

Exames de Ingresso no Ensino Superior, Viés de Seleção e *Accountability* Educacional: Uma Análise do ENEM

Reynaldo Fernandes[#], Elaine Toldo Pazello[#], Luiz Guilherme Scorzafave[#] e Leandro Seiti Anazawa^{*}

Resumo

Exames nacionais de ingresso no ensino superior são comuns ao redor do mundo, mas a divulgação dos resultados desses exames, para efeitos de *accountability* educacional, é controversa. O presente artigo avalia o problema de viés de participação no ENEM, buscando responder duas questões: (i) em que medida o viés de seletividade, potencialmente existente, distorce os resultados do ENEM por escola?; e (ii) como é possível melhorar o Enem usando uma outra avaliação educacional? Esse artigo propõe usar as notas do SAEB (um exame *low-stake* realizados na 4^a, 8^a série do ensino fundamental e no 3^o ano do ensino médio) para corrigir a seletividade do ENEM. Concluímos que a seletividade do ENEM não parece ser um problema para que o mesmo seja utilizado para ordenar escolas em um mesmo momento do tempo. A ordenação de escolas muda muito pouco após utilizar o corretor proposto no artigo. Como a proporção de participantes vem crescendo no tempo, o corretor seria importante para análises temporais. Por fim, a adoção de um corretor tradicional (do tipo Heckman- Gronau) produz uma ordenação de escolas que é mais distante da nossa medida de referência (que usa o corretor proposto no artigo) do que não utilizar qualquer corretor.

Palavras-chaves: Viés de seletividade; ENEM; Corretor de participação; Exames de ingresso no ensino superior. JEL: I23, I28.

I - Introdução

Exames nacionais de ingresso no ensino superior são comuns ao redor do mundo.¹ No entanto, a divulgação dos resultados desses exames, para efeitos de *accountability* educacional, por escolas e/ou jurisdições (municípios, distritos ou estados) é controversa.

Os exames de ingresso no ensino superior possuem algumas vantagens sobre os exames *low-stakes*. De modo geral, eles apresentam uma ampla cobertura de disciplinas (indo além de leitura e matemática) e, por serem instrumento de seleção das universidades/faculdades, os alunos tendem a realizá-los com mais comprometimento. A responsabilidade (e interesse) pela inscrição no exame é do próprio estudante, o que reduz a possibilidade de manipulação por parte das escolas. Um problema com a divulgação dos resultados dos exames de ingresso no ensino superior diz respeito ao seu caráter voluntário. Os que realizam o exame são, de modo geral, aqueles que pretendem ingressar no ensino superior, cujo o desempenho tende a ser mais elevado. A consequência seria a superestimação do desempenho médio da escola/jurisdição.

[#] Departamento de Economia – FEARP/USP.

[#] Departamento de Economia – FEARP/USP.

[#] Departamento de Economia – FEARP/USP.

^{*} Doutor em Economia Aplicada pela FEARP-USP.

¹ Por exemplo, SAT/ACT (EUA), A-Level (Reino Unido), Leaving Certificate (Irlanda), Baccalaureate (França), Abitur (Alemanha), Apolytirion (Grécia), Gaokao (China), Konkoor (Irã), PTU (Chile), ICFES (Colômbia) e ENEM (Brasil).

Em relação ao problema de viés de seleção dos exames de ingresso no ensino superior, vale destacar alguns aspectos. Primeiro, a seletividade tende a superestimar o desempenho médio das escolas/jurisdições, mas não necessariamente tem um impacto significativo na ordenação de desempenho das mesmas. Assim, esses exames podem se mostrar uma boa medida de ranking do desempenho médio dos estudantes das escolas/jurisdições.² Segundo, a literatura especializada dispõe de diversos corretores de participação que, em tese, poderiam ser usados.³ Por fim, exames *low-stakes* não estão isentos de viés de participação, uma vez que muitos estudantes não realizam a prova (e.g. por faltarem na escola no dia do exame).

