

# O efeito do *background* familiar sobre o desempenho dos estudantes do ensino superior<sup>1</sup>

**Resumo:** Este estudo tem como objetivo estudar o efeito do *background* familiar sobre o desempenho acadêmico dos estudantes do Ensino Superior. Utilizamos os microdados do Exame Nacional de Desempenho dos Estudantes (ENADE) e o modelo de Klein e Vella (2009), o qual controla os problemas de simultaneidade. Os resultados evidenciaram que o *background* familiar exerce grande influência no desempenho dos estudantes do Ensino Superior, em particular, a renda da família e a escolaridade do pai.

**Palavras-chave:** Desigualdade; Desempenho escolar; Background familiar.

**Abstract:** *This study aims to study the effect of family background on the academic performance of higher education students. We used microdata from the National Student Performance Examination (ENADE) and the Klein and Vella (2009) model, which controls for concurrency problems. The results showed that the family background exerts a great influence on the performance of higher education students, in particular, the family income and the father's education.*

**Keywords:** *Inequalities; School performance; Family background.*

**Classificação JEL:** I24; C36; B23.

Romilson do Carmo Moreira<sup>2</sup>

Adalmir Marquetti<sup>3</sup>

Carlos Enrique Carrasco Gutierrez<sup>4</sup>

<sup>2</sup> Universidade do Estado da Bahia  
E-mail: rcmoreira@uneb.br.

<sup>3</sup> PUCRS - Pontifícia Universidade Católica do Rio Grande do Sul  
E-mail: aam@pucrs.br.

<sup>4</sup> Universidade Católica de Brasília  
E-mail: carlosenrique@p.ucb.br.

<sup>1</sup> Agradecemos à Fundação de Apoio à Pesquisa e Inovação do Distrito Federal (FAPDF) pelo financiamento desta pesquisa.

## 1. Introdução

As disparidades do desempenho educacional e do aprendizado dos estudantes brasileiros têm origem em fatores decorrentes do baixo padrão de qualidade do sistema de ensino e das diferenças sociais e econômicas vigentes entre as famílias.

A literatura tem investigado os efeitos das condições sociais e econômicas dos indivíduos e do seu núcleo familiar sobre o desempenho educacional. Barros e Mendonça (1995), Barros et al (2001) e Curi et al. (2013) apresentam evidências que as condições socioeconômicas e o conjunto de aspectos relacionados à origem da família aumentam as probabilidades de sucesso no mercado de trabalho e de melhor desempenho nos exames educacionais. Os seus resultados evidenciaram que os indivíduos oriundos de famílias de maior renda e pais com maior escolaridade possuem vantagens competitivas quando comparados com indivíduos sem essas características. Os jovens de melhor origem social têm maior probabilidade de frequentar todo o ciclo escolar. Entretanto, os indivíduos com pais de menor nível de instrução e renda foram mais vulneráveis aos processos de evasão, defasagem nas séries iniciais e apresentaram pior desempenho nas avaliações.

As características associadas à origem da família têm se revelado um importante componente na determinação do desempenho educacional. Essas características são conhecidas como *background* familiar e captam de forma combinada informações sobre a estrutura socioeconômica da família no qual os indivíduos se desenvolveram. São insumos relacionados à estrutura social da família, aos níveis de escolaridade da mãe e do pai e sobre estrato ocupacional dos pais.

O presente trabalho estuda o efeito do *background* familiar sobre o desempenho educacional dos alunos do Ensino Superior, investigamos se as variáveis de origem familiar como a renda e a escolaridade dos pais afetam o desempenho individual. Para tal, utilizamos os microdados do Exame Nacional de Desempenho dos Estudantes, ENADE, do ano 2017 e o modelo de equações simultâneas apresentado por Klein e Vella (2010).

A literatura analisa, em grande parte, os efeitos do *background* familiar no desempenho dos alunos no ensino básico e médio. Há poucos trabalhos que estudam os efeitos da origem da família no desempenho no Ensino Superior. Portanto, o trabalho traz uma importante contribuição para a literatura ao abordar o ensino terciário no Brasil.

A hipótese do presente estudo afirma que há um efeito da origem da família no desempenho estudantil dos alunos do Ensino Superior, sendo o principal canal de transmissão a dependência inerente entre pais e filhos. Em outras palavras, filhos de pais com maior educação e renda podem ter um desempenho educacional melhor do que filhos com pais de baixa escolaridade e renda. Em particular, é plausível considerar que a dependência é maior quando os filhos moram com os pais. Para estudantes que não moram mais com os seus pais essa dependência ainda existe em menor grau. Por exemplo, famílias de renda elevada podem continuar ajudando financeiramente seus filhos e, conseqüentemente influenciar no seu desempenho educacional. Pais com maior escolaridade podem prover uma melhor orientação educacional a seus filhos do que os pais com reduzida escolaridade mesmo quando os filhos estiverem morando fora de casa.

Em relação a metodologia, utilizamos o método de Klein e Vella (2010) para evitar regressões de natureza endógena que geralmente ocorrem no contexto de variáveis que capturam o *background* familiar. A maioria dos trabalhos empíricos sobre desempenho educacional utilizam a função de produção educacional (Hanushek 1970, 2007) onde o estimador de mínimos quadrados ordinários não controla a endogeneidade. Nesses estudos a endogeneidade ocorre ao se supor que quanto maior o nível educacional dos pais, maior será o nível de habilidade dos filhos. De modo similar, também podemos

supor que quanto maior o nível de habilidades dos filhos, maior o nível de renda. A estratégia desenvolvida por Klein e Vella (2010) dispensa a utilização do método de mínimos quadrados ordinários e de variáveis instrumentais. Figueirêdo et al. (2014) é uma aplicação muito interessante deste método no trabalho, os autores estudam os efeitos das circunstâncias sociais no desempenho educacional dos alunos que prestam o Exame Nacional do Ensino Médio, ENEM.

