. Células pouco associadas, se existissem, seriam marcadas com círculos de cor branca.

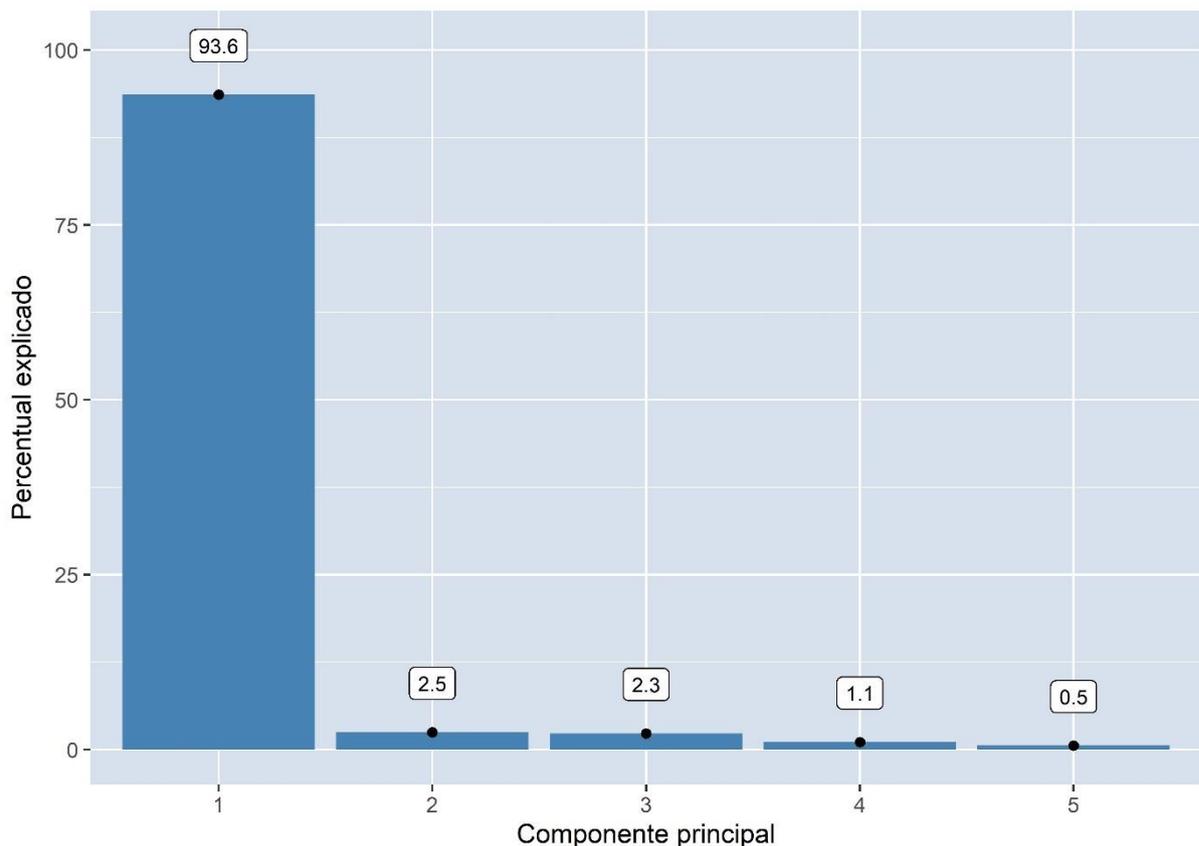
Gráfico 6 – Matriz de correlação entre as notas nas 5 áreas do ENEM



Fonte: elaborado pelos autores

Outro resultado que fortalece o argumento acima é obtido ao realizarmos uma Análise de Componentes Principais (ACP) (MAIR, 2018) sobre as notas nas provas. A saber, esse é um método de análise da mesma família da Análise de Correspondência Múltipla (ACM) explicada no capítulo 4 deste trabalho. Ou seja, os dois métodos operam uma redução de dimensionalidade, porém, enquanto a ACM se aplica a variáveis categóricas (qualitativas), a ACP se aplica preferencialmente a variáveis escalares (quantitativas) (LIMA JUNIOR, 2023). A mesma discussão sobre a escolha da quantidade de dimensões para a ACM também se aplica à escolha da quantidade de componentes principais que devemos reter. O Gráfico 7 contém os percentuais explicados por cada componente principal resultante da ACP aplicada às notas nas 5 provas do ENEM.

Gráfico 7 – Percentual explicado pelas componentes principais da análise de componentes principais das notas nas 5 provas do ENEM



Fonte: elaborado pelos autores

Seguindo o critério de Catell, vemos no Gráfico 7 que apenas a primeira componente principal é capaz de explicar 93,6% das notas nas provas. Ou seja, os desempenhos nas áreas do certame estão associados e podem ser tratados unidimensionalmente, i.e., podem ser tratados como uma única variável. Sabendo disso, calculamos a nota no ENEM como sendo a média das 5 áreas. Dessa forma,

obtivemos um valor do desempenho médio por escola no período analisado. O mesmo procedimento foi realizado para obtermos um valor médio por Unidade da Federação, a fim de ranquearmos, de maneira usual (i.e., calculando médias aritméticas simples e ordenando-as de forma decrescente), tanto as escolas com melhores desempenhos geral e por área, quanto os estados.

## 5.2 RANKINGS ESCOLARES DE CIÊNCIAS DA NATUREZA E DO ENEM

Sabendo a nota em Ciências da Natureza e no ENEM de cada escola, ordenamos de forma decrescente para montarmos um *ranking* da maneira mais usual (i.e., as escolas que obtiveram uma nota bruta maior ocupam as primeiras posições). Criamos um *ranking* com as 10 primeiras escolas, com o objetivo de verificar quais são aquelas que lideram as classificações. A coluna N nas Tabelas 6 e 7 representam a quantidade de alunos que prestaram o ENEM naquele ano.

Tabela 6 – *Ranking* das 10 primeiras escolas em Ciências da Natureza no período de 2015 a 2019

<b>Código da escola</b>	<b>N</b>	<b>Posição social da escola</b>	<b>UF</b>	<b>Tipo de escola</b>	<b>Nota em CN</b>
35399197	205	2,631	SP	Privada	707,48
23235543	68	1,491	CE	Privada	701,19
23246847	118	2,560	CE	Privada	694,86
31350664	1.410	3,178	MG	Privada	694,83
31349720	287	2,668	MG	Privada	687,11
23246880	301	2,886	CE	Privada	684,07
35141240	316	3,576	SP	Privada	683,91
22025740	604	2,951	PI	Privada	683,86
33040516	118	3,119	RJ	Privada	682,68
31351725	30	2,476	MG	Privada	680,46

Fonte: elaborado pelos autores

Tabela 7 – *Ranking* das 10 primeiras escolas no ENEM no período de 2015 a 2019

<b>Código da escola</b>	<b>N</b>	<b>Posição social da escola</b>	<b>UF</b>	<b>Tipo de escola</b>	<b>Nota no ENEM</b>
23246847	118	2,560	CE	Privada	748,84
31350664	1.410	3,178	MG	Privada	746,91
35399197	205	2,631	SP	Privada	746,81
22025740	604	2,951	PI	Privada	743,28
23235543	68	1,491	CE	Privada	740,77
31349720	287	2,668	MG	Privada	739,35
15167275	156	3,036	PA	Privada	739,06
31351725	30	2,476	MG	Privada	738,77
31004812	944	3,080	MG	Privada	732,41
23246871	225	2,229	CE	Privada	730,23

Fonte: elaborado pelos autores

Como é possível observar, no topo das Tabelas 6 e 7, temos escolas da rede privada cuja posição social é bastante elevada (basta consultar o Gráfico 4 e notar que os participantes com as maiores rendas e pais altamente escolarizados

encontram-se a partir da posição 1,5 na Dimensão 1). Porém, como podemos avaliar se as diferenças observadas nos *rankings* acima são significantes? Para isso, devemos olhar para o desvio-padrão da média das notas escolares. Para calculá-lo, basta dividirmos o valor do desvio-padrão das notas individuais pela raiz quadrada do número de alunos de cada escola observada. A média e o desvio-padrão da posição social dos indivíduos e das notas em Ciências da Natureza e no ENEM encontram-se na Tabela 8.