O aumento, em determinados períodos, da taxa de participação dos que realizam o exame (e.g. em períodos de expansão das matrículas no ensino superior) tem sido usado como evidência da falta de validade estatística dos resultados agregados dos exames de ingresso no ensino superior como medida do desempenho dos estudantes das escolas/jurisdições. Grissmer (2000) questiona a publicação dos resultados agregados do SAT (um exame de ingresso usados pelas universidades americanas) por escolas, distritos escolares e estados. Comparando a evolução do SAT com o NAEP (um exame *low-stakes* realizado para uma amostra representativa de estudantes dos EUA) entre 1971 e 1990, ele aponta que enquanto a pontuação do NAEP cresceu a pontuação agregada do SAT apresentou um declínio. Essa queda de pontuação do SAT foi associada ao aumento na taxa de participação: “*In the late 1960s, less than one out of three 17-year-olds had taken an SAT test, whereas by 1990, over 40% had taken an SAT test*”. Admitindo que o NAEP possui validade estatística, Grissmer (2000) conclui que: “*Such use of SAT scores would be beneficial if the scores reflected accurately the quality of schools and school systems. Unfortunately, the aggregate SAT scores, at best, convey no useful information about educational quality and, at worst, convey highly misleading information about educational quality*”.

Um aumento da taxa de participação no exame muda a população de referência e mudanças na população de referência tornam os resultados não comparáveis entre dois períodos de tempo distintos. Entretanto, como apontado anteriormente, isso não significa que, em um período de tempo específico, os resultados agregados não sejam uma boa medida de ranking da proficiência média dos estudantes das escolas/jurisdições. Por outro lado, mudanças na população de referência de uma prova pode ocorrer por diversos motivos e isso é sempre uma preocupação na comparação de resultados de exames ao longo do tempo. Por exemplo, há indicações que as políticas de combate à repetência e à evasão escolar adotadas no Brasil na década de 1990 alterou a população de referência dos que realizavam o SAEB (um exame *low-*

² Clark, Rothstein e Schanzenbach (2009) apresentam, para as escolas do estado de Illinois, estimativas da correlação do desempenho médio dos que realizam as provas do ACT e do desempenho esperado de todos os estudantes da escola e os resultados ficaram entre 0,93 e 0,98. Na próxima seção apresentamos estimativas da correlação entre as pontuações médias das escolas no ENEM e no SAEB (um exame *low-stake* e de caráter amostral que, em tese, estaria menos sujeito ao problema de viés de seleção). Tal correlação é da ordem de 0,92 para matemática e 0,87 para linguagem.

³ Para aplicação de corretores de participação no SAT/ACT nos Estados Unidos ver Dynarski (1987); Dynarski e Gleason (1993); Behrendt, Eisenach e Johnson (1986); Clark, Rothstein e Schanzenbach (2009); e Garlick e Hyman (2021).

stake realizados na 4ª, 8ª série do ensino fundamental e no 3º ano do ensino médio) e isso afetou negativamente a evolução do desempenho observado (Fernandes e Natenzon, 2003).

Exames *low-stakes* podem fazer com que os estudantes ponham pouco esforço na sua realização e esse esforço pode variar entre grupos e no tempo. Assim, um aumento no grau de consequência que o exame acarreta para estudantes e escolas pode tornar o resultado do teste não comparável entre dois períodos de tempo.⁴ Por sua vez, os exames de ingresso no ensino superior, em virtude da pontuação recebida parcialmente determinar se o estudante é admitido pela faculdade a qual se candidata, estão menos sujeitos a flutuações do esforço entre grupos e períodos de tempo.

No Brasil, a crítica à divulgação dos resultados, por escolas/jurisdições, dos exames de ingresso no ensino superior não tem sido diferente. O Exame Nacional de Ensino Médio – ENEM, criado em 1998, passou a ser divulgado por escolas a partir da edição de 2005. Desde a primeira divulgação, essa nova orientação foi alvo de críticas e em 2017 o INEP (órgão do Ministério da Educação encarregado da realização do exame) anunciou que não mais divulgaria os resultados do ENEM por escolas. O viés de participação foi uma das principais justificativas para não mais divulgar os resultados por escolas.

II - Objetivo

O objetivo do presente artigo é o de avaliar o problema de viés de participação no ENEM, buscando responder duas questões: (i) em que medida o viés de seletividade, potencialmente existente, distorce os resultados do ENEM por escola? E (ii) como é possível melhorar o Enem usando uma outra avaliação educacional? Esse artigo propõe usar as notas do SAEB para corrigir a seletividade do ENEM.

