

Fatores socioeconômicos associados ao desempenho no Enem

Alexandre Jaloto

Ricardo Primi

125

Resumo

Diferenças sociais têm sido associadas a diferenças de desempenho em testes educacionais de larga escala. Estudos com os dados do Exame Nacional do Ensino Médio (Enem) sobre a temática são raros e, por vezes, não incluem algumas variáveis (como o atraso escolar) e não abordam os testes das quatro áreas do conhecimento. Com o objetivo de examinar as variáveis sociais que explicam o desempenho nos quatro testes da edição de 2018, analisamos atraso escolar, sexo, cor/raça, nível socioeconômico (NSE) e a dependência administrativa da escola do aluno. Além de avaliar a associação entre as variáveis e o desempenho, verificamos se o efeito do NSE variava entre escolas. Realizamos uma modelagem por meio de regressão linear multinível. Todas as variáveis se mostraram estatisticamente significativas, porém sua relevância variou. O desempenho dos alunos brancos superou o dos demais em menos de 10,0 pontos; ser de escola privada esteve associado a um aumento de até 83,9 em relação a ser de escola estadual. O efeito do NSE sobre o desempenho variou entre escolas. Há uma persistente desigualdade educacional no Enem.

Palavras-chave: avaliação da educação básica; equidade em educação; política da educação.

Abstract

Socio-economic factors associated with student performance on Enem

Social differences have been associated with differences in performance on large-scale educational tests. Studies based on data from the National High School Exam (Enem) on this theme are rare and, at times, do not include some variables (such as school delay) and do not approach tests of the four areas of knowledge. Aiming to examine the social variables that explain the performance in the four tests of the 2018 Enem edition, school delay, gender, color/race, socioeconomic status (SES) and administrative dependence of the student's school were analyzed. In addition to evaluating the association between the variables and performance, this study verifies whether the effects of SES varied across schools. This study also uses a modeling using multilevel linear regression. All variables were statistically significant, but their relevance varied. White students outperformed other students by less than 10.0 points, and studying at a private school was associated with an increase of up to 83.9 as compared to studying at a state school. SES's effect on performance varied across schools. Our findings point to a persistent educational inequality in Enem.

keywords: assessment of basic education; education policy; equity in education.

Resumen

Factores socioeconómicos asociados con el desempeño en Enem

Las diferencias sociales se han asociado con diferencias de rendimiento en pruebas educativas a gran escala. Estudios con datos del Examen Nacional de Enseñanza Media (Enem) que tratan el tema son raros y en ocasiones no incluyen algunas variables (como retraso escolar) y no abordan las pruebas en las cuatro áreas del conocimiento. Este trabajo tuvo como objetivo examinar las variables sociales que explican el desempeño en las cuatro pruebas de la edición 2018 de Enem. Se analizó el retraso escolar, género, color/raza, nivel socioeconómico (NSE) y dependencia administrativa de la escuela del estudiante. Además de evaluar la asociación entre las variables y el desempeño, verificamos si el efecto de NSE variaba entre escuelas. Se realizó un modelado mediante regresión lineal multinivel. Todas las variables fueron estadísticamente significativas, pero su relevancia varió. El desempeño de los estudiantes blancos superó al de los demás por menos de 10.0 puntos, mientras que ser de una escuela privada se asoció con un aumento de hasta 83.9 en comparación con ser de una escuela estatal. Se identificó que el efecto de NSE sobre el rendimiento varió entre las escuelas. Los hallazgos apuntan a una desigualdad educativa persistente en Enem.

Palabras clave: evaluación de la educación básica; equidad en la educación; política educativa.

Introdução

Algumas políticas públicas têm sido desenvolvidas com o objetivo de superar desigualdades educacionais, como a Lei de Cotas (Brasil, 2012). Essa lei garante a reserva de 50% das vagas de cursos das instituições federais de ensino superior (Ifes) para estudantes de escolas públicas. No mínimo 50% dessas vagas são destinadas a estudantes de famílias com renda *per capita* de até 1,5 salário-mínimo. Além disso, as vagas reservadas são destinadas a pessoas pretas, pardas ou indígenas (PPI) pelo menos na mesma proporção em que esses grupos são observados na unidade da Federação em que a instituição se localiza. A partir de 2012, verifica-se um aumento de estudantes egressos de escolas públicas nas Ifes, em especial pessoas PPI (Senkevics; Mello, 2019).

Além da reserva de vagas para determinados grupos, é possível adotar procedimentos matematicamente sofisticados, como a otimização restrita. Esse procedimento, que envolve uma série de cálculos matemáticos, permite maximizar um critério (como a nota em um teste) ao mesmo tempo que impõe algumas restrições na composição das turmas. No caso de seleção para ingresso em um curso de ensino superior com 100 vagas, o programa estatístico busca montar o grupo de participantes que apresente maior média de desempenho e siga as proporções desejadas de características sociodemográficas, como a proporção de grupos étnicos. Primeiro, o programa monta diversos grupos de 100 sujeitos (número de vagas), de modo que cada grupo tenha as proporções desejadas (restrição). Em seguida, ele calcula a média de cada grupo e seleciona aquele com maior valor. Zwick, Ye e Isham (2019) aplicaram o procedimento de maneira ilustrativa e verificaram que ele pode aumentar a diversidade socioeconômica e étnica das turmas de ingressantes de uma universidade dos Estados Unidos, sem afetar o desempenho no primeiro ano de graduação.