Tabela 8 – Média e desvio-padrão das variáveis de posição social, nota em Ciências da Natureza e nota no ENEM no período de 2015 a 2019

	<b>Nota no ENEM</b>	<b>Nota em CN</b>	<b>Posição social</b>
<b>Média</b>	523,56	491,31	0
<b>Desvio-padrão</b>	82,99	77,04	1,36

Fonte: elaborado pelos autores

O desvio-padrão da média é uma medida de dispersão que nos informa a variabilidade dessa média (LIMA JUNIOR, 2023). Por exemplo, se pegarmos a 5ª escola na Tabela 6 e calcularmos o seu desvio-padrão média, encontraremos um valor de aproximadamente 4,5. Como os intervalos de confiança são construídos por um valor médio mais ou menos o dobro do seu desvio-padrão da média, a nota dessa escola, se repetíssemos a medida nas mesmas condições com os mesmos participantes, estaria, com 95% de confiança, no intervalo de  $687,11 \pm 9$ . Ou seja, sua nota estaria entre 678,11 e 696,11. Observe que, a partir dessa flutuação estatística, essa escola, por sorte ou azar, poderia estar melhor ou pior colocada, respectivamente. Isso ilustra como as posições no *ranking* são sensíveis a termos estatísticos aleatórios. O leitor mais curioso pode efetuar o mesmo procedimento para os dados das outras escolas e observar como a configuração das classificações apresentadas nas Tabelas 6 e 7 mudariam devido às flutuações estatísticas.

Porém, ainda que, estatisticamente, algumas diferenças entre as escolas sejam pequenas e não significativas (fruto da aleatoriedade), a criação e a ampla divulgação dos *rankings* tendem a operar uma mágica performativa onde essas pequenas diferenças são exacerbadas, afastando escolas que estão, na verdade, muito juntas. As posições no *ranking* são vivenciadas como se exprimissem uma característica que está na natureza das escolas avaliadas, ou seja, sua qualidade escolar. Além disso, podemos observar, a partir das Tabelas 6 e 7, que as maiores notas em Ciências da Natureza e no ENEM são alcançadas por escolas cuja posição social é mais elevada, indicando uma possível presença do *efeito social da escola*

sobre o desempenho escolar. Consequentemente, os *rankings* tendem a produzir um efeito onde celebram como “mais competentes” as escolas onde se encontram os filhos das classes dominantes.

### 5.3 MODELO DE REGRESSÃO LINEAR MULTINÍVEL

Diante de tudo isso, podemos suspeitar de uma relação de predição entre as variáveis analisadas. Ou seja, a origem social, tanto no nível individual quanto no nível da escola, deve ser capaz de predizer, ao menos em parte, o desempenho na prova de Ciências da Natureza e no ENEM. Para testar essa hipótese e verificar seu impacto no ENEM, realizamos uma Regressão Linear Multinível (HOX, 2010). Nesse tipo de regressão, partimos do pressuposto de que alguns dados podem compartilhar semelhanças entre si, como por exemplo, alunos de uma mesma escola. Portanto, é de se esperar que alunos da mesma instituição tenham desempenhos no ENEM mais similares entre si que entre os demais participantes do certame. Nesses casos, a análise agrupa os dados que tendem a ter mais semelhança entre si e dizemos que há uma hierarquia entre esses dados. Ou seja, os alunos estão agrupados em escolas que, por sua vez, estão agrupadas em estados. Como temos 3 níveis dentro dessa hierarquia (estudante, escola e Unidade da Federação), construímos um modelo hierárquico de 3 níveis onde utilizamos como variável dependente a nota em uma das provas que compõem o ENEM ou a nota geral no exame; e como variável independente utilizamos informações da posição social individual e da escola, o tipo de dependência administrativa da escola e a Unidade da Federação onde a escola está localizada. A esse modelo, denominaremos Modelo Completo.

A modelagem linear hierárquica (ou modelagem multinível) permite que o investigador teste diversas possibilidades estatísticas, como, por exemplo, permitir que cada escola e cada UF tenha uma reta de regressão particular, ajustada aos seus dados agrupados, que poderá servir para comparar seu desempenho com relação à média geral. Ao empregarmos essa possibilidade, podemos verificar, a partir do modelo citado anteriormente, *quais escolas e quais estados que fazem a diferença*, ou seja, quais são as instituições e os territórios que mais se destacaram, mais agregaram valor ao seu público estudantil, quando as condições socioeconômicas estão controladas (ou seja, quando controlamos o efeito da origem social sobre o desempenho). Vale lembrar que, em uma regressão linear, o fato de adicionarmos variáveis independentes ao modelo faz com que o efeito das outras variáveis

permanença constante (controlado). Isso é o que conhecemos como cláusula *coeteris paribus* (LIMA JUNIOR, 2023).

Para podermos verificar a qualidade do modelo empregado nessa pesquisa, construímos um modelo de base (também hierárquico de 3 níveis), chamado Modelo Nulo, que não apresenta variáveis explicativas. O objetivo desse modelo é verificar a variância que está distribuída nos 3 níveis do modelo para que, ao rodar o Modelo Completo, sejamos capazes de calcular o percentual dessa variância que foi explicada. No fundo, o procedimento que estamos fazendo visa averiguar, no nível das escolas, quanto nosso modelo é capaz de explicar as notas obtidas no ENEM. A Tabela 9 traz as informações da variância de cada modelo distribuída nos 3 níveis para cada prova do certame, bem como para a nota geral.

Tabela 9 – Variância residual dos Modelos Nulo e Completo para cada prova do ENEM

	Modelo Nulo (sem variáveis explicativas)			Modelo Completo (com variáveis explicativas)		
	Aluno	Escola	Estado	Aluno	Escola	Estado
<b>CN</b>	3710	1848	173	3638,93	219,55	56,86
<b>MT</b>	7935,5	3890,7	369,2	7741,9	542,7	194,7
<b>CH</b>	4587	1580	289	4497,16	241,48	101,01
<b>LC</b>	3422,8	1180,7	254,7	3334,37	165,02	94,22
<b>RD</b>	21558,9	7332,7	547,8	21272,6	1198	558,3
<b>ENEM</b>	3643,9	2669,9	296,3	3508,2	320,8	148,9

Fonte: elaborado pelos autores

De posse dos dados presentes na Tabela 9, é possível calcularmos o percentual que cada nível do modelo é capaz de explicar com relação às variações nos desempenhos dos alunos. Por exemplo, se tomamos a prova de Ciências da Natureza (CN), no Modelo Nulo, a escola é capaz de explicar 32,2% da variação nas notas (basta dividir a variância no nível da escola pela soma das variâncias nos 3 níveis, ou seja, 1848 dividido por 5731). No entanto, quando as variáveis socioeconômicas e o tipo de dependência administrativa da escola são controlados, esse efeito cai para 5,6% (219,55 dividido por 3915,34). Já no caso da nota geral no ENEM (última linha da Tabela 9), vemos que o percentual explicado pela escola cai de 40,3% (Modelo Nulo) para 8% (Modelo Completo). Essa magnitude do efeito no nível das escolas é o que vemos na literatura como “efeito-escola” e está próximo dos valores encontrados na literatura (TRAVITZKI, 2013). Ou seja, controlando algumas

variáveis de contexto, as escolas explicam pouco (até 8%) da variação nas notas dos estudantes participantes do ENEM no período analisado. O efeito-escola calculado para as outras provas do certame ficaram entre 4,5% e 6,5%. A Tabela 10 traz o efeito-escola após o controle das variáveis de contexto para todas as provas do certame.

Tabela 10 – Efeito-escola para cada prova do ENEM calculado a partir do Modelo

Completo

	CN	MT	CH	LC	RD	ENEM
Efeito-escola	5,6%	6,4%	5,0%	4,6%	5,2%	8,0%

Fonte: elaborado pelos autores

Diante desse resultado, podemos suspeitar do uso do resultado no ENEM como indicador de qualidade escolar, já que, de fato, a escola tem pouco efeito sobre os desempenhos dos estudantes quando controlamos as variáveis contextuais (nível socioeconômico individual e da escola, tipo de dependência administrativa e Unidade Federativa da escola).

Outra informação interessante que podemos obter comparando as variâncias dos Modelos Nulo e Completo diz respeito ao percentual explicado pela inserção das variáveis contextuais (independentes) no modelo. Ou seja, se, no nível da escola, a variância era de 100 e, após a inserção das variáveis explicativas, a variância cai para 20, quer dizer que nosso modelo foi capaz de explicar 80% desse resultado ao nível da escola. No nosso caso, olhando para a prova de Ciências da Natureza, vemos que o percentual explicado, ao nível da escola, é de 88,1%. Isto significa que aproximadamente 88% da nota das escolas em Ciências da Natureza pode ser calculada a partir das variáveis contextuais utilizadas no modelo (i.e., nível socioeconômico individual e da escola, tipo de dependência administrativa e Unidade Federativa da escola).