III - Metodologia - O problema do viés de participação e as estratégias para contorná-lo

Seja α_{ij} a proficiência do concluinte i da escola j cujo Exame de Ingresso no Ensino Superior busca produzir uma medida. O problema consiste em identificar $E[\alpha_{ij}|J = j] = \alpha_j$, o qual é dado por (1). Em (1), A_{ij} é uma variável indicadora que assume valor 1 quando o estudante participa do exame e 0 quando não participa. Por sua vez, $p_j = P[A_{ij} = 1|J = j] = E[A_{ij}|J = j]$ é a taxa esperada de participação no exame dos concluintes da escola j . Os dados de exames de ingresso permitem, de modo geral, uma estimativa de $E[\alpha_{ij}|J = j, A_{ij} = 1]$ e de p_j , mas não permitem a identificação de α_j e $E[\alpha_{ij}|J = j, A_{ij} = 0]$.

$$\alpha_j = E[\alpha_{ij}|J = j, A_{ij} = 1]p_j + E[\alpha_{ij}|J = j, A_{ij} = 0](1 - p_j) \quad (1)$$

⁴ Estudos têm constatado que, por vezes, a implementação de um programa de *accountability* educacional tende a promover importante crescimento na pontuação do exame adotado pelo programa, mas tal crescimento não é detectado em exames *low-stakes*, sugerindo a presença de um fenômeno que tem sido denominado por *score inflation*: crescimento da medida, mas não dos conhecimentos e habilidades que a medida busca aferir (ver, por exemplo, Koretz e Barron, 1998; Klein et al., 2000; e Figlio e Rouse, 2006). Quando a prova traz consequências para diretores e professores, mas não para os estudantes, a introdução de um programa de *accountability* de escolas pode ter impactos que variam entre escolas: algumas escolas conseguem convencer os estudantes a fazer a prova com mais seriedade enquanto outras não conseguem. Isso aumentaria o problema de erro de medida das estimativas da proficiência média das escolas.

Definindo $D_j = (1 - p_j)(E[\alpha_{ij}|J = j, A_{ij} = 1] - E[\alpha_{ij}|J = j, A_{ij} = 0])$, a equação (1) pode ser reescrita como (2). A estratégia usual para identificar α_j consiste em considerar D_j como uma função de p_j e, então, obter uma estimativa para ele. Assim, $\hat{\alpha}_j = t_j - \hat{D}_j(p_j)$, onde t_j é o desempenho médio dos concluintes da escola j que realizaram o exame e $\hat{D}_j(p_j)$ é o termo de correção de participação. Há duas dificuldades nessa estratégia. A primeira é que a forma funcional de $D_j(p_j)$ é desconhecida e, assim, precisamos impor uma forma funcional para ele. A segunda é que, mesmo de posse da forma funcional de $D_j(p_j)$, α_j não é observado. Assim, temos que substituí-lo por uma variável proxy ou por uma função controle. Em todos os casos surgiria um termo erro que, se correlacionado com p_j , introduziria um viés na estimação dos parâmetros de $D_j(p_j)$, obtida a partir de (2).

$$E[\alpha_{ij}|J = j, A_{ij} = 1] = \alpha_j + D_j \quad (2)$$

O procedimento padrão na literatura de viés de seletividade (abordagem Heckman-Gronau) é considerar que $\alpha_j = X_j\beta + \eta_j$, onde X é um vetor de características das escolas e dos estudantes, β é o vetor de coeficientes e η é o termo de erro. A proposta desse artigo é utilizar as notas do SAEB como proxy para α_j . Considerando a correção com base no SAEB como *benchmark*, iremos avaliar o desempenho da correção com base no corretor tradicional. Queremos avaliar se é melhor corrigir usando o corretor Heckman-Gronau ou se é melhor não fazer nada, ou seja, simplesmente usar o ENEM observado.

IV - Resultados

Além dos dados do ENEM e SAEB, utilizamos o Censo Escolar para variáveis relativas à qualidade da escola. Todos os dados são de 2011⁵. Como o SAEB de 2011 é amostral, trabalhamos com um subconjunto das escolas que compõem a base do ENEM, ou seja, só aquelas que faziam parte da amostra do SAEB (1890 escolas, correspondendo a 7,4% das escolas que compõem nosso banco do ENEM 2011). Desse total, excluímos as escolas com menos de dez alunos concluintes participando do ENEM. Todos os exercícios são realizados com esse subconjunto de escolas – 1654 escolas.