As variáveis utilizadas para produção de políticas afirmativas (nível socioeconômico – NSE, cor/raça e dependência administrativa da escola) e outras têm sido associadas ao desempenho em testes educacionais de larga escala. Por exemplo, em 2001, o atraso escolar e não ser branco tiveram uma associação negativa com o desempenho nos testes do Sistema de Avaliação da Educação Básica (Saeb). Ao contrário, o NSE do aluno e ser do sexo masculino foram positivamente associados (Andrade; Laros, 2007).

Na edição de 2012 do Programa Internacional de Avaliação de Estudantes (Pisa, em inglês *Programme for International Student Assessment*), o nível de escolaridade da mãe e ser do sexo feminino estiveram positivamente associados ao desempenho nos testes de Leitura e Matemática. Por outro lado, a repetência e estudar em escola pública estiveram negativamente associados ao desempenho nesses testes e no de Ciências. Adicionalmente, ser do sexo feminino se relacionou negativamente com o desempenho em Ciências (Aristizabal; Caicedo; Parra, 2017).

No Exame Nacional do Ensino Médio (Enem) de 2012, verificou-se que o NSE e ser branco ou amarelo estavam positivamente associados à nota em Matemática e Linguagens e Códigos (Travitzki; Ferrão; Couto, 2016). No entanto, outras variáveis

comumente utilizadas nesse tipo de análise não foram incluídas, como atraso escolar, sexo e dependência administrativa da escola.

A forma como o NSE é empregado nas pesquisas varia, dada sua natureza abstrata (Broer; Bai; Fonseca, 2019). Há alguns estudos que utilizam uma medida composta por alguns indicadores (Alves; Soares; Xavier, 2016; Laros; Marciano; Andrade, 2012; Souza Vieira, 2020) e outros utilizam uma variável como aproximação do NSE, como a escolaridade dos pais (Aristizabal; Caicedo; Parra, 2017; Bonamino *et al.*, 2010; Travitzki; Ferrão; Couto, 2016). No geral, o NSE tem incluído como principais componentes a escolaridade dos pais, sua ocupação e a renda familiar (Cowan *et al.*, 2012; Sirin, 2005). Adotamos a recomendação de Cowan *et al.* (2012) de utilizar uma medida composta.

Como descrito, algumas variáveis sociais se mostram estatisticamente importantes na predição do desempenho, como NSE, escolaridade dos pais, atraso escolar, sexo, cor/raça e dependência administrativa (Aristizabal; Caicedo; Parra, 2017; Karino; Laros, 2017). No entanto, nos estudos citados que têm o Enem como objeto, algumas delas não são incluídas (atraso escolar e dependência administrativa). Ainda, somente dois dos quatro testes são utilizados (Linguagens e Códigos; e Matemática). Assumindo a potencialidade dos estudos sobre os fatores associados ao desempenho para a produção de políticas públicas, este trabalho objetiva examinar as variáveis sociais que explicam o desempenho nos quatro testes do Enem, a saber: Ciências Humanas (CH); Ciências da Natureza (CN); Matemática (MT); e Linguagens e Códigos (LC). Para tal, responderemos a três perguntas:

- 1) Diferenças entre escolas explicam quanto do desempenho?
- 2) Variáveis no nível do aluno e da escola explicam quanto do desempenho?
- 3) O efeito do NSE sobre o desempenho varia entre escolas?

Método

Participantes

Os dados deste estudo foram extraídos dos microdados do Enem 2018 (Brasil. Inep, 2018) e obtidos em janeiro de 2020. Seleccionamos os participantes com pelo menos 16 anos em 31 de dezembro de 2018 concluintes do ensino médio (ensino regular e educação de jovens e adultos).

Incluímos nas análises de cada prova todos os participantes que tiveram nota nela, ainda que tenham faltado à aplicação de outra. Excluímos os que deixaram os 45 itens em branco e os que não tinham informações sobre sexo ou cor/raça. Por fim, excluímos os estudantes de escolas com menos de 30 participantes presentes na prova. Tratamos esses casos como se os alunos fossem de escolas diferentes, uma vez que uma mesma escola pode oferecer diferentes modalidades de ensino (ensino regular e educação de jovens e adultos). A Tabela 1 traz as características sociodemográficas dos participantes.

Tabela 1 – Características sociodemográficas dos sujeitos incluídos

	Ciências Humanas		Ciências da Natureza		Linguagens e Códigos		Matemática	
	M	DP	M	DP	M	DP	M	DP
Variáveis intervalares								
Desempenho	565,4	81,7	495,4	75,9	525,6	73,7	540,9	105,0
NSE	0,2	0,9	0,2	0,9	0,2	0,9	0,2	0,9
Variáveis categóricas	N	%	N	%	N	%	N	%
Atraso escolar								
Aluno sem atraso escolar	756252	85,7	716331	86,3	756252	85,7	716331	86,3
Aluno com atraso escolar	126162	14,3	113569	13,7	126162	14,3	113569	13,7
Sexo								
Feminino	507266	57,5	475947	57,3	507266	57,5	475947	57,3
Masculino	375148	42,5	353953	42,7	375148	42,5	353953	42,7
Cor/raça								
Branca	339085	38,4	319742	38,5	339085	38,4	319742	38,5
Preta	99483	11,3	93050	11,2	99483	11,3	93050	11,2
Parda	418195	47,4	393070	47,4	418195	47,4	393070	47,4
Amarela	20469	2,3	19257	2,3	20469	2,3	19257	2,3
Indígena	5182	0,6	4781	0,6	5182	0,6	4781	0,6
Dependência administrativa da escola								
Municipal	6040	0,7	5520	0,7	6040	0,7	5520	0,7
Estadual	683232	77,4	635192	76,5	683232	77,4	635192	76,5
Privada	153169	17,4	149908	18,1	153169	17,4	149908	18,1
Federal	6040	0,7	5520	0,7	6040	0,7	5520	0,7

Fonte: Elaboração própria.