Esse resultado corrobora a afirmação de que um bom desempenho no certame é um privilégio social. Quase todo o desempenho médio da escola está determinado pela faixa de renda e escolaridade das famílias que frequentam essa escola. Consequentemente, certas posições de prestígio no Ensino Superior estariam "destinadas" a um grupo privilegiado, enquanto a classe dominada estaria relegada aos cursos universitários de menor prestígio (BOURDIEU, 1998, 2014). A Tabela 11 apresenta os percentuais não explicados pelo Modelo Completo para cada prova do certame, bem como para a nota geral.

Tabela 11 – Percentual da variância residual não explicada pelo modelo para cada prova do ENEM

	CN	MT	CH	LC	RD	ENEM
<b>Variância não explicada</b>	11,9%	14,0%	15,3%	14,0%	16,3%	12,0%

Fonte: elaborado pelos autores

Devemos nos atentar para um cuidado importante ao interpretarmos esses resultados. Ainda que cerca de 12% da variação das notas em Ciências da Natureza não sejam explicados pelas variáveis do modelo, isso não implica que esse percentual se refere ao mérito da instituição escolar, como vemos relatados em alguns trabalhos (cf. TRAVITZKI, 2013). Devemos lembrar que há diversos fatores que influenciam no desempenho em uma prova e nosso modelo empregado não é capaz de, nem pretende, dar conta de todos as variáveis que podem exercer influência na nota no ENEM. Portanto, consideramos arriscado imputar o percentual não explicado da variância de nosso modelo como mérito da instituição escolar.

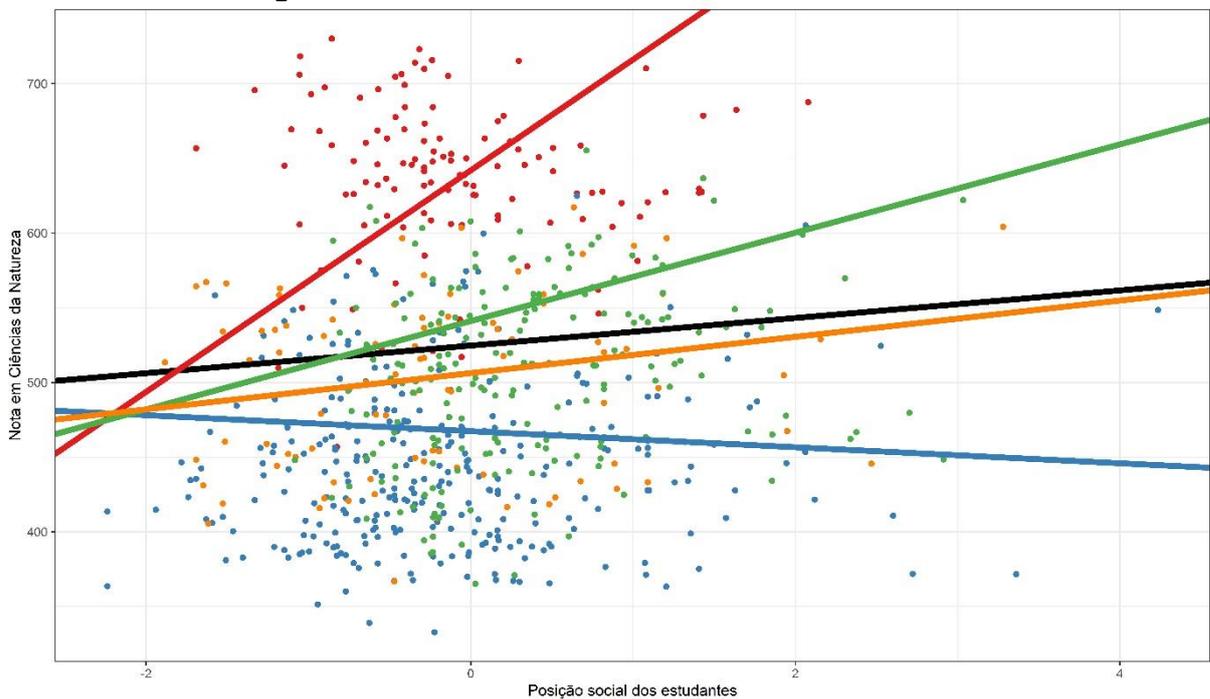
### 5.3.1 Escolas que fazem a diferença

Nesse contexto, as *escolas que fazem a diferença* são aquelas que, por seu trabalho pedagógico, contornam a tendência geral do efeito social sobre o desempenho escolar. Ou seja, são escolas que obtêm um desempenho acima do que seria esperado dado seu contexto. Para encontrar essas escolas, permitimos, em nosso Modelo Completo, que cada escola obtivesse uma reta de regressão ajustada aos resultados de seus alunos. Com isso, pudemos comparar o desempenho específico das escolas com a tendência geral da amostra utilizada em nossa pesquisa.

Para melhor compreensão dessa discussão, o Gráfico 8 apresenta um diagrama de dispersão da nota em Ciências da Natureza contra a origem social de estudantes de 4 escolas e, em preto, a reta de regressão linear que mais se ajusta aos dados da amostra. Como é possível perceber, cada público discente foi representado em cores distintas, bem como suas retas de regressão. A comparação feita para encontrar as escolas que fizeram a diferença diz respeito ao intercepto das escolas e o intercepto médio geral. A saber, o intercepto de um modelo de regressão linear representa a média da variável dependente, no nosso caso, a média de cada escola na prova de Ciências da Natureza. Além disso, diz respeito ao valor no eixo vertical (nota em Ciências da Natureza) quando o valor no eixo horizontal (posição social) é zero.

Vale observar que as 4 escolas representadas no Gráfico 8 possuem um corpo discente parecido entre si (com uma posição social média perto de zero), no entanto, obtiveram resultados diferentes no período analisado. Se olharmos para a reta vertical que passa pela origem do gráfico (ponto zero), veremos que a escola em vermelho obteve uma média de aproximadamente 650 pontos, enquanto as demais escolas obtiveram média de cerca de 550, 500 e 450 (em verde, laranja e azul, respectivamente). Quanto mais a escola se afasta para cima da reta em preto (média geral), maior foi sua nota e consideramos essas as escolas que fizeram a diferença, pois, retirando o efeito da origem social sobre o desempenho, essas escolas despontaram, alcançando resultados acima do que seria esperado dado seu corpo discente. Vale ressaltar que a presença de escolas que fazem a diferença (ou de escolas que performam abaixo do esperado por sua origem social) não invalida a tendência geral de associação entre posição social e desempenho da escola.

Gráfico 8 – Diagrama de dispersão da nota em Ciências da Natureza contra a origem social de estudantes de 4 escolas brasileiras



Fonte: elaborado pelos autores

A partir de um ordenamento decrescente dos interceptos (desempenhos médios) de cada escola, geramos uma lista das 10 escolas que mais fizeram a diferença no período analisado (2015 a 2019) tanto na prova de Ciências da Natureza quanto no ENEM como um todo. Essas e outras informações relevantes encontram-se resumidas nas Tabelas 12 e 13, respectivamente.

Tabela 12 – Listagem das 10 escolas que fizeram a diferença na prova de Ciências da Natureza no período de 2015-2019

Código da escola	N	Posição social escolar	UF	Tipo de escola	Nota em CN
22164219	38	-0,287	PI	Privada	618,95
35805555	807	0,667	SP	Privada	612,44
35480617	483	0,080	SP	Privada	610,86
23238933	433	-0,500	CE	Pública	602,80
35434619	287	-0,089	SP	Pública	600,20
23041790	419	-0,569	CE	Pública	600,04
35330401	367	-0,058	SP	Privada	591,10
35428905	312	0,060	SP	Pública	590,21
23235543	68	1,491	CE	Privada	589,77
23264071	585	-0,653	CE	Pública	589,31

Fonte: elaborado pelos autores

Tabela 13 – Listagem das 10 escolas que fizeram a diferença no ENEM no período de 2015-2019

Código da escola	N	Posição social escolar	UF	Tipo de escola	Nota no ENEM
22164219	38	-0,287	PI	Privada	689,85
35805555	807	0,667	SP	Privada	686,51
23041790	419	-0,569	CE	Pública	678,55
35434619	287	-0,089	SP	Pública	673,59
23238933	433	-0,500	CE	Pública	672,76
35480617	483	0,080	SP	Privada	665,17
31258393	86	0,199	MG	Privada	661,83
23047860	463	-0,766	CE	Pública	661,41
23255528	477	-0,823	CE	Pública	654,01
35330401	367	-0,058	SP	Privada	653,07

Fonte: elaborado pelos autores

Olhando para as Tabelas 12 e 13, observamos uma presença maior de escolas públicas nas primeiras posições. É interessante notar como os *rankings* anteriormente apresentados (Tabelas 6 e 7) e as listagens aqui presentes (Tabelas 12 e 13) diferem com relação às informações apresentadas. Enquanto, nos *rankings*, as 10 primeiras posições são ocupadas por escolas que atendem frações de classe média mais abastadas socialmente (com posições sociais que vão de 1.6 a 3.0), aqui predominam escolas de fração de classe mais baixa (com posições que vão desde -0.8 a 0.7). Vale observar também que, de acordo com o nosso modelo empregado de classificar as escolas (controlando os efeitos da origem social), nenhuma escola que estava presente nos *rankings* anteriores figuraram entre as primeiras posições nas Tabelas 12 e 13. Isso corrobora, mais uma vez, para que olhemos com suspeita para ranqueamentos de dados brutos, principalmente aqueles que são produzidos a partir de dados escolares, já que, pelo que foi apresentado nessa pesquisa, por trás do “mérito” escolar esconde-se privilégio social.