Outra vantagem do método de Klein e Vella (2010) é a identificação do efeito direto da qualificação dos pais que refletem o efeito indireto de variáveis não observadas no desempenho dos filhos. Mais especificamente, o efeito direto está relacionado a variáveis não observáveis relacionadas a educação dos pais. Essas podem incluir outras características que tem um impacto no desempenho educacional dos filhos, tais como motivação, transferência de habilidade intergeracional não observada, talentos, entre outras. Os pais mais qualificados os tornam um tipo "diferente de pai", e, assim, induzem seus filhos a resultados educacionais mais elevados. Assim, a existência dessa componente indica que a educação dos pais é um fator não exógeno na educação dos filhos e que fatores não observáveis que afetam o desempenho educacional podem ser correlacionados entre as gerações.

Os resultados do presente estudo evidenciam o efeito do *background* familiar no desempenho educacional dos alunos do Ensino Superior, em particular, as variáveis associadas à renda familiar e a escolaridade do pai.

O trabalho está organizado em quatro seções, contando a introdução. Na seção 2 apresentamos o modelo econométrico de Klein e Vella (2009, 2010) e a estratégia de estimação. Na seção 3 descrevemos o banco de dados e discutimos os resultados das estimativas. Finalmente, na seção 4, concluímos o presente estudo.

## 2. Revisão de literatura

A literatura tem mostrado que além dos fatores que compõem a oferta educacional, o *background* familiar é um importante fator para explicar o resultado educacional. Um dos primeiros estudos científicos sobre o impacto das condições da família e as ofertas educacionais no aprendizado na escola foi elaborado por Coleman et al. (1966). Os autores, utilizando uma amostra de 567.148 alunos e professores em mais de 3.000 escolas nos Estados Unidos, mostraram que os efeitos do *background* familiar foram mais importantes do que os escolares na determinação do desempenho educacional dos alunos. Lee e Barro (1997), Hanushek (1996), entre outros estenderam as pesquisas em diferentes dimensões, chegando a conclusões similares aos de Coleman et al. (1966).

Estudos realizados para realidade brasileira sobre os determinantes educacionais têm mostrado que o ambiente familiar possui grande influência no resultado educacional. Alguns trabalhos relacionados ao tema são: Silva e Hasenbalg (2002), Riani e Rios-Neto (2008), Menezes Filho (2007), Rocha, Teixeira e Mello (2015), Mendes e Karruz (2016), Vieira, Nogueira e Souza (2017) e Souza, Oliveira e Annegues (2018).

Silva e Hasenbalg (2002) usando a base de dados da Pesquisa Nacional por Amostragem a Domicílio, PNDA, para 1999 mostraram que os fatores de *background* familiar, principalmente a educação dos pais, possuem grande influência no resultado educacional. Em particular, a origem da família e o contexto social têm sido variáveis estratégicas para explicar as chances de progressão escolar e de sucesso na carreira. Por outro lado, Riani e Rios-Neto (2008) estudaram os determinantes dos resultados educacionais nos ensinos fundamental e médio no Brasil. Os autores consideraram fatores relacionados ao *background* familiar e a estrutura escolar dos municípios. Os resultados evidenciaram a grande importância da educação materna nos indicadores

educacionais, revelando que a trajetória escolar do aluno está relacionada com a sua origem social.

Menezes Filho (2007) considerando as variáveis relacionadas ao *background* familiar e as condições de oferta educacional conclui que o desempenho as características familiares e dos alunos, educação da mãe, número de livros, acesso a computador e cor do aluno são as que mais explicam o desempenho escolar de alunos da quarta e da oitava séries. Em outro estudo, Rocha, Teixeira e Mello (2015) ao investigarem os fatores que afetam o aprendizado escolar dos alunos do ensino fundamental no Rio Grande do Norte encontram relação de causalidade entre as variáveis qualificação do corpo docente, insumos escolares e *background* familiar no desempenho dos estudantes. Os resultados apontaram que o *background* familiar exerce maior impacto no desempenho dos estudantes, com grande destaque para o efeito do nível de escolaridade da mãe.

Figueirêdo et al. (2014) estudaram se a origem familiar tem papel relevante no desempenho dos alunos que concluíram o Ensino Médio no Brasil e no Nordeste, usando os microdados do ENEM e do Sistema de Avaliação da Educação Básica, SAEB, para 2010. Os resultados indicaram que variáveis como renda familiar e escolaridade dos pais são fatores fundamentais para o desempenho dos alunos.

Mendes e Karruz (2016) testaram a relação entre o *background* familiar e o desempenho educacional no ENEM para o ano 2012. Os resultados indicaram que quanto maior a escolaridade dos pais e a renda familiar, melhor foi o desempenho dos indivíduos no ENEM.

Souza, Oliveira e Anegues (2018) pesquisaram o efeito do *background* familiar sobre a proficiência escolar com os microdados do SAEB de 2011 e do Censo Escolar. O público alvo foram os de alunos do quinto e do nono anos do ensino fundamental de escolas das redes pública e privada. Os resultados mostraram que as variáveis ligadas ao *background* familiar exerceram grande influência no desempenho dos alunos em comparação com as variáveis ligadas à escola, aos docentes e aos diretores, sobretudo para os estudantes de escolas públicas.

Vieira, Nogueira e Souza (2017) investigaram quais os fatores são essenciais na determinação do desempenho educacional dos estudantes matriculados nas universidades federais localizadas nas regiões de menor desenvolvimento econômico. Os resultados apontaram forte influência das condições sociais no desempenho dos estudantes. Também, verificaram que os modelos captaram o elevado esforço necessário dos estudantes de baixa renda para atingir um desempenho médio nos exames. Esses estudantes necessitam de esforço e dedicação para superar as condições adversas.

Como visto acima, a maioria dos trabalhos sobre o tema utiliza os dados para os estudantes do Ensino Fundamental ou Ensino Médio. Os resultados indicam que o *background* familiar exerce grande influência nos resultados educacionais.