Nota: M = média; DP = desvio padrão.

Análise de dados

Realizamos uma regressão linear multinível para prever o desempenho em cada prova. As variáveis preditoras de primeiro nível (estudante) foram: atraso escolar (considerado neste trabalho como equivalente a possuir mais de 18 anos no fim de 2018), cor/raça, sexo e NSE. As variáveis atraso escolar, cor/raça e sexo foram usadas como dicotômicas, sendo as referências respectivamente: participantes com até 18 anos no fim de 2018, cor/raça não branca e sexo feminino. A resposta “não declarado” foi transformada em dado faltante (NA). O NSE foi calculado por meio do modelo de resposta graduada da Teoria de Resposta ao Item – TRI (Samejima, 1969). Para o cálculo dessa medida (construída a partir de cinco itens), utilizamos a escolaridade dos pais, sua ocupação e a renda familiar. A resposta “não sei” foi transformada em NA. As variáveis de ocupação dos pais possuem categorias que representam grupos de ocupação, as quais consideramos ordinais. Adotamos o mesmo procedimento para a escolaridade dos pais. Consideramos a variável NSE no nível do estudante, porque para incluí-la no nível da escola devemos observar a existência de outras variáveis da escola (como clima escolar) e o fato de alguns estudantes frequentarem escolas longe de sua vizinhança.

Ainda que o NSE do estudante e o NSE da escola tenham alta correlação, existem informações únicas em cada uma dessas variáveis e, neste estudo, focamos o NSE do estudante (Cowan *et al.*, 2012). Por último, a variável preditora de segundo nível (escola) foi a dependência administrativa, tratada como *dummy*. A referência foram as escolas municipais. No cálculo do NSE, as marcações NA foram consideradas como itens não apresentados ao sujeito. Na modelagem multinível, sujeitos com pelo menos um NA nas variáveis foram excluídos da análise. A Tabela 2 apresenta as variáveis utilizadas e suas correlações com os desempenhos.

Tabela 2 – Variáveis incluídas no estudo e suas correlações com os desempenhos

Nível	Variável	Descrição	Correlação com o desempenho			
			CH	CN	LC	MT
Estudante	Atraso escolar	1 = Sim 0 = Não	-0,18	-0,16	-0,21	-0,17
	Sexo	1 = Masculino 0 = Feminino	0,05	0,08	0,03	0,20
	Cor/raça branca	1 = Branca 0 = Outras	0,21	0,23	0,23	0,23
	NSE	Calculado a partir da renda familiar, escolaridade e ocupação dos pais	0,41	0,46	0,46	0,47
Escola	Estadual	1 = Rede estadual 0 = Outras (referência é municipal)	-0,40	-0,48	-0,41	-0,46
	Privada	1 = Rede privada 0 = Outras (referência é municipal)	0,36	0,44	0,37	0,42
	Federal	1 = Rede federal 0 = Outras (referência é municipal)	0,14	0,15	0,15	0,15

Fonte: Elaboração própria.

Nota: CH = Ciências Humanas; CN = Ciências da Natureza; LC = Linguagens e Códigos; MT = Matemática.

Construímos quatro modelos para cada prova e adotamos o mesmo protocolo de análise. No Modelo 1 (modelo nulo), em que somente o desempenho foi incluído, observamos a correlação intraclasse (ICC). Essa medida indica o quão homogêneas são as escolas e, por isso, diferentes umas das outras. Quanto maior esse parâmetro, mais a variação no desempenho é devida às diferenças entre escolas e mais se justifica uma análise multinível. Nos modelos 2 e 3, incluímos as variáveis do aluno e da escola, respectivamente. Os coeficientes das variáveis tiveram suas significâncias estatísticas avaliadas. No Modelo 4, permitimos que o efeito do NSE sobre o desempenho variasse entre escolas. As equações de cada modelo são as que seguem:

Modelo 1:

$$Y_{ij} = (\beta_0 + u_{0j}) + e_{ij}$$

Modelo 2:

$$Y_{ij} = (\beta_0 + u_{0j}) + \beta_1 \text{atraso} + \beta_2 \text{sexo} + \beta_3 \text{cor} + \beta_4 \text{nse} + e_{ij}$$

Modelo 3:

$$Y_{ij} = (\beta_0 + u_{0j}) + \beta_1 \text{atraso} + \beta_2 \text{sexo} + \beta_3 \text{cor} + \beta_4 \text{nse} + \beta_5 \text{estadual} + \beta_6 \text{privada} + \beta_7 \text{federal} + e_{ij}$$

Modelo 4:

$$Y_{ij} = (\beta_0 + u_{0j}) + \beta_1 \text{atraso} + \beta_2 \text{sexo} + \beta_3 \text{cor} + (\beta_4 + u_{4j}) \text{nse} + \beta_5 \text{estadual} + \beta_6 \text{privada} + \beta_7 \text{federal} + e_{ij}$$