De fato, o *ranking* é muito sensível aos critérios de classificação empregados. Por essa razão, é sempre possível criar um critério alternativo que, confrontado ao

primeiro, revela o caráter arbitrário do ponto de vista que o sustenta. Ao realizar uma poderosa mudança no enfoque da classificação (controlando o nível socioeconômico e permitindo que cada escola tenha uma reta de regressão ajustada aos resultados de seus alunos), vemos como algumas escolas perdem suas posições enquanto outras despontam. A forma como instituições, pessoas e práticas são classificadas sempre traz consigo intenções mais ou menos dissimuladas que estão na base dos interesses objetivos de quem opera a classificação (BOURDIEU, 2020).

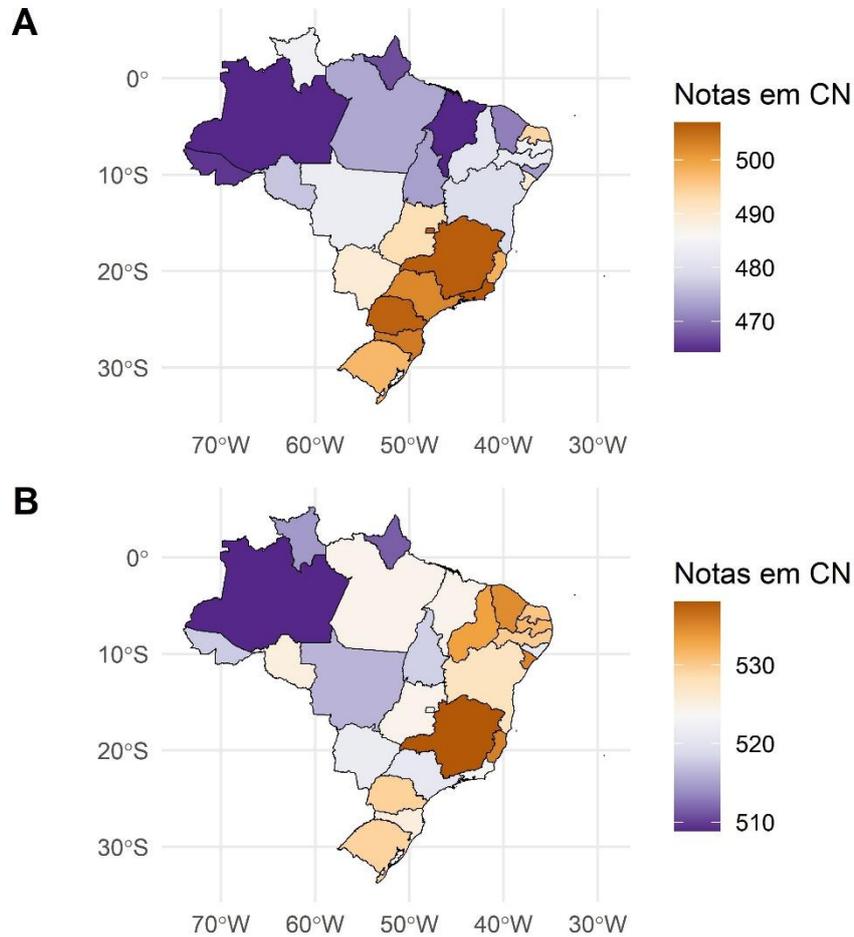
#### 5.4 RANKINGS ESTADUAIS E ESTADOS QUE FAZEM A DIFERENÇA

Os procedimentos para ranquear os estados com e sem o efeito social são similares aos descritos nas seções anteriores. Realizando uma média simples dos desempenhos brutos em Ciências da Natureza das escolas localizadas em cada Unidade da Federação, chegamos a um desempenho por estados. Ordenando-os de forma decrescente, obtemos então um *ranking* dos desempenhos alcançados por estados sob efeito da origem social na nota.

Assim como permitimos que cada escola tivesse uma reta de regressão que melhor se ajustasse ao seu corpo discente, também demos liberdade ao modelo para que ajustasse as melhores retas de regressão para cada Unidade da Federação. Portanto, similarmente ao que foi feito com as escolas, comparamos os estados com a média geral. Os estados que se sobressaíram, ou seja, as Unidades da Federação que fizeram a diferença são aquelas cuja nota média em Ciências da Natureza foi acima da nota média geral quando controlamos as variáveis contextuais (nível socioeconômico individual e escolar e o tipo de dependência administrativa da escola).

A Figura 8 apresenta o mapa do Brasil com escalas de cor mostrando dois tipos de classificações. A primeira figura (A) apresenta o *ranking* usual dos estados com base na nota bruta em Ciências da Natureza (com efeito do NSE sobre a nota). Já na segunda figura (B), podemos ver quais estados (e regiões) assumem as primeiras posições na classificação quando o efeito da origem social é controlado (i.e., estados que fizeram a diferença).

Figura 8 – Mapa do Brasil com os desempenhos dos estados em Ciências da Natureza de 2015 a 2019: A – sem controle do NSE sobre o desempenho; B – com controle do NSE sobre o desempenho



Fonte: elaborado pelos autores

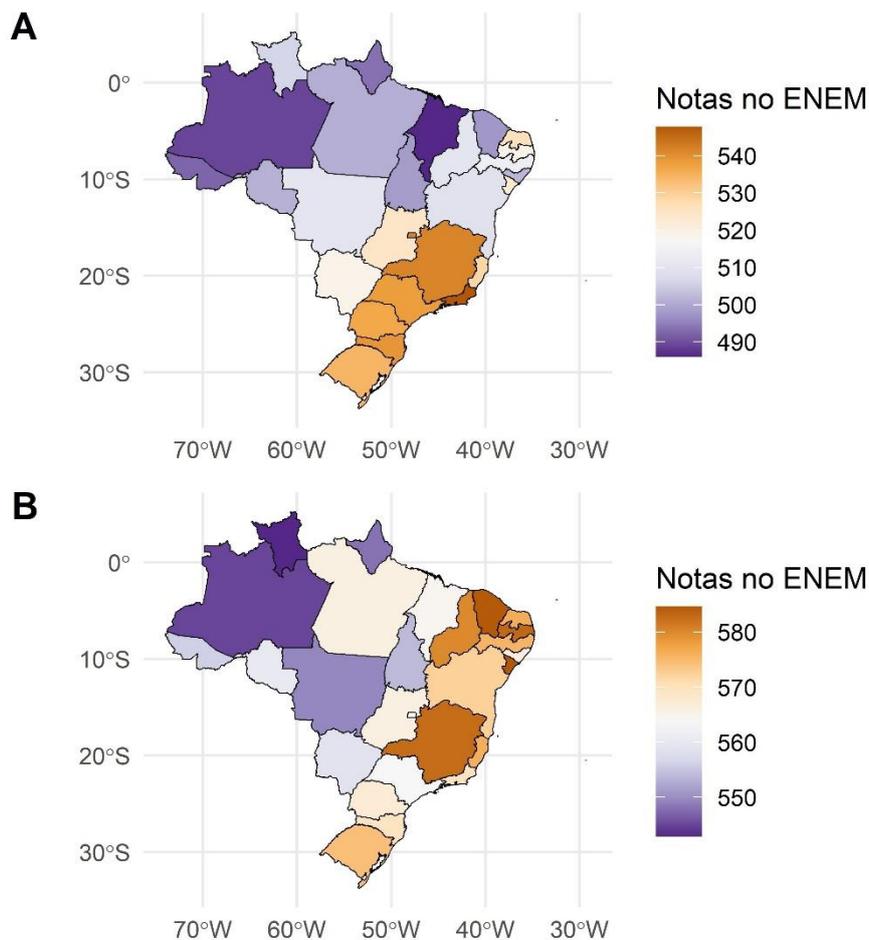
Na Figura 8A, os estados em laranja possuem as maiores notas, enquanto os estados em roxo possuem as menores. Vemos que as regiões Sul e Sudeste e o Distrito Federal se sobressaem e lideram o *ranking* do desempenho bruto (i.e., sem controlar o efeito da origem social das escolas). Esses resultados já eram esperados dada a distribuição desigual do capital em território nacional. Vemos, portanto, como há vantagem em estudar em instituições escolares que estão localizadas em certas partes do Brasil em detrimento de outras. Esses resultados estão de acordo com outros publicados na literatura (DUTRA et al., 2019; JUSTINIANO; QUEIROZ, 2021; PAULA; NOGUEIRA, 2018; ZACCHI; NEY; PONCIANO, 2016).