### 3. Procedimentos metodológicos

A Função de Produção Educacional (FPE) foi utilizada nos trabalhos de Krueger (1999), Rios-Neto et. al (2002), Hanushek (1970, 2002), Hanushek et al. (2007), entre outros, como ponto de partida para estudar os impactos de algumas variáveis no resultado educacional. Hanushek (1970, 2002) define a FPE por:

$$D_i^t = f(F_i^t, P_i^t, I_i^t, S_i^t) \quad (1)$$

Em que  $D_i^t$  é a variável escolaridade que capta o desempenho educacional do estudante  $i$  no tempo  $t$ ;  $F_i^t$  é vetor com as características individuais do estudante em  $t$ ;  $P_i^t$  é o vetor que capta as variáveis socioeconômicas e do *background* familiar em  $t$ ;

$I_i^t$  é o vetor de dotações iniciais do estudante que identifica as habilidades individuais;  $S_i^t$  é o vetor que capta as características e os insumos da escola em  $t$ .

Apesar do emprego da Função de Produção Educacional-FPE em muitos trabalhos, o uso dessa abordagem na estimação do modelo traz algumas limitações. Uma crítica é a presença de endogeneidade entre as variáveis. A FPE não considera a simultaneidade entre a renda dos pais e o nível de habilidades dos filhos. É razoável supor que quanto maior o nível educacional dos pais, maior será o nível de habilidade dos filhos. Também podemos supor que quanto maior o nível de habilidades dos filhos, maior o nível de renda familiar. A aplicação do método de mínimos quadrados ordinários na FPE produzirá estimadores viesados e inconsistentes. O procedimento mais utilizado para tratar dos problemas de endogeneidade é o método de variáveis instrumentais. Embora muito utilizada, há frequentemente discordância na literatura sobre as restrições impostas em aplicações empíricas específicas, em particular, quando ocorrem escolhas de instrumentos considerados fracos.

Para tratar do problema de endogeneidade adotamos o modelo simultâneo de transferência no desempenho educacional proposto por Klein e Vella (2009, 2010). Figueirêdo et al. (2014) utilizaram esse modelo para determinar como as circunstâncias sociais influenciam o desempenho educacional dos alunos que prestaram o ENEM. A seguir apresentamos o modelo de Klein e Vella (2009):

$$D_i^f = \beta_{10} + \beta_1 X_i + \beta_m E_i^m + \beta_p E_i^p + u_i, \quad i = 1, \dots, N \quad (2)$$

$$E_i^m = \beta_{20} + \gamma_m X_i + v_i^m \quad (3)$$

$$E_i^p = \beta_{20} + \gamma_p X_i + v_i^p \quad (3')$$

Em que  $D_i^f$  é a variável relacionada ao desempenho educacional do estudante,  $E_i^m$  denota o nível educação da mãe,  $E_i^p$  o nível de educação do pai,  $X_i$  representa um vetor de variáveis exógenas, que em geral se assume idêntico nas equações de desempenho e nas equações dos pais. Os  $\beta_s$  e os  $\gamma_j$  são parâmetros desconhecidos,  $u_i$  e  $v_i^j$  são os erros do modelo com a seguinte propriedade:

$$E(v_i^j | X_i) = E(u_i | X_i) = 0 \quad j = m, p \quad (4)$$

$$\text{corr}(u_i | v_i^j) \neq 0 \quad j = m, p \quad (5)$$

A endogeneidade na variável  $E_i^j$  pode ter como causa a correlação entre  $u_i$  e  $v_i^j$ , e quando estimado pelo método de MQO produz estimadores viesados e inconsistentes. O procedimento apresentado por Klein e Vella (2009) sugere eliminar a correlação do erro da equação (2) com o erro da equação (3), para os valores de  $j=m,p$ . Assim, o erro da equação (2) pode ser reescrito como:

$$u_i = \lambda^m v_i^m + \lambda^p v_i^p + \varepsilon_i \quad (6)$$

Em que  $\lambda^j = \frac{\text{cov}(v_i^j, u)}{\text{var}(v_i^j)}$ , para  $j=m,p$ . A função controle definida em (6) é o método de estimação alternativo ao de variáveis instrumentais para este modelo. A equação (2) pode ser reescrita por:

$$D_i^f = \beta_{10} + \beta_1 X_i + \beta_m E_i^m + \beta_p E_i^p + \lambda^m v_i^m + \lambda^p v_i^p + \varepsilon_i \quad (7)$$

Em que o termo erro  $\varepsilon_i$  não é correlacionado com as variáveis independentes do modelo  $M = [1, X_i, E_i^m, E_i^p, v_i^m, v_i^p]$ . A estimação da equação (7) não é possível com a ausência de restrições da forma reduzida apresentada em (3) que garantam que a matriz  $M$  tenha o posto máximo. Assumimos que os erros da distribuição dependem

de  $X_i^j$ , ou seja, consideraremos a existência da heteroscedasticidade. Os coeficientes da equação (6) são expressos como:

$$\lambda^j = \frac{\text{cov}(v^j, u | X)}{\text{var}(v^j | X)} = A^j(X_i) \quad \text{para } j = m, p$$

No entanto, Klein e Vella (2009) observam que, quando a distribuição do termo erro não depende de  $X_i$  podemos condicionar em  $X_i$  fazendo o erro da equação (6) como:

$$u_i = \lambda^m v_i^m + \lambda^p v_i^p + \varepsilon_i = A^m(X_i)v_i^m + A^p(X_i)v_i^p + \varepsilon_i \quad (8)$$

Para identificar o modelo os autores assumiram a presença de heteroscedasticidade impondo restrições adicionais entre os termos do erro. Mais exatamente, consideramos as funções  $S_u^2(X_i) = \text{Var}(u | X)$  e  $S_{v^j}^2(X_i) = \text{Var}(v^j | X)$  que denotam as funções de variância condicional de  $u$  e  $v$  respectivamente. A estrutura dos erros pode ser definida como:

$$u_i \equiv S_u(X_i)u_i^* \\ v_i^j \equiv S_{v^j}(X_i)v_i^{j*}, \quad j = m, p$$

Em que  $u_i^*$  and  $v_i^{j*}$  são erros homocedásticos.

$$E(u_i^* | X) = E(X | v_i^{j*}) = 0 \\ \rho^j \equiv E(u_i^* v_i^{j*} | X) = E(u_i^* v_i^{j*}) \quad j = m, p \quad (9)$$

E a correlação  $\rho^j$  é uma constante, sendo a identificação possível se a relação é satisfeita:

$$\lambda^j = A^j(X_i) = \rho^j \frac{S_u(X_i)}{S_{v^j}(X_i)} \quad \text{para } j = m, p$$