Avaliamos se um modelo era estatisticamente diferente do anterior por meio do teste da razão de verossimilhança, que utiliza as informações da *deviance* e os graus de liberdade. Caso a diferença fosse verificada ($p < 0,05$), adotamos aquele com melhor ajuste. Consideramos um ajuste melhor se o valor do Critério de Informação de Akaike (AIC) reduzisse e a variância explicada em cada nível aumentasse. O AIC mede a qualidade do ajuste de um modelo. Ao se comparar dois modelos diferentes, aquele que apresentar menor valor possui melhor ajuste. A variância explicada corresponde à porcentagem de redução da variância do modelo nulo, ou seja, o quanto o modelo em voga explica a variância do desempenho em cada nível (Valentini *et al.*, 2015). As análises foram realizadas no R 4.0.3 (R Core Team, 2019) por meio dos pacotes *mirt* (v1.33.2; Chalmers, 2012) e *lme4* (v1.1-23; Bates *et al.*, 2015). R é um ambiente de programação gratuito para computação estatística e gráficos. Seus pacotes contêm programações adequadas para realizar diferentes análises, como as adotadas neste trabalho. Especificamente, o pacote *mirt* permite realizar análises no escopo da TRI e o pacote *lme4* permite realizar modelagens com regressão multinível.¹

¹ Os comandos utilizados nas análises estão disponíveis no repositório <https://github.com/alexandrejaloto/EmAbertoEnem>.

Resultados

Para responder nossas perguntas de pesquisa, construímos quatro modelos de regressão linear multinível para cada prova. As Tabelas de 3 a 6 mostram os parâmetros dos modelos das provas CH, CN, LC e MT, respectivamente. Essas tabelas apresentam os efeitos fixos, seus erros e seus parâmetros padronizados. Além desses, apresentamos os efeitos randômicos nos dois níveis, incluindo a variância do efeito do NSE no Modelo 4. As tabelas também incluem as estatísticas de cada modelo.

Para responder a primeira pergunta de pesquisa, investigar o quanto as diferenças entre as escolas explicam o desempenho, analisamos a ICC do Modelo 1. Em CH, esse valor foi de 0,28. Por isso, 28% da variância do desempenho nessa prova podem ser atribuídos a diferenças entre escolas; em CN, 37%; em LC, 32%; e em MT, 35%. Esses valores justificam uma análise multinível.

Para responder a segunda pergunta de pesquisa, investigar o quanto as variáveis no nível do aluno e no nível da escola explicam o desempenho, construímos os Modelos 2 e 3. No Modelo 2, os coeficientes das variáveis no nível do aluno se mostraram estatisticamente significativos para as quatro provas. O atraso escolar se associou a uma redução no desempenho nas quatro provas. Ao contrário, ser do sexo masculino ou branco aumenta a nota. O NSE também ficou positivamente associado ao desempenho.

Para todas as provas, os Modelos 1 e 2 se mostraram estatisticamente diferentes ($p < 0,001$), sendo este último com valor de AIC menor, o que indica melhor ajuste. A variância explicada no nível da escola ficou superior a 32,0% nas quatro provas. Dada a melhora no modelo, procedemos à construção do Modelo 3. Nele, mantivemos as variáveis de primeiro nível, em virtude de suas significâncias estatísticas, e incluímos as variáveis do nível da escola.

Para as quatro provas, os Modelos 3 e 2 foram estatisticamente diferentes ($p < 0,001$) e observamos uma melhora (o valor de AIC reduziu e a variância explicada aumentou no nível da escola). Os coeficientes das variáveis de primeiro nível ficaram semelhantes nos Modelos 3 e 2. Quanto às variáveis do nível da escola, ser de escola estadual associou-se a uma redução no desempenho, ao passo que ser de escola privada ou federal se associou a um aumento (em relação a ser de escola municipal). A variância explicada no nível da escola ficou superior a 72,3% nas quatro provas.

Para responder a terceira pergunta de pesquisa, se o efeito do NSE sobre o desempenho varia entre escolas, construímos o Modelo 4. A diferença entre os Modelos 3 e 4 é que, neste, o valor do efeito do NSE variou entre escolas. A variância desse efeito variou de 15,0 (LC) a 56,6 (MT). Em todas as provas, o Modelo 4 teve melhor ajuste e foi estatisticamente diferente ($p < 0,001$) do Modelo 3.

Tabela 3 – Parâmetros dos modelos de regressão linear multinível da prova de Ciências Humanas

Variáveis	Modelo 1			Modelo 2			Modelo 3			Modelo 4		
	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta
Efeitos fixos												
Intercepto	563,3*	0,4	-	560,0*	0,3	-	570,3*	2,8	-	570,0*	2,8	-
Atrazo escolar	-	-	-	-23,2*	0,2	-0,10	-23,2*	0,2	-0,10	-23,3*	0,2	-0,10
Sexo masculino	-	-	-	5,1*	0,1	0,03	5,1*	0,1	0,03	5,1*	0,1	0,03
Cor/raça branca	-	-	-	6,3*	0,2	0,04	6,4*	0,2	0,04	6,4*	0,2	0,04
NSE	-	-	-	13,5*	0,1	0,15	12,7*	0,1	0,14	-	-	-
Escola estadual	-	-	-	-	-	-	-23,9*	2,8	-0,12	-24,0*	2,8	-0,12
Escola privada	-	-	-	-	-	-	34,0*	2,8	0,16	34,8*	2,8	0,16
Escola federal	-	-	-	-	-	-	36,8*	3,0	0,09	36,5*	3,0	0,09
Efeitos randômicos												
NSE	-	-	-	-	-	-	-	-	-	-	-	-
Variância do efeito do NSE	-	-	-	-	-	-	-	-	-	-	-	-
Variância (nível do aluno)	4771,3			4643,3			4641,6			4635,4		
Variância (nível da escola)	1828,6			1070,0			493,9			480,1		
Correlação entre o efeito do NSE e o intercepto da escola	-			-			-			0,30		
Estatísticas do modelo												
ICC	0,28			0,19			0,10			0,09		
Variância explicada (nível do aluno)	-			2,7%			2,7%			2,8%		
Variância explicada (nível da escola)	-			41,5%			73,0%			73,7%		
Graus de liberdade	3			7			10			11		
Deviance	10017061,7			9987345,0			9978719,1			9978525,4		
AIC	10017067,7			9987359,0			9978739,1			9978549,4		

Fonte: Elaboração própria.