Já na Figura 8B, vemos que, controlando o efeito social sobre o desempenho escolar, alguns estados da região Nordeste despontam, assumindo as primeiras posições. Minas Gerais e Espírito Santo são os únicos estados da região Sudeste que permanecem nas primeiras colocações nesses *rankings*. Essa diferença vista entre

os gráficos corroboram para a visualização de como os *rankings* usuais, que usam o desempenho bruto como parâmetro classificatório, são sensíveis à origem social. Esses resultados nos fazem refletir sobre o papel dos *rankings* na sociedade e como certas instituições e regiões são (des)valorizadas pela forma usual como são classificadas a partir de seu desempenho no certame.

A Figura 9 apresenta, tal como a Figura 8, o mapa do Brasil com escalas de cor mostrando dois tipos de classificações: o *ranking* usual dos estados com base na nota bruta no ENEM (A) e quais estados (e regiões) assumem as primeiras posições na classificação quando o efeito da origem social é controlado (B) (estados que fizeram a diferença).

Figura 9 – Mapa do Brasil com os desempenhos dos estados no ENEM de 2015 a 2019: A – sem controle do NSE sobre o desempenho; B – com controle do NSE sobre o desempenho



Assim como a Figura 8, a Figura 9 mostra que, sem controlar o efeito das variáveis contextuais, os estados que assumem as primeiras posições no *ranking* do ENEM

estão localizados no eixo Sul-Sudeste, com exceção do Distrito Federal (Figura 9A). No entanto, quando controlamos essas variáveis contextuais (posição social individual e média escolar e a dependência administrativa da escola), vemos que a região Nordeste passa a ter tons mais alaranjados, o que significa que estão entre as maiores notas no exame. Minas Gerais e Espírito Santo também foram os únicos estados da região Sudeste que permaneceram entre as primeiras colocações quando controlados os efeitos da origem social.

## 5.5 SÍNTESE E DISCUSSÃO DOS RESULTADOS

Neste capítulo, vimos que as notas nas 5 provas do ENEM possuem uma alta correlação. Diante disso, e do resultado na Análise de Componentes Principais, calculamos a nota no ENEM como uma média aritmética simples das 5 provas. Para o desempenho médio por escola em cada prova e no ENEM como um todo, agrupamos os candidatos que frequentaram a mesma escola e calculamos a média de suas notas. O mesmo procedimento foi realizado para chegarmos aos desempenhos por Unidade da Federação.

Quando ordenamos as notas brutas, seja por escola, seja por estado, vimos que as primeiras posições eram ocupadas por instituições escolares cuja posição social eram as maiores da amostra. Isso vai de encontro ao entendimento de que a cultura escolar, enquanto cultura dominante, será mais familiar e, portanto, mais fácil de aprender, por aqueles que estão mais próximos dela, ou seja, os filhos das classes dominantes (BOURDIEU; PASSERON, 2010). No caso dos estados, os primeiros colocados foram aqueles cujo IDH são os maiores no país (UNITED NATIONS DEVELOPMENT PROGRAMME; FUNDAÇÃO JOÃO PINHEIRO; INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2013).

Portanto, sabendo da influência da origem social nos desempenhos escolares, decidimos averiguar, por meio de modelos de Regressão Linear Multinível (HOX, 2010), o tamanho do efeito-escola (percentual imputado às escolas pela variação nas notas dos seus estudantes) nas provas do ENEM e no certame em geral. Além disso, verificamos o poder explicativo após a inserção de variáveis de contexto (como posição social individual e média da escola e o tipo de dependência administrativa) no modelo.

Como primeiro resultado, encontramos que, no Modelo Completo (com variáveis explicativas), o efeito-escola cai para 8%, chegando até mesmo nos 4,6%

na prova de Linguagens e Códigos. Esse dado aponta para a baixa influência das escolas no desempenho dos seus estudantes e levanta questões sobre o uso e a interpretação dos *rankings* baseados no ENEM como indicadores de qualidade escolar, já que a maior parte da variação nos desempenhos é explicada por variações a nível individual, ou seja, entre os estudantes. Esse baixo percentual também foi encontrado por Travitzki (2013) ao analisar os dados do ENEM no ano de 2009. No entanto, ainda assim há escolas que contornam a tendência geral observada entre a origem social e desempenho. Essas escolas alcançam desempenhos acima do que seria esperado dado seu contexto. Denominamos essas instituições como *escolas que fazem a diferença*.

Como segundo resultado, verificamos que as variáveis explicativas utilizadas em nosso Modelo Completo (posição social individual e média escolar e o tipo de dependência administrativa da escola) foi capaz de explicar aproximadamente 88% das notas das escolas em Ciências da Natureza no período de 2015 a 2019, restando, portanto, 12% não explicado pelo modelo. Ainda que seja tentador imputar esses 12% a um suposto “mérito institucional”, devemos ter em mente que nosso modelo não, nem pretendeu ser, é capaz de dar conta de todos os fatores que influenciam de alguma forma o desempenho dos estudantes. Dessa forma, o mais correto é afirmar que nosso modelo, que utilizou apenas variáveis socioeconômicas e o tipo de dependência administrativa da escola, não foi capaz de explicar 12% da variação das notas no nível das escolas.

Os resultados apresentados até aqui nos ajudam a visualizar, em um primeiro momento, como as desigualdades sociais no Brasil ainda possuem uma forte influência no sistema escolar. Essa relação exerce um papel importante na manutenção de privilégios sociais via desigualdades escolares (BOURDIEU; PASSERON, 2010). Em um segundo momento, vemos que os *rankings* escolares não fogem dessa influência, classificando as melhores escolas aquelas onde estão filhos de famílias mais abastadas. Essas classificações podem influenciar na escolha da instituição escolar dos filhos, onde os pais, a depender de seu conhecimento do sistema educacional e de outras variáveis, como a própria localização da escola, julgam aquelas que atendem a seus objetivos e expectativas (NOGUEIRA; RESENDE; VIANA, 2015; RESENDE; NOGUEIRA; NOGUEIRA, 2011).

Uma vez que os *rankings* são interpretados como divulgadores da qualidade do ensino das escolas brasileiras, as pessoas tendem a acreditar que nas escolas

mais bem colocadas há, de fato, qualidade escolar. As escolas que não estão bem classificadas tendem, portanto, a ser mal vistas. Esses resultados em avaliações sistêmicas influenciam também no investimento que as escolas recebem, onde as escolas mais centrais e em regiões mais abastadas recebem mais investimentos (GARCIA DA SILVA; LIMA JUNIOR, 2020). Com isso, aquelas escolas que estão localizadas em regiões mais periféricas tenderão a receber menos investimentos, porque “se saíram mal” na avaliação.

Essa desigualdade presente na distribuição das famílias pelo espaço territorial, bem como nos investimentos públicos destinados às escolas, também contribui para que certos estados tenham vantagens sobre os demais. Com isso, as oportunidades e sua ciência, bem como as percepções e aspirações frente a essas oportunidades, variam de acordo com o contexto geográfico. Essa relação é compreendida pelo termo “geografia de oportunidades” (NOGUEIRA; RESENDE; VIANA, 2015) e pode ser, juntamente com outras variáveis, crucial na escolha de instituições escolares (PAULA; NOGUEIRA, 2018).

Sabendo dessa relação de predição entre origem social e desempenho escolar, decidimos mudar o enfoque que os *rankings* geralmente dão. Não optamos por olhar para as notas brutas (obtidas a partir das médias dos estudantes e habitantes), mas para o quanto a nota alcançada por determinada instituição escolar foi além do que seria esperado pelo modelo para aquela realidade (ou seja, sua posição social). Sendo assim, como terceiro resultado, vimos que escolas com posição social mais baixa (próximas de zero) despontaram, assumindo as primeiras colocações.

Essa mudança corrobora para a visualização do caráter arbitrário dos *rankings* e de como eles são sensíveis às intenções de quem opera a classificação (BOURDIEU, 2020). O mesmo ocorreu com relação aos estados. Quando o efeito da posição social sobre o desempenho foi controlado, vimos que alguns estados da região Nordeste despontaram e apresentaram maiores contribuições para o desempenho no ENEM do que alguns estados das regiões mais ricas do país (Sul, Sudeste e Distrito Federal).