$A^j$  é conhecido como uma função não linear de  $X_i$  e a não linearidade em  $A^j(X_i)$  é uma fonte de identificação do modelo. Klein e Vella (2009) mostra que isso pode ser feito impondo para o erro  $u_i$  a estrutura:

$$u_i = A^m(X_i)v_i^m + A^p(X_i)v_i^p + \varepsilon_i = \rho^m \frac{S_u(X_i)}{S_{v^m}(X_i)} v_i^m + \rho^p \frac{S_u(X_i)}{S_{v^p}(X_i)} v_i^p + \varepsilon_i \quad (10)$$

Considerando esta última expressão para o erro, finalmente o modelo a ser estimado será:

$$D_i^f = \beta_0 + \beta_1 X_i + \beta_m E_i^m + \beta_p E_i^p + \rho^m \frac{S_u(X_i)}{S_{v^m}(X_i)} v_i^m + \rho^p \frac{S_u(X_i)}{S_{v^p}(X_i)} v_i^p + \varepsilon_i \quad (11)$$

### 3. Estratégia empírica de identificação

Klein e Vella (2010) propõem um estimador para o modelo apresentado acima. Seguimos o procedimento de Farré et al. (2013) para identificar as funções  $S_u$ ,  $S_{v^m}$  e  $S_{v^p}$ :

$$S_u = \sqrt{\exp(Z_u \theta_1)} \quad (12)$$

$$S_{v^j} = \sqrt{\exp(Z_{v^j} \theta_{2j})} \quad \text{para } j = m, p \quad (13)$$

Em que  $Z_u$  e  $Z_{vj}$  são vetores das variáveis que produzem heteroscedasticidade em cada equação e  $\theta_j$  um vetor de parâmetros desconhecidos para  $i=1,2$  e  $j=m,p$ . Em particular,  $Z_u$  e  $Z_{vj}$  podem ser as variáveis consideradas em  $X_i$ . Estimando essas funções podemos definir o espaço das variáveis como:

$$\tilde{M} = \left[ 1, X_i, E_i^m, E_i^p, \frac{S_u}{S_{vm}} \hat{v}_i^m, \frac{S_u}{S_{vp}} \hat{v}_i^p \right]$$

A estimação agora é possível com a matriz  $\tilde{M}$  que apresenta posto cheio devido a não linearidade induzida pelas covariadas  $X_i$ . O modelo da função controle é estimado de forma robusta. O procedimento de estimação se distribui nas seguintes etapas:

i) Estime por mínimos quadrados ordinários a equação da mãe e do pai, apresentadas em (3), nas covariadas  $X_i$  e obtenha os resíduos  $\hat{v}_i^m$  e  $\hat{v}_i^p$  respectivamente.

ii) Assumindo que a forma funcional para  $S_{vj}$  é

$$S_{vj} = \sqrt{\exp(Z_{vj}\theta_{2j})} \quad \text{para } j = m, p$$

estime o parâmetro  $\theta_{2j}$  usando o método de mínimo quadrados não lineares para a variável dependente  $\ln[(\hat{v}_i^m)^2]$ . Logo encontre as estimativas para as funções:

$$\hat{S}_{vm}(X_i) = \sqrt{\exp(X_j\theta_{2m})} \quad \hat{S}_{vp}(X_i) = \sqrt{\exp(X_j\theta_{2p})}$$

iii) Assume a seguinte forma funcional para  $S_u$ :

$$S_u = \sqrt{\exp(Z_u\theta_1)}$$

iv) Substituindo os valores de  $\hat{v}_i^m, \hat{v}_i^p, S_u, \hat{S}_{vm}$  e  $\hat{S}_{vp}$  no modelo apresentado pela equação 10, temos a equação a ser estimada:

$$D_i^f = \beta_0 + X_i'\beta_1 + \beta_m E_i^m + \beta_p E_i^p + \rho^m \left( \sqrt{\exp(z_{ui}\theta_u)} \right) \frac{\hat{v}_i^m}{\hat{S}_{vm}(X_i)} + \rho^p \left( \sqrt{\exp(z_{ui}\theta_u)} \right) \frac{\hat{v}_i^p}{\hat{S}_{vp}(X_i)} + \varepsilon_i$$

v) Estime os parâmetros do modelo usando o método de mínimos quadrados não lineares:

$$\min_{(\beta_0, \beta_1, \beta_m, \beta_p, \rho^m, \rho^p, \theta_u)} \sum_{i=1}^n \left( \begin{aligned} & D_i^f - \beta_0 - X_i'\beta_1 - \beta_m E_i^m - \beta_p E_i^p - \rho^m \left( \sqrt{\exp(z_{ui}\theta_u)} \right) \frac{\hat{v}_i^m}{\hat{S}_{vm}(X_i)} + \\ & \rho^p \left( \sqrt{\exp(z_{ui}\theta_u)} \right) \frac{\hat{v}_i^p}{\hat{S}_{vp}(X_i)} + \varepsilon_i \end{aligned} \right)^2$$

A metodologia de Klein e Vella (2010) é usada para corrigir possíveis vieses decorrentes do uso do método de mínimos quadrados ordinários (MQO) devido a especificação errada do modelo e para evitar o uso de variáveis instrumentais em aplicações em que os instrumentos não estão facilmente disponíveis, como é o caso deste estudo. Outra vantagem do método de Klein e Vella (2010) é a identificação do efeito direto da educação dos pais filtrada pelos efeitos indiretos das variáveis não observadas representadas pelos parâmetros  $\rho^m$  e  $\rho^p$ .

## 4. Resultados

### 4.1 Descrição dos dados

A base de dados utilizada corresponde aos microdados do Exame Nacional de Desempenho do Estudante (ENADE) para 2017. Os dados contemplam um conjunto de características com objetivo de captar o perfil socioeconômico dos estudantes, sobretudo com informações da sua origem familiar, o nível de escolaridade dos pais e a renda da família, variáveis fundamentais para a construção do *background* familiar.