Nota: * p < .001

Tabela 4 – Parâmetros dos modelos de regressão linear multinível da prova de Ciências da Natureza

Variáveis	Modelo 1			Modelo 2			Modelo 3			Modelo 4		
	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta
Efeitos fixos												
Intercepto	493,2*	0,4	-	488,1*	0,4	-	497,7*	3,0	-	496,5*	2,8	-
Atraso escolar	-	-	-	-16,6*	0,2	-0,08	-16,5*	0,2	-0,07	-16,7*	0,2	-0,08
Sexo masculino	-	-	-	7,8*	0,1	0,05	7,8*	0,1	0,05	7,8*	0,1	0,05
Cor/raça branca	-	-	-	5,8*	0,2	0,04	5,8*	0,2	0,04	5,8*	0,2	0,04
NSE	-	-	-	11,0*	0,1	0,14	10,4*	0,1	0,13	-	-	-
Escola estadual	-	-	-	-	-	-	-26,2*	3,1	-0,15	-25,4*	2,9	-0,14
Escola privada	-	-	-	-	-	-	43,0*	3,1	0,22	43,7*	2,9	0,22
Escola federal	-	-	-	-	-	-	38,6*	3,3	0,11	35,5*	3,1	0,10
Efeitos randômicos												
NSE	-	-	-	-	-	-	-	-	-	10,1*	0,1	0,12
Variância do efeito do NSE	-	-	-	-	-	-	-	-	-	22,5	-	-
Variância (nível do aluno)	3554,1	-	-	3462,6	-	-	3461,8	-	-	3457,2	-	-
Variância (nível da escola)	2064,0	-	-	1403,9	-	-	572,6	-	-	485,5	-	-
Correlação entre o efeito do NSE e o intercepto da escola	-	-	-	-	-	-	-	-	-	0,65	-	-
Estatísticas do modelo												
ICC	0,37	-	-	0,29	-	-	0,14	-	-	0,12	-	-
Variância explicada (nível do aluno)	-	-	-	2,6%	-	-	2,6%	-	-	2,7%	-	-
Variância explicada (nível da escola)	-	-	-	32,0%	-	-	72,3%	-	-	76,5%	-	-
Graus de liberdade	3	-	-	7	-	-	10	-	-	11	-	-
Deviance	9181606,2	-	-	9155990,0	-	-	9146211,9	-	-	9145066,8	-	-
AIC	9181612,2	-	-	9156004,0	-	-	9146231,9	-	-	9145090,8	-	-

Fonte: Elaboração própria.

Nota: * p < .001

Tabela 5 – Parâmetros dos modelos de regressão linear multinível da prova de Linguagens e Códigos

Variáveis	Modelo 1			Modelo 2			Modelo 3			Modelo 4		
	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta
Efeitos fixos												
Intercepto	523,3*	0,4	-	521,4*	0,3	-	534,0*	2,5	-	534,0*	2,6	-
Atraso escolar	-	-	-	-25,0*	0,2	-0,12	-25,0*	0,2	-0,12	-25,0*	0,2	-0,12
Sexo masculino	-	-	-	1,5*	0,1	0,01	1,5*	0,1	0,01	1,5*	0,1	0,01
Cor/raça branca	-	-	-	6,5*	0,1	0,04	6,6*	0,1	0,04	6,6*	0,1	0,04
NSE	-	-	-	15,5*	0,1	0,19	14,8*	0,1	0,19	-	-	-
Escola estadual	-	-	-	-	-	-	-24,7*	2,6	-0,14	-24,7*	2,6	-0,14
Escola privada	-	-	-	-	-	-	25,7*	2,6	0,13	26,3*	2,6	0,14
Escola federal	-	-	-	-	-	-	33,8*	2,7	0,10	33,9*	2,8	0,10
Efeitos randômicos												
NSE	-	-	-	-	-	-	-	-	-	14,8*	0,1	0,19
Variância do efeito do NSE	-	-	-	-	-	-	-	-	-	15,0	-	-
Variância (nível do aluno)	3681,5	-	-	3523,9	-	-	3522,7	-	-	3516,1	-	-
Variância (nível da escola)	1699,8	-	-	881,4	-	-	425,1	-	-	424,1	-	-
Correlação entre o efeito do NSE e o intercepto da escola	-	-	-	-	-	-	-	-	-	0,04	-	-
Estatísticas do modelo												
ICC	0,32	-	-	0,20	-	-	0,11	-	-	0,11	-	-
Variância explicada (nível do aluno)	-	-	-	4,3%	-	-	4,3%	-	-	4,5%	-	-
Variância explicada (nível da escola)	-	-	-	48,1%	-	-	75,0%	-	-	75,1%	-	-
Graus de liberdade	3	-	-	7	-	-	10	-	-	11	-	-
Deviance	9790384,8	-	-	9744833,8	-	-	9736622,0	-	-	9736464,6	-	-
AIC	9790390,8	-	-	9744847,8	-	-	9736642,0	-	-	9736488,6	-	-