## 6 CONCLUSÃO

As desigualdades escolares sofreram transformações ao longo dos anos. Se antes tínhamos uma exclusão da classe popular no acesso à educação, hoje essa exclusão ocorre na escola, adiando a eliminação dessa parcela da população (FREITAS, 2002, 2007). Essas desigualdades, que têm por base as desigualdades sociais (BOURDIEU; PASSERON, 2010), foram transladas para o acesso e conclusão do Ensino Superior (GONÇALVES; RAMOS, 2019; KNOP; COLLARES, 2019; LIMA JUNIOR; FRAGA JUNIOR, 2021).

Hoje, após uma reforma em sua estrutura, uma das formas de ingresso ao Ensino Superior de abrangência nacional é o Exame Nacional de Ensino Médio (ENEM) (INEP, 2020), cuja nota pode ser utilizada tanto para ingresso em universidades públicas e privadas no Brasil, quanto em universidades portuguesas. Sabendo do efeito que a origem social carrega sobre os desempenhos e sucesso escolares, propusemo-nos a seguinte pergunta: *controlando o efeito da origem social sobre o ENEM, quais escolas e Unidades da Federação fazem a diferença na educação em Ciências no Brasil?*

Para responder a essa pergunta, adotamos por referencial teórico a Sociologia de Pierre Bourdieu (BOURDIEU, 1980, 1998, 2007, 2020; BOURDIEU; PASSERON, 2010), com o objetivo de compreender seu quadro explicativo das relações entre origem social e escola, bem como entender o caráter arbitrário das avaliações escolares. Além disso, utilizamos pesquisas sobre eficácia escolar para entender seu surgimento, o que consideram como escolas eficazes e suas metodologias de análise (BROOKE; SOARES, 2008). Para os modelos estatísticos, utilizamos a Análise de Correspondência Múltipla (GREENACRE, 2007) para encontrarmos uma medida da posição social dos candidatos e, posteriormente, das escolas da amostra, e a Regressão Linear Multinível (HOX, 2010), para verificar o percentual de explicação das escolas sobre a nota de seus estudantes (efeito-escola), bem como o poder de explicação das variáveis contextuais (nível socioeconômico e dependência administrativa da escola) sobre o desempenho das escolas nas provas do ENEM e, em específico, de Ciências da Natureza.

Fizemos duas espécies de classificação. Na primeira, mais usual, ordenamos as notas brutas no certame de forma decrescente. Vimos que as primeiras posições são ocupadas sobretudo por escolas onde estão localizados os filhos de frações de

classe mais abastadas. Na segunda, proposta por nós, decidimos ordenar as escolas com base nos interceptos das escolas, ou seja, no quão distante, para mais, sua nota alcançada foi da nota prevista pelo modelo, dada sua posição social. A essas escolas denominamos *escolas que fazem a diferença*. Ao realizar essa mudança, houve uma mudança radical nas primeiras posições, onde escolas de posição social mais baixa despontaram e ocuparam o topo do *ranking*.

Vale lembrar que, em nosso modelo de Regressão Linear Multinível (HOX, 2010), a nota na prova de Ciências da Natureza foi explicada em aproximadamente 88% pelas variáveis contextuais inseridas quando controlamos o efeito da origem social. Já com relação ao efeito-escola, nessa mesma prova, apenas 5,6% da variação na nota dos estudantes pode ser atribuída à escola. No caso do ENEM como um todo, esses valores são de 88% e 8%, respectivamente.

Similarmente ao ocorrido com as escolas, os *rankings* estaduais também se mostraram sensíveis à origem social de seus habitantes. Quando operada sob efeito da posição social, a classificação dos estados privilegiou aqueles localizados nas regiões Sul e Sudeste do país, além do Distrito Federal. Quando controlamos o efeito da origem social sobre a nota em Ciências da Natureza e no ENEM geral, vimos que os estados da região Nordeste despontaram, apresentando grandes contribuições para a nota nessas provas.

Essas mudanças nos *rankings* decorrentes das alterações na forma de classificar explicitam o caráter arbitrário dos ranqueamentos exercidos na nossa sociedade. Ou seja, não há um princípio universal que sustente a forma como as classificações educacionais são operadas, elas são fruto de lutas e disputas travadas entre agentes e instituições no interior do campo educacional. Um outro ponto que desejamos levantar é sobre o caráter performativo dos *rankings* escolares. Tal como imperativos de excelência (LIMA JUNIOR et al., 2020a), eles são atos de instituição (BOURDIEU, 2020) e participam ativamente da criação e manutenção do que se entende por qualidade escolar e qualidade de ensino de Ciências. Portanto, a qualidade escolar é arbitrária, uma ficção, já que não está dada na natureza das escolas e redes de ensino avaliados.

Esperamos, com os resultados aqui apontados, contribuir para uma visão que critica o atrelamento da qualidade do ensino, em específico o de Ciências, ao desempenho em uma avaliação padronizada, como o ENEM. Também foi nossa intenção mostrar as raízes da compulsão pelos ranqueamentos que vemos nos dias

de hoje, a partir de um panorama das transformações que a educação brasileira e o ENEM passaram ao longo dos anos. Não buscamos aqui solucionar o problema da qualidade escolar, apenas mostrar, propondo uma forma alternativa de classificação, que a maneira vigente de ranquear tende a privilegiar e celebrar instituições escolares frequentadas por filhos de frações de classe mais abastadas.

## REFERÊNCIAS

ALMEIDA, D. C. de. et al. Enem: uma análise do efeito das políticas públicas educacionais em Alagoas e a comparação do resultado do ranking com os demais estados do Nordeste. **Revista Brasileira de Política e Administração da Educação**, v. 33, n. 2, p. 399–419, 2017.

ANDRIOLA, W. B. Doze motivos favoráveis à adoção do Exame Nacional do Ensino Médio (ENEM) pelas Instituições Federais de Ensino Superior (IFES). **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 19, n. 70, p. 107–125, 2011.

BARROSO FILHO, G. (2000). Universalização da Escola Pública do “Para que?” ao “Quanto?”. **Revista Contexto & Educação**, 59, 07-20.

BOURDIEU, P. The forms of capital. In: RICHARDSON, J. (org.). **Handbook of Theory and Research for the Sociology of Education**. Westport: Greenwood, 1986, p. 15-29.

BOURDIEU, P. **Escritos de educação**. Petrópolis: Ed. Vozes, 1998.

BOURDIEU, P. **A Distinção: crítica social do julgamento**. Porto Alegre: Zouk, 2007.

BOURDIEU, P. **Razões práticas: sobre a teoria da ação**. 9. ed. Campinas, SP: Papyrus, 2008.

BOURDIEU, P. **O senso prático**. Petropolis (RJ): Vozes, 2009.

BOURDIEU, P.; PASSERON, J.-C. **A reprodução: elementos para uma teoria do sistema de ensino**. 3. ed ed. Petropolis (RJ): Vozes, 2010.

BOURDIEU, P.; PASSERON, J. C. **Os herdeiros: os estudantes e a cultura**. 1. ed. Florianópolis: Editora UFSC, 2014.

BOURDIEU, P. **Sociologia Geral vol.1: Lutas de Classificação**. Petrópolis: Vozes, 2020.

BRESSOUX, P. As pesquisas sobre o efeito-escola e o efeito-professor. **Educação em Revista**, v. 38, p. 17–88, 1 dez. 2003.

BROOKE, N.; SOARES, J. F. **Pesquisa em eficácia escolar: origem e trajetórias**. 1a edição ed. Belo Horizonte: Editora UFMG, 2008.

CASTRO, M.H.G; TIEZZI, S. A reforma do ensino médio e a implantação do Enem no Brasil. In: BROCK, C.; SCHWARTZMAN, S. (Org.). **Os desafios da educação no Brasil**. Rio de Janeiro, 2002.

COLEMAN, J. S. EQUALITY OF EDUCATIONAL OPPORTUNITY. **Equity & Excellence in Education**, v. 6, n. 5, p. 19–28, set. 1968.

CORDEIRO, R. S. et al. BOT NICA, CAI TANTO 'ENEM' SABIA! UMA ANÁLISE DO PERFIL DOS ITENS NO ENEM. **Revista Brasileira de Ensino de Ciências e Matemática**, v. 4, n. 1, 2020.

DARDOT, P.; LAVAL, C. **A nova razão do mundo**: ensaio sobre a sociedade neoliberal. São Paulo: Boitempo, 2016.

DUTRA, R. S. et al. Determinantes do desempenho educacional dos Institutos Federais do Brasil no Exame Nacional do Ensino Médio. **Educação e Pesquisa: Revista da Faculdade de Educação da Universidade de São Paulo**, v. 45, 2019.