O ENADE é utilizado para avaliar o rendimento dos alunos do último ano dos cursos de graduação em relação aos conteúdos programáticos previstos nas diretrizes curriculares dos cursos, o desenvolvimento de competências e habilidades necessárias ao aprofundamento da formação geral e profissional, bem como o nível de atualização dos estudantes com relação à realidade brasileira e mundial. Os resultados do ENADE, aliados às respostas do Questionário do Estudante, são insumos para o cálculo dos Indicadores de Qualidade da Educação Superior usados para medir o conceito ENADE nas instituições de ensino superior, atribuindo notas entre 1 a 5. O Quadro 1 apresenta uma breve descrição das variáveis.

A Tabela 1 apresenta a distribuição de frequências das principais variáveis utilizadas na análise. Após o processo pré-tratamento dos dados, excluindo os missing e outliers, o número disponível para a estimação econométrica foi de 430.008 observações. Os dados da tabela indicam que a participação feminina no ENADE foi de aproximadamente 56%. Quando comparamos o desempenho, os estudantes do sexo feminino obtiveram média de 45,15 pontos, enquanto os estudantes masculinos tiveram a média de 44,00 pontos.

Quando a análise considera o critério raça, verificamos que os estudantes que se autodeclararam brancos representam aproximadamente 54% do total dos indivíduos que fizeram o ENADE em 2017. Comparando a nota média no exame, os estudantes brancos obtiveram 3,08 pontos a mais do que os não brancos. Quando os resultados são analisados por tipo de escola, 73,02% dos estudantes que prestaram o exame eram egressos do ensino médio público e tiveram nota média inferior em 5,36 pontos em relação aos estudantes egressos do ensino privado.

Quando a análise é pelo tipo de universidade, observa-se que aproximadamente 65% eram alunos matriculados em universidade particulares. Comparando a nota de desempenho, os estudantes das universidades públicas obtiveram média 4,9 pontos superior aos estudantes das universidades particulares.

O recorte dos dados por regiões mostra que 68,08% dos estudantes que fizeram o exame estavam nas Regiões Sudeste e Sul, enquanto as Regiões Norte e Centro Oeste representavam 6,71% e 7,52% do total dos estudantes. Analisando o desempenho pela nota média, as médias dos estudantes das Regiões Sul e Sudeste foram 44,20 e 44,18 pontos, enquanto a média dos estudantes do Norte e do Nordeste foram 39,98 e 42,89 pontos.

No que tange as variáveis que compõem o *background* familiar, os dados da amostra revelam que os estudantes de renda familiar mínima e baixa, correspondendo a renda familiar até 3 salários mínimos, representavam aproximadamente 52% da totalidade dos estudantes. Houve um aumento das notas médias com a elevação do nível de renda familiar. A nota média dos estudantes com renda familiar máxima foi 10,75 pontos superior à média dos alunos de famílias com renda mínima.

Com relação as escolaridades do pai e da mãe, a amostra revelou que os estudantes cujo pai é analfabeto ou completou apenas o ensino fundamental representou 53,23% dos indivíduos da amostra. Por sua vez, os estudantes cuja mãe é analfabeta ou



completou apenas o ensino fundamental representavam 46.63% dos indivíduos que prestaram o exame. A nota média dos estudantes cujos pais têm ensino superior foram 10,26 pontos maior do que os estudantes com pais analfabetos.

#### Quadro 1. Descrição das variáveis

Variáveis	Descrição
Nota	Nota Final - Média ponderada da formação geral (25%) e componente específico (75%). A nota tem valores entre 0 e 100.
Escola Pública	Se o estudante estudou o ensino médio em escola pública = 1, e zero, caso contrário.
Branco	Se o estudante se auto declarou branco = 1 e zero, caso contrário.
Homem	Se o estudante for o sexo masculino = 1 e zero se for mulher.
Pesmoram	Indica o número de pessoas que moram na mesma residência que o estudante.
Trabalho	Se o estudante declarou que trabalha = 1 e zero, caso contrário.
Renda familiar	A renda familiar é distribuída em: 1) Renda familiar mínima - Até 1,5 salário mínimo 2) Renda familiar baixa – de 1,5 a 3 salários mínimos 3) Renda familiar média – de 3 a 6 salários mínimos 4) Renda familiar alta – de 6 a 9 salários mínimos 5) Renda familiar máxima - acima de 9 salários mínimos
Anos estudos pai	i) Pai analfabeto com zero ano de estudo ii) Pai com ensino fundamental I com 5 anos de estudos, iii) Pai com ensino fundamental II com 9 anos de estudos iv) Pai com ensino médio com 12 anos de estudos v) Pai com ensino superior com 17 anos de estudos vi) Pai com pós-graduação com 20 anos de estudos
Anos estudos mãe	i) Mãe analfabeta com zero ano de estudo ii) Mãe com ensino fundamental I com 5 anos de estudos iii) Mãe com ensino fundamental II com 9 anos de estudos iv) Mãe com ensino médio com 12 anos de estudos v) Mãe com ensino superior com 17 anos de estudos vi) Mãe com pós-graduação com 20 anos de estudos

Fonte: Elaboração a partir de dados do ENADE.

Os dados mostram que a maioria dos estudantes que fizeram o ENADE em 2017 eram egressos do ensino médio das escolas públicas, mas que no ensino superior estavam matriculados nas universidades particulares. Em sua maioria eram de famílias de renda baixa e mínima com renda que tiveram desempenho médio no ENADE inferior os estudantes de estrato de renda mais elevados. Por sexo, eram em sua maioria mulheres e por raça eram em sua maioria brancos.

A análise descritiva simples apresentada acima apontou uma falha latente do modelo educacional brasileiro. A maior parcela dos estudantes que concluíram o ensino médio eram alunos egressos do sistema público, entretanto a maioria dos alunos que fizeram o ENADE estavam matriculados no ensino superior privado. Como os alunos que frequentam a escola privada recebem melhor escolaridade que os alunos da escola pública, os primeiros podem estar ocupando as vagas das instituições públicas de ensino superior. Isso pode estar ocorrendo mesmo no contexto de políticas públicas de ação afirmativa que muitas universidades públicas adotaram ao longo dos últimos anos.