Fonte: Elaboração própria.
Nota: * p < .001

Tabela 6 – Parâmetros dos modelos de regressão linear multinível da prova de Matemática

Variáveis	Modelo 1			Modelo 2			Modelo 3			Modelo 4		
	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta	Coefficiente	Erro	Beta
Efeitos fixos												
Intercepto	537,4*	0,6	-	519,7*	0,5	-	530,2*	4,1	-	525,8*	3,8	-
Atraso escolar	-	-	-	-28,8*	0,3	-0,09	-28,8*	0,3	-0,09	-29,0*	0,3	-0,10
Sexo masculino	-	-	-	35,9*	0,2	0,17	35,9*	0,2	0,17	36,0*	0,2	0,17
Cor/raça branca	-	-	-	8,8*	0,2	0,04	8,7*	0,2	0,04	8,7*	0,2	0,04
NSE	-	-	-	18,0*	0,1	0,16	17,1*	0,1	0,15	-	-	-
Escola estadual	-	-	-	-	-	-	-30,8*	4,1	-0,12	-26,8*	3,8	-0,11
Escola privada	-	-	-	-	-	-	53,9*	4,2	0,20	57,1*	3,8	0,21
Escola federal	-	-	-	-	-	-	53,2*	4,4	0,11	47,3*	4,0	0,10
Efeitos randômicos												
NSE	-	-	-	-	-	-	-	-	-	16,4*	0,2	0,15
Variância do efeito do NSE	-	-	-	-	-	-	-	-	-	56,6	-	-
Variância (nível do aluno)	6961,3	-	-	6421,6	-	-	6419,9	-	-	6406,2	-	-
Variância (nível da escola)	3745,7	-	-	2287,6	-	-	1031,1	-	-	845,8	-	-
Correlação entre o efeito do NSE e o intercepto da escola	-	-	-	-	-	-	-	-	-	0,68	-	-
Estatísticas do modelo												
ICC	0,35	-	-	0,26	-	-	0,14	-	-	0,12	-	-
Variância explicada (nível do aluno)	-	-	-	7,8%	-	-	7,8%	-	-	8,0%	-	-
Variância explicada (nível da escola)	-	-	-	38,9%	-	-	72,5%	-	-	77,4%	-	-
Graus de liberdade	3	-	-	7	-	-	10	-	-	11	-	-
Deviance	9738672,8	-	-	9667166,4	-	-	9658468,4	-	-	9656740,2	-	-
AIC	9738678,8	-	-	9667180,4	-	-	9658488,4	-	-	9656764,2	-	-

Fonte: Elaboração própria.
Nota: * p < .001

Discussão

Nosso objetivo principal foi examinar fatores socioeconômicos associados ao desempenho no Enem. Os achados indicam que as diferenças entre escolas explicam as diferenças de desempenho, e as variáveis no nível do aluno e no nível da escola estão associadas ao desempenho. Todas as associações examinadas são estatisticamente significativas, porém sua relevância é variada. Além disso, identificamos que o efeito do NSE sobre o desempenho varia entre escolas.

O atraso escolar se associou negativamente ao desempenho nas quatro provas. Os alunos com pelo menos um ano de atraso escolar tiveram em média de 16,7 (CN) a 29,0 (MT) pontos a menos. Esse resultado corrobora achados de outros estudos que investigaram o efeito do atraso escolar no Saeb (Alves; Ferrão, 2020; Andrade; Laros, 2007; Laros; Marciano; Andrade, 2012). Isso é uma evidência de que essa variável se associa negativamente ao desempenho em avaliações somativas e em testes de seleção de alto impacto.

Neste estudo, a variável sexo se associou a uma diferença em média de 36,0 pontos na prova de MT, com vantagem para os homens. Somente nesta prova o coeficiente padronizado ficou maior do que 0,10. Nas demais, esse coeficiente não passou de 0,05, e ser do sexo masculino aumentou em média 7,8 pontos em CN e 5,1 pontos em CH. Estudos anteriores também identificaram médias de homens superiores em testes de MT em avaliações nacionais (Alves; Ferrão, 2020; Barbetta; Andrade; Tavares, 2018) e internacionais (Aristizabal; Caicedo; Parra, 2017). No entanto, nossos resultados se diferenciam de achados anteriores que apontam vantagens para mulheres em provas da área de LC (Alves; Ferrão, 2020; Aristizabal; Caicedo; Parra, 2017; Barbetta; Andrade; Tavares, 2018). Para essa área, observamos uma pequena diferença de 1,5 ponto em média entre os sexos, com vantagem para os homens. Cabe destacar que o Enem é um exame de alto impacto, ao contrário dos testes utilizados nesses estudos anteriores. Por isso, recomendamos investigações sobre a redução da diferença de desempenho entre sexos ao longo do tempo em testes de alto impacto da área de LC.