O principal exercício realizado no presente artigo consiste em corrigir as notas das escolas no ENEM com base no desempenho dessas escolas no SAEB. O corretor de participação é obtido a partir da estimação de uma equação derivada da equação (2). Para definir a especificação do corretor de participação ($D(p)$), foram consideradas 8 especificações para $D(p)$ e adotamos um procedimento de validação cruzada para definir a melhor especificação (menor raiz quadrada do erro quadrado médio). As estimativas de benchmark selecionaram o polinômio $1/p$ de ordem 2.

⁵ Enquanto o SAEB 2011 tem caráter amostral, é possível identificar todas as escolas da amostra e, assim, pareá-las com o banco de dados do ENEM. Para o SAEB de 2013 e 2015 não é mais possível identificar as escolas participantes. Para 2017, quando o SAEB se torna censitário para as escolas públicas, só é possível identificar as escolas que tenham tido pelo menos dez alunos presentes na aplicação e tenham alcançado a taxa de participação de, pelo menos, 80% dos alunos matriculados. Apenas 36,3% das escolas participantes em 2017 possuíam identificação no banco de dados. Em 2011 o ENEM foi aplicado dias 22 e 23 de outubro, enquanto o SAEB foi realizado entre 7 e 18 de novembro.

O gráfico 1 a seguir mostra o impacto da correção sobre a nota média da escola (nota_corrigida – nota_observada) de acordo com a taxa de participação no Enem. As escolas de menor participação são as mais penalizadas, com perdas de 20 pontos, por exemplo, para as escolas com taxas de participação da ordem de 20% no Enem.

Gráfico 1: Impacto da correção na nota média de matemática – modelo benchmark



No entanto, o impacto sobre a distribuição geral de notas é pequeno como pode ser observado pelos gráficos a seguir. O primeiro gráfico mostra as densidades das notas antes e depois da correção; e o segundo, a variação na ordenação das escolas de acordo com a taxa de participação.

Gráfico 2: Impacto da correção sobre a densidade da nota média de matemática

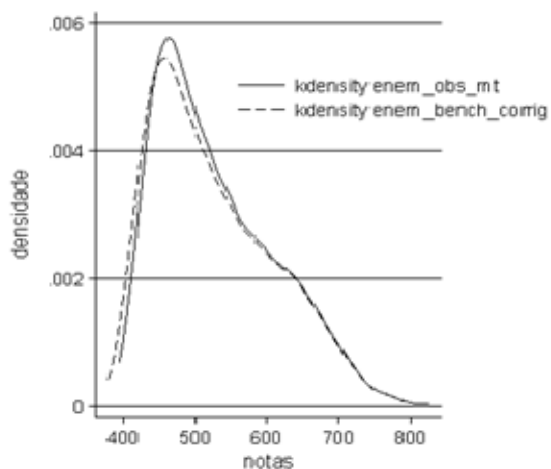
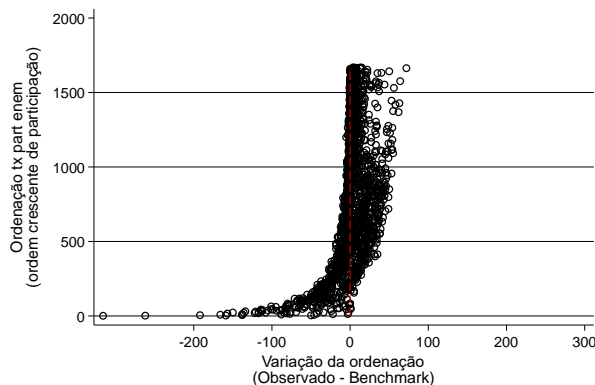


Gráfico 3: Impacto da correção sobre a ordenação das escolas de acordo com a taxa de participação



V - Conclusão

O presente artigo avaliou o problema de viés de participação no ENEM, propondo usar o SAEB como medida da habilidade não-observada dos alunos para corrigir a seletividade. A nota média observada no ENEM de uma determinada escola pode ser escrita como a nota média de todos os estudantes da escola (variável latente) mais um termo de correção. Como colocado, o artigo propôs usar justamente o desempenho das escolas no SAEB como uma medida do desempenho latente dos alunos. Embora no SAEB uma parte dos alunos também não compareça para realizar a prova, a direção do viés não é clara. Comparamos nossos resultados com os obtidos quando usamos o corretor tradicional (Heckman-Gronau), que é o procedimento padrão nessa literatura de viés de seletividade.