**Tabela 1. Participação e nota média por variáveis de background familiar e de controle**

<i>Variáveis do Background familiar</i>					
<b>Renda familiar e a nota média de desempenho</b>					
	Mínima	Baixa	Média	Alta	Máxima
Participação	22,58%	29,05%	31,02%	10,55%	6,81%
Nota média	39,96	42,22	45,52	47,38	50,71
<b>Escolaridade da mãe e a nota média de desempenho</b>					
	Analfabeta	Fundamental	Médio	Superior	
Participação	5,54%	41,09%	30,72%	22,64%	
Nota média	38,57	42,80	44,29	47,66	
<b>Escolaridade do pai e a nota média de desempenho</b>					
	Analfabeto	Fundamental	Médio	Superior	
Participação	7,92%	45,31%	29,01%	17,76%	
Nota média	39,47	43,39	44,51	49,73	
<i>Variáveis de Controle</i>					
<b>Alunos conforme sexo, cor e a nota média de desempenho</b>					
	Masculino	Feminino	Branco	Não Brancos	
Participação	44,48%	55,52%	53,38%	46,62%	
Nota média	44,00	45,15	44,96	41,88	
<b>Tipo de escola, tipo de universidade e a nota média de desempenho</b>					
	Escola Pública	Escola Privada	Univ. Pública	Universidade Privada	
Participação	73,02%	26,98%	35,27%	64,73%	
Nota média	42,08	47,44	46,70	41,8	
<b>Regiões dos alunos e a nota média de desempenho ( )</b>					
	Norte	Sul	Centro-Oeste	Sudeste	Nordeste
Participação	6,71%	21,73%	7,52%	46,35%	17,69%
Nota média	39,98	44,20	42,24	44,14	42,98

Fonte: Elaboração a partir de dados do ENADE.

## 4.2 O modelo econométrico empírico

O modelo econométrico de Klein e Vella (2009) adaptado para nosso estudo toma a seguinte estrutura:

$$D_i^f = \beta_0 + \beta_1 X_i + \beta_m E_i^m + \beta_p E_i^p + \rho^m \frac{S_u(X_i)}{S_{vm}(X_i)} v_i^m + \rho^p \frac{S_u(X_i)}{S_{vp}(X_i)} v_i^p + \varepsilon_i$$

Em que as variáveis são definidas como:

(a) Nota ( $D_i^f$ ): é o desempenho educacional na prova do ENADE (expressos em logaritmo);

(b) Escolapub ( $X_1$ ): variável dummy que identifica o tipo de escola, 1 se o estudante cursou ensino médio em escola pública, zero caso contrário;

(c) Branco ( $X_2$ ): variável dummy que identifica como o estudante se autodeclarou, 1 se for branco e zero caso contrário;

(d) Homem ( $X_3$ ): variável dummy que identifica o sexo do estudante, 1 se for do sexo masculino e zero se for do sexo feminino;

(e) Trabalho ( $X_4$ ): variável dummy que identifica o status ocupacional do estudante,

1 se o estudante trabalho e zero caso contrário;

(f) Pesmoram ( $X_j$ ): identifica quantidade de pessoas que moram na mesma residência com o estudante;

(g) RendaFam ( $X_g$ ): a renda mensal familiar;

(h) Anosestudospai ( $E^p$ ): anos de estudos do pai;

(i) Anosestudosmãe ( $E^m$ ): anos de estudos da mãe;

(j)  $\rho^p$ : identifica efeito indireto relacionada a educação do pai;

(k)  $\rho^m$ : identifica efeito indireto relacionada a educação da mãe;

(l)  $\varepsilon_i$ : termo de erro da regressão.

### 4.3 Análise dos resultados

Os resultados da estimação não paramétrica de Klein e Vella (2010) e o método do Mínimo Quadrados Ordinários, MQO, são apresentados na Tabela 2. Observamos que há um efeito negativo estatisticamente significativo (ao nível de 5%) entre o desempenho educacional e as seguintes variáveis independentes: o aluno ter estudado em escola pública, ser do sexo masculino, estar trabalhando, e maior número de componentes da família; há um efeito positivo estatisticamente significativo com as variáveis ser branco, possuir renda alta e ser filho de pai com elevado grau de instrução. Por outro lado, a estimação pelo método de Klein e Vella (2010) não evidencia os efeitos da educação da mãe no desempenho individual, como acontece com o método de MQO. Não foram também encontradas evidências de que as variáveis têm influência indireta da educação dos pais sobre o desempenho educacional.

A análise econométrica evidencia que os estudantes oriundos de famílias, cujos pais tem elevada escolaridade e maior renda obtiveram os melhores desempenhos no exame. Observamos que os indivíduos egressos do ensino médio privado e com menor número de moradores no domicílio, indicadores, em boa medida, da renda familiar, tiveram um desempenho melhor. Esses resultados revelam a importância do efeito do background familiar na nota obtida pelos alunos. Em particular, os indivíduos do sexo feminino, que se autodeclararam branco, de renda familiar alta e cujo pai tem escolaridade superior tiveram um desempenho superior aos estudantes que não possuem essas características. O resultado está em linha com a literatura sobre os efeitos do background familiar sobre o aprendizado dos indivíduos.

Os indivíduos de cor branca tiveram uma nota média superior aos não brancos, resultado similar os estudos desenvolvidos por Fernandes (2001), Silva e Hasenbalg (2001). Esses estudos apontam que indivíduos pretos e pardos são submetidos a um ciclo de desvantagens cumulativas em termos de mobilidade intergeracional, produzindo efeitos em todas as fases dos processos educacionais e na estrutura das ocupações do mercado de trabalho.

Por sua vez, os estudantes do sexo feminino obtiveram desempenho superior aos estudantes do sexo masculino, revelado pelo sinal negativo da variável dummy "homem". Estudos mostram que, nos últimos anos, as mulheres têm maior número de matrículas no ensino superior. Portanto, além de predominarem no número de pessoas matriculadas no ensino superior, as mulheres apresentaram desempenho melhor do que os homens. Em que medida a maior educação e empenho das mulheres se refletirá na redução das desigualdades no mercado de trabalho é uma questão em aberto.