EDMONDS, R. R. Effective Schools for the Urban Poor. **Educational Leadership**, v. 37, p. 15-24, 1979.

FERNANDES, L. D. S.; CAMPOS, A. F. Análise das questões sobre radioatividade no Exame Nacional do Ensino Médio - ENEM. **Amazônia (Universidade Federal do Pará)**, v. 13, n. 25, p. 62–74, 2016.

FIGUEIRÊDO, E.; NOGUEIRA, L.; SANTANA, F. L. Igualdade de Oportunidades: Analisando o Papel das Circunstâncias no Desempenho do ENEM. **Revista Brasileira de Economia**, v. 68, n. 3, p. 373–392, 2014.

FREITAS, L. C. DE. A internalização da exclusão. **Educação & Sociedade**, v. 23, n. 80, p. 299–325, set. 2002.

FREITAS, L. C. DE. Eliminação adiada: o ocaso das classes populares no interior da escola e a ocultação da (má) qualidade do ensino. **Educação & Sociedade**, v. 28, n. 100, p. 965–987, out. 2007.

FREITAS, L. C. DE. **A reforma empresarial da educação: nova direita, velhas ideias**. 1a edição ed. São Paulo: Expressão Popular, 2018.

GADOTTI, M. **História das idéias pedagógicas**. São Paulo, SP: Editora Atica, 1993.

GARCIA DA SILVA, S.; LIMA JUNIOR, P. A Educação Científica das Periferias Urbanas: Uma Revisão sobre o Ensino de Ciências em Contextos de Vulnerabilidade Social (1985–2018). **Revista Brasileira de Pesquisa em Educação em Ciências**, p. 221–243, 2020.

GONÇALVES, F. G. DE; RAMOS, M. P. SUCESSO NO CAMPO ESCOLAR: CONDICIONANTES PARA ENTRADA NA UNIVERSIDADE NO BRASIL. **Educação & Sociedade**, v. 40, p. e0188393, 2019.

GOUVÊA, M. C. S. DE. A escolarização da criança brasileira no século XIX: apontamentos para uma re-escrita. **Revista Educação em Questão**, v. 28, n. 14, 15 jun. 2007.

GRAY, J.; WILCOX, B. **Good School, Bad School: Evaluating Performance and Encouraging Improvement**. Buckingham: Open University Press, 1995.

GREENACRE, M. J. **Correspondence analysis in practice**. 2nd ed ed. Boca Raton: Chapman & Hall/CRC, 2007.

HOX, J. J. **Multilevel analysis: techniques and applications**. 2. ed. New York: Routledge, Taylor & Francis, 2010.

INEP [INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA]. **Exame Nacional do Ensino Médio (Enem)**. Brasília, DF: INEP, 2020a. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>. Acesso em: 2 de novembro de 2022.

INEP [INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA]. **Pisa 2018 revela baixo desempenho escolar em Leitura, Matemática e Ciências no Brasil**. Brasília, DF: INEP, 2020b. Disponível em: [https://www.gov.br/mec/pt-br/assuntos/noticias\\_1/pisa-2018-revela-baixo-desempenho-escolar-em-leitura-matematica-e-ciencias-no-brasil](https://www.gov.br/mec/pt-br/assuntos/noticias_1/pisa-2018-revela-baixo-desempenho-escolar-em-leitura-matematica-e-ciencias-no-brasil). Acesso em: 15 de maio de 2023.

INEP [INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA]. **Microdados**. Brasília, DF: INEP, 2022. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados>. Acesso em: 2 de novembro de 2022.

JUSTINIANO, E. F.; QUEIROZ, A. P. DE. Desempenho dos alunos de escolas públicas no Exame Nacional do Ensino Médio (2012-2018): análise e representação. **Revista do Departamento de Geografia**., v. 41, p. e185791, 2021.

KLEINKE, M. U. Influência do status socioeconômico no desempenho dos estudantes nos itens de física do Enem 2012. **Revista Brasileira de Ensino de Física**, v. 39, n. 2, 2017.

KNOP, M.; COLLARES, A. C. M. A influência da origem social na probabilidade de concluir os diferentes cursos de ensino superior. **Sociedade e Estado**, v. 34, n. 2, p. 351–380, maio 2019.

KUJAWA, D. R.; MARTINS, A. R. DE Q.; PATIAS, N. D. A EVOLUÇÃO HISTÓRICA DA EDUCAÇÃO E DA ESCOLA NO BRASIL. **Revista Sociais e Humanas**, v. 33, n. 3, 9 dez. 2020.

LAUTÉRIO, A. Q. M. R.; NEHRING, C. M. REESTRUTURAÇÃO DO CURRÍCULO ESCOLAR: A TRAJETÓRIA DO ENSINO MÉDIO E O CONCEITO DE CONTEXTUALIZAÇÃO. *In: IX ANPED SUL, 2012, Caxias do Sul. Anais [...]* Caxias do Sul: ANPED, 2012, p. 1-16.

LAVAL, C. **A escola não é uma empresa**: o neo-liberalismo em ataque ao ensino público. Londrina: Editora Planta, 2004.

LEE, V. Using Hierarchical Linear Modeling to Study Social Contexts: the Case of School Effects. **Educational Psychologist**, v. 35(2), p. 125-141, 2000.

LIMA JUNIOR, P. **Métodos quantitativos da pesquisa educacional**: uma introdução baseada em R. Brasília: Editora da UnB, 2023.

LIMA JUNIOR, P.; OSTERMANN, F.; REZENDE, F. **Razões para desistir**: análise sociológica da evasão do curso de Física. Curitiba: Appris, 2018.

LIMA JUNIOR, P.; OSTERMANN, F.; REZENDE, F. Análise dos condicionantes sociais do sucesso acadêmico em cursos de graduação em física à luz da sociologia de Bourdieu. **Ensaio Pesquisa em Educação em Ciências**, v. 15, n. 1, p. 113–129, 2013.

LIMA JUNIOR, P. et al. A Física como uma construção cultural arbitrária: Um exemplo da controvérsia sobre o status ontológico das forças inerciais. **Revista Brasileira de Pesquisa em Educação em Ciências**, v. 15, n. 1, p. 195–217, 2014.

LIMA JUNIOR, P. et al. EXCELÊNCIA, EVASÃO E EXPERIÊNCIAS DE INTEGRAÇÃO DOS ESTUDANTES DE GRADUAÇÃO EM FÍSICA. **Ensaio Pesquisa em Educação em Ciências (Belo Horizonte)**, v. 22, p. e12165, 2020a.

LIMA JUNIOR, P. et al. A Integração dos Estudantes de Periferia no Curso de Física: razões institucionais da evasão segundo a origem social. **Ciência & Educação (Bauru)**, v. 26, p. e20030, 2020b.

LIMA JUNIOR, P.; FRAGA JUNIOR, J. C. QUAL É O EFEITO DA DESIGUALDADE SOCIAL NO DESEMPENHO EM CIÊNCIAS DOS ESTUDANTES BRASILEIROS? UMA ANÁLISE DO EXAME NACIONAL DO ENSINO MÉDIO (2012-2019). **Investigações em Ensino de Ciências**, v. 26, n. 1, p. 110–126, 2021.

LOPES, A. C.; LÓPEZ, S. B. A performatividade nas políticas de currículo: o caso do ENEM. **Educação em revista**, v. 26, n. 1, p. 89–110, 2010.

LUCENA, J. P. O.; DOS SANTOS, H. N. L. A relação entre desempenho no Exame Nacional do Ensino Médio e o perfil socioeconômico: um estudo com os microdados de 2016. **GeSec : Revista de Gestão e Secretariado**, v. 11, n. 2, p. 1–23, 2020.

MADAUS, G.; AIRASIAN, P.; KELLAGHAN, T. **School Effectiveness: a Reassessment of the Evidence**. New York: McGraw-Hill Book Company, 1980.

MAIR, P. **Modern Psychometrics with R**. Cham: Springer International Publishing, 2018.

MARCOM, G. S.; KLEINKE, M. U. Indicadores Formativos para o Ensino de Física através do Exame Nacional do Ensino Médio (ENEM). **Caderno Brasileiro de Ensino de Física**, v. 38, n. 3, p. 1388–1419, 2021.

MALUSÁ, S.; ORDONES, L. L. DE M.; RIBEIRO, E. ENEM: PONTOS POSITIVOS PARA A EDUCAÇÃO BRASILEIRA. **Revista Educação e Políticas em Debate**, v. 3, n. 2, 2015.

MASSI, L.; MUZZETI, L. R.; SUFICIER, D. M. A pesquisa sobre trajetórias escolares no Brasil. **Revista Ibero-Americana de Estudos em Educação**, v. 12, n. 3, p. 1854–1873, 5 set. 2017.