O desempenho dos alunos brancos superou o dos demais em menos de 10,0 pontos em média nas quatro provas, controlando as demais variáveis. Estudo anterior verificou que esse grupo teve de 20,1 a 37,2 pontos a mais na prova de MT do Enem 2012 (Travitzki; Ferrão; Couto, 2016). Isso pode ser um indício de redução das desigualdades nesse exame, porém é preciso cautela ao se chegar a essa conclusão. Diferentemente deste trabalho, o estudo citado agrupou os brancos e os amarelos. Além disso, não incluiu o atraso escolar e as características da escola no modelo. Portanto, é possível que a inclusão dessas variáveis reduza o efeito de outras, como a cor/raça. Isso endossa a importância de pesquisas que utilizem as variáveis incluídas neste estudo em tempos diferentes.

A maior desigualdade observada está relacionada à dependência administrativa das escolas, nesse sentido estudar em escola estadual se associa a uma redução de 26,8 pontos em MT, ao passo que estudar em escola privada se associa a um aumento de 57,1 pontos nessa área. Ou seja, alunos dessas duas redes têm em média uma

diferença de aproximadamente 83,9 pontos nessa prova, quando controladas as outras características. O desempenho superior de alunos de escola privada também foi identificado na prova de CN do Enem 2011 (Gomes; Amantes; Jelihovschi, 2020), o que indica que essa desigualdade persiste. Em termos de políticas públicas em educação, esse achado se torna ainda mais relevante ao considerarmos que as redes estaduais devem oferecer prioritariamente o ensino médio (Brasil, 1996).

Quanto ao efeito do NSE, os achados corroboram os de pesquisas anteriores (Alves; Ferrão, 2020; Barbetta; Andrade; Tavares, 2018; Laros; Marciano; Andrade, 2012; Travitzki; Ferrão; Couto, 2016), pois sua associação com o desempenho nas quatro provas foi positivo e com efeito padronizado superior a 0,10. É importante destacar que não há uniformidade no cálculo do NSE nos trabalhos, portanto não é possível saber o impacto que haveria ao incluir ou excluir determinada informação na composição dessa variável. Além disso, alguns estudos incluem essa variável no primeiro nível de análise e outros, no segundo. Apesar dessas diferenças, em todos os estudos o NSE se associou positivamente ao desempenho. Isso mostra que o desempenho acadêmico de uma maneira geral tem sido associado aos recursos (sejam culturais ou materiais) de que a família do aluno dispõe. Dada essa diferença entre os estudos, recomendamos pesquisas que avaliem a variação do efeito do NSE ao longo do tempo e que utilizem o mesmo método de obtenção dessa variável.

O fato de esse efeito variar entre escolas corrobora achados de estudo anterior (Travitzki; Ferrão; Couto, 2016) e reforça a ideia de que em algumas escolas a relação entre o NSE e o desempenho é menor do que em outras. Essa informação é importante, pois nessas escolas pode haver práticas que reduzam desigualdades. Longe de assumirmos que a qualidade educacional se resume ao desempenho no Enem, recomendamos estudos que identifiquem escolas em que o efeito do NSE ficou próximo de zero. Nelas, podem ocorrer práticas pedagógicas que reduzam desigualdades em testes de acesso ao ensino superior.

Destacamos que a prova de MT foi a que apresentou o maior efeito de escola privada, de sexo e a maior variância do efeito do NSE. Isso pode indicar que essa área possui algumas especificidades associadas a essas diferenças. Por isso, recomendamos investigações que estabeleçam relações entre as características dessa prova e os contextos de ensino e aprendizagem que contribuam para elucidar essa peculiaridade.

Por último, destacamos as correlações observadas entre as variáveis socioeconômicas e o desempenho nas quatro provas (Tabela 2). O NSE, ser de escola estadual ou ser de escola privada apresentaram correlações de alta magnitude (Gignac; Szodorai, 2016) com o desempenho nas quatro provas. Ser branco apresentou correlações de magnitude moderada com o desempenho. As correlações do atraso escolar foram de pequena a moderada e as de ser de escola federal, pequenas. Já para o sexo masculino, as correlações foram triviais ou moderada (em MT). Todas essas variáveis possuem sua contribuição única para a explicação de desempenho, observada nos modelos finais de regressão. Independentemente disso, nossos achados indicam que elas por si só apresentam uma associação relevante com o desempenho.

Considerações finais

Neste estudo, verificamos a associação de fatores socioeconômicos ao desempenho nas quatro provas do Enem 2018. O atraso escolar, o NSE dos alunos e a dependência administrativa da escola estiveram associados a diferenças no desempenho nas quatro áreas. Em MT, o sexo também se mostrou uma variável relevante, foi nessa área que observamos diferenças entre os grupos mais expressivas. Este estudo aponta para uma persistente desigualdade no desempenho em testes educacionais de larga escala e a necessidade de se examinar o efeito de algumas variáveis em testes passados. Esperamos que os resultados possam contribuir para o aprimoramento das políticas públicas para a promoção da equidade, aparentemente ainda distante de ser alcançada. Consideramos necessária a manutenção de mecanismos que reduzam as desigualdades no acesso ao ensino superior no Brasil.

Referências bibliográficas

ALVES, M. T. G.; FERRÃO, M. E. Uma década da Prova Brasil: evolução do desempenho e da aprovação. *Estudos em Avaliação Educacional*, São Paulo, v. 30, n. 75, p. 688-720, jan. 2020.

ALVES, M. T. G.; SOARES, J. F.; XAVIER, F. P. Desigualdades educacionais no ensino fundamental de 2005 a 2013: hiato entre grupos sociais. *Revista Brasileira de Sociologia*, Porto Alegre, v. 4, n. 7, p. 49-81, jan./jun. 2016.