Com relação à variável que indica o status ocupacional do estudante, as estimativas mostraram que os estudantes que trabalhavam tiveram desempenho inferior aos que

não trabalhavam. A literatura sobre desempenho educacional tem apontado que os estudantes que não trabalham, dedicam mais tempo às atividades acadêmicas, sobretudo, às horas direcionadas aos estudos excluindo o tempo em sala de aula, assim, alcançam melhores resultados. O desempenho educacional é também afetado negativamente pela maior quantidade de pessoas que moram com o estudante no mesmo domicílio, registrando um melhor aproveitamento no exame para os indivíduos que residem com menor quantidade de pessoas na mesma casa. Esse resultado é consistente com Sandra et al. (2015).

**Tabela 2. Resultados das estimações Klein e Vella (2010) e MQO**

MODELO	MQO	KV	KV (bootstrap)
Nota	$\beta_{MQO}$	$\beta_{KV}$	$\beta_{KV}$
Escola Pública	-0,035206*** (0,001)	-0,0532799*** (0,004)	-0,0532799*** (0,003)
Branco	0,0285473*** (0,001)	0,0324956*** (0,001)	0,0324956*** (0,001)
Homem	-0,0168336*** (0,001)	-0,0113714*** (0,001)	-0,0113714*** (0,001)
Trabalho	-0,0527112*** (0,001)	-0,0582784*** (0,001)	-0,0582784*** (0,001)
Pesmoram	-0,0153115*** (0,000)	-0,0160862*** (0,000)	-0,0160862*** (0,000)
Renda Familiar	0,031443*** (0,000)	0,0329753*** (0,000)	0,0329753*** (0,000)
Anos Estudos Pai	0,0035942*** (0,000)	0,0041776*** (0,001)	0,0041776*** (0,001)
Anos Estudos Mãe	0,0031179*** 0,000	-0,002 0,001	-0,002 0,001
$\rho^p$		0,016 (0,003)	0,016 (0,003)
$\rho^m$		-0,003 (0,002)	-0,003 (0,002)
Constante	3,66817*** (0,003)	3,7282*** (0,011)	3,7282*** (0,010)
TOTAL: 430,008 Observações			

Nota: A variável Nota é expressa em logaritmo. KV refere-se ao estimador de Klein e Vella (2009, 2010). O método KV robusto é obtido usando o procedimento bootstrap. Os valores entre parênteses representam os erros padrão dos coeficientes estimados. \*\*\*, \*\* e \* representam níveis de significância 1%, 5% e 10%.

Em relação às variáveis: renda familiar e anos de escolaridades dos pais, que caracterizam o *background* familiar do estudante, pôde-se afirmar que as famílias com renda elevada afetam positivamente o desempenho do estudante no exame. A literatura traz evidências dos efeitos da renda familiar na decisão do investimento na formação da capital humano. Famílias dotadas de maior poder aquisitivo escolhem maior investimento em educação, gastam com mensalidades em escolas de melhor qualidade; o que possibilita aquisição de livros acesso a tecnologias que favorecem o aprendizado, com maior gasto na formação de habilidades e domínio de outras línguas. Entretanto, o efeito da escolaridade da mãe se mostrou estatisticamente significativo no desempenho somente na estimativa de MQO, enquanto foi não significava nas estimativas que controlaram a endogeneidade.

Diferente da literatura voltada ao Ensino Fundamental e Ensino Médio (Guimarães e Albuquerque 2010; Figueirêdo et al., 2014), na qual o efeito das mães com alta qualificação influenciam fortemente o desempenho escolar dos filhos e, em muitos casos, maior do que a influência dos pais, nesse estudo a educação do pai mostrou-se mais relevante do que a educação da mãe. Uma possível explicação para esse fato é que a influência paterna na escolha profissional seja maior do que a materna.

A adoção da metodologia de Klein e Vella (2010) permitiu analisar a existência de efeitos indiretos dos pais e das mães no desempenho educacional dos indivíduos, o que não ocorre com os métodos tradicionais de estimação por MQO. Os resultados da Tabela 2 mostram que os efeitos indiretos,  $\rho^m$  e  $\rho^p$ , no desempenho educacional foram não significativos estatisticamente. É importante destacar que, mesmo não existindo os efeitos indiretos, não podemos afirmar que os resultados apresentados pelo método MQO são melhores aos do método de Klein e Vella (2010). Embora os resultados sejam parecidos, observamos que a educação da mãe não tem um efeito significativo no desempenho quando é utilizado o método Klein e Vella (2010).

Entre as limitações do trabalho podemos mencionar que não se consideramos um universo maior de variáveis exógenas, optando por empregar as mais utilizadas em outros trabalhos. Certamente, a omissão de variáveis estatisticamente significativas poderia levar a algum viés na estimação dos modelos e na conclusão dos resultados. Estudos futuros poderiam considerar outras variáveis exógenas, como por exemplo, as relacionadas ao apoio financeiro, a Bolsa Família, o FIES, a migração, a composição da família, entre outras.

## 5. Considerações finais

O trabalho teve por objetivo estudar a relação entre o background familiar e o desempenho educacional dos alunos do Ensino Superior que realizaram o ENADE em 2017. A amostra foi composta por 430.008 indivíduos. O procedimento desenvolvido por Klein e Vella (2010) foi utilizado para corrigir o problema de endogeneidade presente neste tipo de trabalho.

Os resultados relevaram que as variáveis do background familiar possuem uma forte influência no desempenho educacional. Em particular, evidenciou-se que a renda familiar e a educação do pai proporcionam efeitos positivos e estatisticamente significantes. Resultado similar foi obtido para os alunos que estudaram em escolas privadas, as quais oferecem melhor qualidade. Contudo, somente famílias de renda maior podem pagar a educação privada nos ensinos fundamental e médio. Outro resultado que também se associa ao background familiar é o número de moradores do domicílio. Famílias de maior renda e educação têm menor número de filhos. Os resultados mostraram que, quanto maior número de moradores do domicílio menor foi a nota dos indivíduos.