MELO, R. O. et al. Impacto das variáveis socioeconômicas no desempenho do Enem: uma análise espacial e sociológica. **Revista de administração pública (Rio de Janeiro)**, v. 55, n. 6, p. 1271–1294, 2021.

MORAES, C. P. de et al. Equidade e desempenho no Exame Nacional do Ensino Médio: Um estudo sobre sexo e raça nos municípios brasileiros. **Education policy analysis archives**, v. 30, 2022.

MORTIMORE, P.; SAMMONS, P.; STOLL, L.; LEWIS, D.; ECOB, R. **School Matters**. California: University of California Press, 1988.

MOSTELLER, F.; MOYNIHAN, D. **On Equality of Educational Opportunity**. New York: Vintage Books, 1972.

NASCIMENTO, M. M.; CAVALCANTI, C.; OSTERMANN, F. Uma busca por questões de Física do ENEM potencialmente não reprodutoras das desigualdades socioeconômicas. **Revista Brasileira de Ensino de Física**, v. 40, n. 3, 2018.

NAVARRO, D. et al. Impactos do Conhecimento das Ciências Naturais para o Desempenho no ENEM: Considerações sobre a Desigualdade Científico-tecnológica para a Justiça Social. **Revista Brasileira de Pesquisa em Educação em Ciências**, v. 21, n. u, p. 1–30, 2021.

NOGUEIRA, C. M. M.; RESENDE, T. D. F.; VIANA, M. J. B. Escolha do estabelecimento de ensino, mobilização familiar e desempenho escolar. **Revista Brasileira de Educação**, v. 20, n. 62, p. 749–772, set. 2015.

NOGUEIRA, M. A. O CAPITAL CULTURAL E A PRODUÇÃO DAS DESIGUALDADES ESCOLARES CONTEMPORÂNEAS. **Cadernos de Pesquisa**, v. 51, p. e07468, 2021.

NOGUEIRA, M. A. A construção da excelência escolar – Um estudo de trajetórias feito com estudantes universitários provenientes das camadas médias intelectualizadas. In: NOGUEIRA, M. A.; ROMANELLI, G.; ZAGO, N. (Org.). **Família e Escola: trajetórias de escolarização em camadas médias e populares**. Editora Vozes: Petrópolis, 2000, p. 125-154.

NOGUEIRA, M. A.; LACERDA, W. G. Os rankings de estabelecimentos de Ensino Médio e as lógicas de ação das escolas: o caso do Colégio de Aplicação da UFV. In: KRAWCZYK, N. (Org.). **Sociologia do Ensino Médio: crítica ao economicismo na política educacional**. Cortez Editora: São Paulo, 2014, p. 127-161.

NOGUEIRA, M. A. Estratégias de escolarização em famílias de empresários. In: ALMEIDA, A. M. F.; NOGUEIRA, M. A. (Org.). **A escolarização das elites: um panorama internacional da pesquisa**. Editora Vozes: Petrópolis, 2002, p. 49-65.

NOGUEIRA, M. A.; ROMANELLI, G.; ZAGO, N. (EDS.). **Família e escola: trajetórias de escolarização em camadas médias e populares**. Petrópolis: Editora Vozes, 2000.

PAULA, G. B. de; NOGUEIRA, M. A. de L. G. Desigualdades socioespaciais e escolhas escolares. **Educação**, p. 55–74, 29 mar. 2018.

RAFTERY, A. E.; HOUT, M. Maximally Maintained Inequality: Expansion, Reform, and Opportunity in Irish Education, 1921-75. **Sociology of Education**, v. 66, n. 1, p. 41, jan. 1993.

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

RUTTER, M.; MAUGHAN, B.; MORTIMORE, P.; OUSTON, J.; SMITH, A. **Fifteen Thousand Hours: Secondary Schools and their Effect on Children**. London: Open Books, 1979.

SCHNORR, S. M.; PIETROCOLA, M. Educação em Ciências e Matemática no Brasil: uma Revisão Sistemática de 25 Anos de Pesquisa (1994–2018). **Revista Brasileira de Pesquisa em Educação em Ciências**, p. e37242-30, 20 jul. 2022.

SILGE, J.; ROBINSON, D. **Text mining with R: a tidy approach**. 1ª ed. Boston: O'Reilly, 2017.

SOARES, J. F.; ALVES, M. T. G.; OLIVEIRA, R. M. DE. O efeito de 248 escolas de nível médio no vestibular da UFMG nos anos de 1998 1999 e 2000. **Estudos em Avaliação Educacional**, n. 24, p. 69–118, 30 dez. 2001.

TORRES, R. et al. Determinantes do Desempenho dos Participantes da Prova do Enem: Evidências para o Rio Grande Do Sul. **Desenvolvimento em questão**, v. 18, n. 53, p. 352–368, 2020.

TRAVITZKI, R.; FERRÃO, M. E.; COUTO, A. P. Desigualdades educacionais e socioeconômicas na população brasileira pré-universitária: Uma visão a partir da análise de dados do ENEM. **Education policy analysis archives**, v. 24, p. 74, 2016.

TRAVITZKI, Rodrigo. **ENEM**: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar. 2013. Tese (Doutorado em Educação) - Faculdade de Educação, Universidade de São Paulo, São Paulo, 2013. doi:10.11606/T.48.2013.tde-28062013-162014. Acesso em: 2023-07-21.

UNITED NATIONS DEVELOPMENT PROGRAMME; FUNDAÇÃO JOÃO PINHEIRO; INSTITUTO DE PESQUISA ECONÔMICA APLICADA (EDS.). **O Índice de Desenvolvimento Humano Municipal Brasileiro**. Brasília, Distrito Federal, Brazil: PNUD, 2013.

VIGGIANO, E.; MATTOS, C. O desempenho de estudantes no Enem 2010 em diferentes regiões brasileiras. **Revista Brasileira de Estudos Pedagógicos**, v. 94, n. 237, p. 417–438, 2013.

VIZZOTTO, P. Análise e classificação das questões de Biologia do ENEM segundo suas características psicométricas. **Revista da SBEnBIO**, v. 15, n. 1, 2022a.

VIZZOTTO, P. A. Qual é a qualidade psicométrica das questões de Química do ENEM? Uma análise dos itens de 2009 a 2019. **Revista de Ensino de Ciências e Matemática**, v. 13, n. 1, p. 1–21, 2022b.

VIZZOTTO, P. A. As “piores” questões de Física do ENEM: Uma análise psicométrica das edições de 2009 a 2019. **Revista Brasileira de Ensino de Física**, v. 44, 2022c.

VIZZOTTO, P. A.; MACKEDANZ, L. F. Como a Física aplicada ao trânsito foi abordada no ENEM de 1998 até 2018? **Revista Brasileira de Ensino de Ciências e Matemática**, v. 3, n. 1, 2020.

WILLMS, J. **Monitoring School Performance: a Guide for Educators**. London: The Falmer Press, 1992.

ZACCHI, R. C.; NEY, M. G.; PONCIANO, N. J. Desigualdades educacionais na educação básica: uma investigação a partir do Exame Nacional do Ensino Médio. **Vértices (Campos dos Goitacazes)**, v. 18, n. 1, p. 79–108, 2016.

ZAGO, N. Processos de escolarização nos meios populares – As contradições da obrigatoriedade escolar. In: NOGUEIRA, M. A.; ROMANELLI, G.; ZAGO, N. (Org.). **Família e Escola: trajetórias de escolarização em camadas médias e populares**. Editora Vozes: Petrópolis, 2000, p. 17-44.

**DECLARAÇÃO DE ORIGINALIDADE DE DISSERTAÇÃO DE Mestrado OU TESE DE  
DOUTORADO**

Declaro que a presente dissertação/tese é original, elaborada especialmente para este fim, não tendo sido apresentada para obtenção de qualquer título e que identifico e cito devidamente todas as autoras e todos os autores que contribuíram para o trabalho, bem como as contribuições oriundas de outras publicações de minha autoria.

Declaro estar ciente de que a cópia ou o plágio podem gerar responsabilidade civil, criminal e disciplinar, consistindo em grave violação à ética acadêmica.

Brasília, 24 de agosto de 2023.

Assinatura do/a discente: \_\_\_\_\_

Programa: Pós-Graduação em Educação em Ciências

Nome completo: Jailton Correia Fraga Junior

Título do Trabalho: Desigualdades sociais e excelência na educação científica:  
uma busca por escolas que fazem a diferença a partir do Exame Nacional do  
Ensino Médio

Nível: ( X ) Mestrado    ( ) Doutorado

Orientador/a: Paulo Roberto Menezes Lima Junior