ANDRADE, J. M.; LAROS, J. A. Fatores associados ao desempenho escolar: estudo multinível com dados do SAEB/2001. *Psicologia: Teoria e Pesquisa*, Brasília, DF, v. 23, n. 1, p. 33-41, mar. 2007.

ARISTIZABAL, G. C.; CAICEDO, M. C.; PARRA, J. C. M. Factores asociados a la adquisición de competencias en América Latina. *Revista de Ciencias Sociales*, Maracaibo, Venezuela, v. 23, n. 4, p. 33-52, oct./dic. 2017.

BARBETTA, P. A.; ANDRADE, D. F.; TAVARE.S, H. R. Estudo de fatores associados através de regressão quantílica hierárquica. *Estudos em Avaliação Educacional*, São Paulo, v. 29, n. 71, p. 320-349, 31 ago. 2018

BATES, D. et al. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, Innsbruck, Austria, v. 67, n. 1, p. 1-48, 2015.

BONAMINO, A. et al. Os efeitos das diferentes formas de capital no desempenho escolar: um estudo à luz de Bourdieu e de Coleman. *Revista Brasileira de Educação*, Rio de Janeiro, v. 15, n. 45, p. 487-499, dez. 2010.

BRASIL. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. *Diário Oficial da União*, Brasília, DF, 23 dez. 1996. Seção 1, p. 27834-27841.

BRASIL. Lei nº 12.711, de 29 de agosto de 2012. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. *Diário Oficial da União*, Brasília, DF, 30 ago. 2012. Seção 1, p. 1-2.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Microdados do Enem 2018*. Brasília, DF, 2018. Available in: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>>. Access in: 23 mar. 2021.

BROER, M.; BAI, Y.; FONSECA, F. Socioeconomic inequality and educational outcomes: evidence from twenty years of TIMSS. Cham: Springer International Publishing, 2019. v. 5 Available in: <<http://link.springer.com/10.1007/978-3-030-11991-1>>. Access in: 23 mar. 2021.

CHALMERS, R. P. MIRT: a multidimensional item response theory package for the R environment. *Journal of Statistical Software*, Innsbruck, Austria, v. 48, n. 6, p. 1-29, 2012.

COWAN, C. D. et al. *Improving the measurement of socioeconomic status for the national assessment of educational progress: a theoretical foundation*. [S.l]: [s.n.], 2012. Available in: <https://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf>. Access in: 24 mar. 2021.

GIGNAC, G. E.; SZODORAI, E. T. Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, v. 102, p. 74-78, Nov. 2016.

GOMES, C. M. A.; AMANTES, A.; JELIHOVSCHI, E. G. Applying the regression tree method to predict students' science achievement. *Trends in Psychology*, v. 28, n. 1, p. 99-117, mar. 2020.

KARINO, C. A.; LAROS, J. A. Estudos brasileiros sobre eficácia escolar: uma revisão de literatura. *Revista Examen*, Brasília, v. 1, n. 1, p. 95-126, 2017.

LAROS, J. A.; MARCIANO, J. L.; ANDRADE, J. M. Fatores associados ao desempenho escolar em Português: um estudo multinível por regiões. *Ensaio: avaliação e políticas públicas em educação*, Rio de Janeiro, v. 20, n. 77, p. 623-646, dez. 2012.

R CORE TEAM. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019. Available in: <<https://www.R-project.org/>>. Access in: 25 Oct. 2019.

SAMEJIMA, F. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, n. 17, 1969. Available in: <<https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf>>. Access in: 15 ago. 2018.

SENKEVICS, A. S.; MELLO, U. M. O perfil discente das universidades federais mudou pós-lei de cotas? *Cadernos de Pesquisa*, São Paulo, v. 49, n. 172, p. 184-208, abr./jun. 2019.

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

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. *Arquivos Analíticos de Políticas Educativas*, Arizona, USA, v. 24, n. 74, p. 1-32, jul. 2016.

VALENTINI, F. et al. Análise de regressão multinível: exemplo prático. In: SILVA, M. C. R. da et al. (Org.). *Aplicações de métodos estatísticos avançados aplicados à avaliação psicológica e educacional*. São Paulo: Vetor, 2015. p. 135-152.

VIEIRA, I. S. Oportunidades educacionais no Brasil: o que dizem os dados do Saeb. *Estudos em Avaliação Educacional*, São Paulo, v. 30, n. 75, p. 748, jan. 2020.

ZWICK, R.; YE, L.; ISHAM, S. Using constrained optimization to increase the representation of students from low-income neighborhoods. *Applied Measurement in Education*, v. 32, n. 4, p. 281-297, Sept. 2019.

Alexandre Jaloto, doutorando no Departamento de Psicologia da Universidade São Francisco (USF), é psicometrista na Diretoria de Avaliação de Educação Básica (Daeb) do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), e atua nas análises das avaliações educacionais em larga escala em âmbito nacional.

alexandre.jaloto@inep.gov.br

Ricardo Primi, doutor em Psicologia Escolar e do Desenvolvimento Humano pela Universidade de São Paulo (USP), é professor associado do Programa de Pós-Graduação em Psicologia da Universidade São Francisco e membro do comitê diretivo do EduLab21 Centro de Conhecimento do Instituto Ayrton Senna. É membro da Comissão Assessora de Estatística e Psicometria da Diretoria de Avaliação de Educação Básica (Daeb), do Inep, e do grupo de especialistas em questionários para o Pisa 2021, coordenado pelo *Educational Testing Service* (ETS).

rprimi@mac.com

Recebido em 29 de junho de 2021

Aprovado em 5 de novembro de 2021