Em relação às outras variáveis houve um efeito negativo e estatisticamente significativo entre o desempenho individual e o aluno ser do sexo masculino. Por outro lado, houve um efeito positivo e estatisticamente significativo entre o desempenho e a variável que representa a cor branca. Combinando os resultados econométricos com os dados da Tabela 1, temos que os indivíduos com pior desempenho no ENADE em 2017 foram as pessoas do sexo masculino, não brancos que estudaram em escola pública, filhos de pais analfabetos e com familiar até 1,5 salários mínimos.

Embora os resultados apresentem evidências de background familiar devemos ter presentes os limites do trabalho. Há a possibilidade de alguma variável omitida ter impacto nos resultados, como informações sobre migração, bolsa família, etc. Outro ponto é que usamos apenas dados referentes ao ano 2017, o que poderia ser estendido

para um período maior de tempo.

Os resultados abrem caminhos para novos trabalhos e a consideração de outras questões. Dentre elas, podemos pontuar as políticas públicas direcionadas aos estudantes com baixo *background* para avaliar como as intervenções estatais poderia para melhorar a igualdade de oportunidades. A redução dos efeitos negativos do *background* familiar sobre o desempenho acadêmico é questão fundamental para melhorar a qualidade de educação no Brasil.

### Referências

BARROS, R.; MENDONÇA, R. Os Determinantes da Desigualdade no Brasil. Texto para discussão n. 377, Rio de Janeiro, IPEA, 1995.

BARROS, R.; MENDONÇA, R.; SANTOS, D.; QUINTAES, G. Determinantes do desempenho educacional no Brasil. *Revista Pesquisa e Planejamento Econômico*, v.31, n.1, p.1-42, 2001.

COLEMAN, J.; CAMPBELL, E.; HOBSON, C.; MCPARTLAND, J.; MOOD, A.; WEINFELD, F. D.; YORK, R. *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office, 1966.

CURI, A.; MENEZES-FILHO, N. Mensalidade escolar, Background familiar e os Resultados do Exame Nacional do Ensino Médio (enem). *Pesquisa e Planejamento Econômico*, v. 43, p. 223, 2013.

FARRÉ, L.; KLEIN, R.; VELLA, F. A parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: an application to the NLSY. *Empirical Economics*, p 44:111–133, 2013.

FERNANDES, C. Desigualdade de rendimentos e educação no Brasil: alguns indicadores de diferenças regionais. *Econômica*, v. 3, n. 2, p. 231–250, 2001.

FIGUEIRÊDO, E.; NOGUEIRA, L.; SANTANA, F. Igualdade de Oportunidades: Analisando o papel das circunstâncias no desempenho do ENEM. *Revista Brasileira de Economia*, vol .68, n.3, pp.373-392, 2014.

GUIMARÃES, D.; ALBUQUERQUE, R. Atributos individuais, background familiar e as chances de sucesso dos candidatos ao vestibular da UFC. *Contextus-Revista Contemporânea de Economia e Gestão*, v. 8, n. 2, p-81, 2010.

HANUSHEK, E. The production of education, teacher quality and efficiency. In U.S. Office of Education, *Do Teachers Make a Difference?* pages 79–99. Government Printing Office, Washington, DC, 1970.

HANUSHEK, E. Measuring investment in education. *Journal of Economics Perspectives*, v.10, n.4, p. 9-30, 1996.

HANUSHEK, E.; WOESSMANN, L.; PETERSON, P. *Schools and the equal opportunity*

problem. editors, *Schools and the Equal Opportunity Problem*, pages 159–190. MIT Press, Cambridge, 2007.

KLEIN, R.; VELLA, F. Estimating the return to endogenous schooling decisions via conditional second moments. *Journal of Human Resources*, v. 44, p. 1047–1065, 2009.

KLEIN, R./ VELLA, F. Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, v. 154, p. 154–164, 2010.

KRUEGER, A. Experimental estimates of education production functions. *The Quarterly Journal of Economics*, v.114, n.2, 1999.

LEE, J.; BARRO, R. *Schooling Quality in a Cross Section of Countries*, NBER Working Papers 6198, National Bureau of Economic Research, 1997.

MENDES, B.; KARRUZ, A. Background familiar, desigualdade regional e o desempenho no exame nacional do ensino médio (ENEM). I Encontro Nacional de Ensino e Pesquisa do Campo de Públicas, UnB – ESAF (Brasília), 2015.

MENEZES-FILHO, N. *Os determinantes do desempenho escolar do Brasil*. São Paulo: Instituto Futuro Brasil, 2007.

RIANI, J.L.R., RIOS-NETO, E.,L.G. Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros? *Revista Brasileira de Estudos de População*, v. 25, n. 2, p. 251-269, jul./dez. 2008.

RIOS-NETO, E.; CESAR, C.; RIANI, J. Estratificação educacional e progressão escolar por série no Brasil. *Pesquisa e Planejamento Econômico*, v. 32, n.3, 2002.

ROCHA, F.; TEIXEIRA, J.; MELO, F. Análise dos Fatores que Influenciam o Desempenho Escolar dos Alunos do Ensino Fundamental no Estado do Rio Grande do Norte. *Interface - Revista do Centro de Ciências Sociais Aplicadas*, v. 12, n. 1, p. 89-108, 2015.

SANDRA E.; BLACK, P.; DEVEREUX, K.; SALVANES, G. The More the Merrier? The Effect of Family Size and Birth Order on Children's Education, *The Quarterly Journal of Economics*, Volume 120, Issue 2, Pages 669–700, 2015.

SANTOS, J. Ações afirmativas e educação superior no Brasil: um balanço crítico da produção. *Revista Brasileira de Estudos Pedagógicos*, v.93, n.234, p.401-422, 2010.

SILVA, N.; HASENBALG, C. Recursos familiares e transições educacionais. *Cadernos Saúde Pública*, v. 18 (Supl.), Rio de Janeiro, 2002.

SOUZA, W.; OLIVEIRA, C.; ANNEGUES, A. Background Familiar e Desempenho Escolar: uma abordagem Não Paramétrica. *Pesquisa e Planejamento Econômico*, v. 48. n. 2, s/p. Ago. 2018.

VIEIRA, B.; NOGUEIRA L.; SANTOS W. Os determinantes do índice de desempenho acadêmico dos estudantes do semiárido potiguar. HOLOS, v. 01, s/p., 2017.