As análises realizadas sugerem que a seletividade do ENEM parece não impedir que o mesmo seja utilizado para ordenar escolas num mesmo momento do tempo. A correlação entre as notas corrigidas pelo SAEB – nosso resultado benchmark – e as notas observadas é de 0,9970. As principais mudanças de ordenação são para as escolas de menor nota média, menor taxa de participação no ENEM e com perfil mais difícil de alunado. Além disso, nossos resultados mostraram que na ausência de uma prova do tipo SAEB, o melhor a se fazer é usar as notas observadas. O corretor tradicional com base nas características típicas de qualidade da escola e dos estudantes não teve bom desempenho para corrigir as notas.

Referências

- Ahn, H. e Powell, J. (1993). Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism. *Journal of Econometrics*, 58, 3–29.
- Behrendt, A.; Eisenach, J. e Johnson, W. R. (1986). Selectivity Bias and the Determinants of SAT Scores. *Economics of Education Review*, 5 (4), 363-371.
- Clark, M.; Rothstein, J. e Schanzenbach, D. W. (2009). Selection Bias in College Admissions Test Scores. *Economics of Education Review*, 28 (3), 295-307.
- Dynarski, M. (1987). The Scholastic Aptitude Test: Participation and performance. *Economics of Education Review*, 6 (3), 263–273.
- Dynarski, M. e Gleason, P. (1993). Using Scholastic Aptitude Test Scores as indicators of state educational performance. *Economics of Education Review*, 12 (3), 203–211.
- Fernandes, R. e Gremaud, A. (2009). Qualidade da Educação: Avaliação, Indicadores e Metas. In: Veloso, F., Pessoa, S., Henriques, R. e Giambiagi, F. (orgs). *Educação Básica no Brasil: Construindo o País do Futuro*. Editora Campus/Elsevier, 213-238.
- Fernandes, R. e Natenzon, P. E. (2003). *A Evolução Recente do Rendimento Escolar das Crianças Brasileiras: Uma Reavaliação dos Dados do SAEB*. *Estudos em Avaliação Educacional*, 28, 3-22.
- Figlio, D., Rouse, C. E. (2006). Do Accountability and Voucher Threats Improve Low-Performing Schools? *Journal of Public Economics*, 90 (1–2), 239–255.
- Garlick, R. e Hyman, J. (2021). Quasi-Experimental Evaluation of Alternative Sample Selection Corrections. *Journal of Business & Economic Statistics*, publicado online em 16 de abril.
- Grissmer, D. W. (2000). The Continuing Use and Misuse of SAT Scores. *Psychology, Public Policy, and Law*, 6 (1), 223-232.
- Gronau, R. (1974). Wage Comparisons – A Selectivity Bias. *Journal of Political Economy*, 82(6), 1119-1143.
- Heckman, J. (1974). Shadow Prices, Market Wages, and Labor Supply. *Econometrica*, 42(4), 679-694.
- Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161.
- Klein, S., Hamilton, L., McCaffrey, D. e Stecher, B. (2000). What Do Test Scores in Texas Tell Us? *Educational Policy Analysis Archives*, 9 (49), 1-22.
- Koretz, D. M., Barron, S. (1998). *The Validity of Gains on the Kentucky Instructional Results Information System (KIRIS)*. RAND Corporation, Working paper.
- Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Economics*, New York: Cambridge University Press.
- Powell, J. (1987). *Semiparametric Estimation of Bivariate Latent Variable Models*. Working Paper 8704, Social Systems Research Institute, University of Wisconsin, Madison.
- Puhani, P. (2002). The Heckman Correction for Sample Selection and its Critique. *Journal of Economic Surveys*, 14 (1), 53-68.