

TÍTULO: EFETIVIDADE NO ENSINO SUPERIOR BRASILEIRO: APLICAÇÃO DE MODELOS MULTINÍVEL À ANÁLISE DOS RESULTADOS DO EXAME NACIONAL DE CURSOS

AUTORA: MARIA DOLORES MONTOYA DIAZ

Prof.^a Dr.^a da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto/USP.

E-MAIL para correspondência: madmdiaz@usp.br

ENDEREÇO COMERCIAL: Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Universidade de São Paulo
Av. dos Bandeirantes, 3900
14049-900 – Ribeirão Preto – SP
TEL: (0xx16) 6023910 ou (0xx16) 6024477

RESUMO: Este estudo tem por objetivo analisar o impacto de características institucionais sobre o desempenho dos alunos no Exame Nacional de Cursos (Provão). Uma amostra de mais de 92500 alunos das áreas de Administração, Direito e Engenharia Civil que realizaram os exames no ano de 2000 foi analisada. Foram estimados modelos multinível pois estes dados apresentam uma nítida estrutura hierárquica. Relativamente aos aspectos individuais, verificou-se a existência de um padrão não-linear na relação entre a condição econômica dos alunos e o seu desempenho. A contribuição de níveis superiores de renda sobre o desempenho dos alunos possui um limite, a partir do qual, observam-se, inclusive, impactos negativos. Em relação aos aspectos institucionais, encontraram-se efeitos positivos provenientes da maior qualificação e das melhores condições de trabalho para o corpo docente e da utilização de atividades de pesquisa como estratégia de ensino/aprendizagem. A participação em atividades de extensão, além de afetar positivamente o desempenho dos alunos, gera um efeito atenuador sobre o impacto negativo resultante da condição socioeconômica desfavorável do aluno. Verificou-se que se o aluno tinha uma renda familiar até R\$ 454,00, seu desempenho, em média, tenderia a ser 3,6 pontos inferior ao de alunos com renda familiar superior a R\$ 7.550,00. Porém, se ele participar de atividades de extensão essa diferença se reduzirá para aproximadamente 2 pontos e, se os todos os docentes mostrarem empenho e dedicação, a diferença se situará na faixa de 1 ponto.

ABSTRACT: This study analyses the impact of institutional characteristics on the students' performance in the National Examination of Higher Education Courses (Provão). A sample of more than 92500 students from Management, Law and Civil Engineering who had made the tests in the year 2000 was analyzed. Multilevel models have been fitted because these data present an evident hierarchical structure. Regarding individual aspects, one noted a nonlinear relation between economic condition and students' performance. The contribution of higher levels of income on the performance had a limit, from which, the impacts were negative. As far as the institutional aspects are concerned, one found a positive impact on pupils' performance of a teaching staff with higher percentage of PhDs and Masters, where there were better work conditions for the faculty and where research activities were used as a teaching/learning strategy. The students' participation in additional activities called "de extensão" had positive effects on their performance. Besides, this kind of activities generated an attenuating effect on the negative impact of the less favorable socioeconomic condition of the student. If the student had a familiar income less than R\$454,00, its performance, in average, tend to be 3,6 points inferior to the one with familiar income higher than of R\$ 7.550,00. However, if he participated in additional activities this difference will be reduced to approximately 2 points. If, additionally, the teaching staff shows more dedication and hard work, this difference will be reduced to 1 point.

Palavras-chave: Ensino Superior, Modelos Multinível, Provão, Efetividade Institucional

Keywords: Higher Education, Multilevel Models, Institutional Effectiveness

Área ANPEC: ECONOMIA SOCIAL E DEMOGRAFIA ECONÔMICA

Classificação JEL: I23, C21

1. Introdução

Os debates acerca das tendências mundiais em relação ao trabalho e emprego convergem em um aspecto: a inserção no mercado de trabalho será cada vez mais difícil em decorrência da progressiva competitividade nos mercados globalizados e das constantes e, não raras vezes, abruptas transformações. Esta realidade tem reflexos diretos sobre a situação e as políticas na área educacional, especialmente no que tange ao ensino superior.

No caso brasileiro, nos últimos anos ocorreram profundas mudanças no sistema de ensino superior. Para se ter uma noção da magnitude desse processo, basta verificar que as matrículas no ano de 2001 atingiram a faixa dos 3 milhões, o que representa um crescimento de 93,7% em relação aos números de 1991. Relativamente ao número de concluintes, foram aproximadamente 400 mil em 2001, correspondendo a uma variação de 67,5% na década.

Não se pode deixar de mencionar, ainda, que das 1391 instituições de ensino superior existentes em 2001, aproximadamente 87% eram privadas, tendo no período 1991-2001, crescido à notável taxa de 80,0%. As instituições privadas passaram a responder por aproximadamente 67% dos concluintes e 69% das matrículas em 2001.

Algumas questões que vêm permeando as discussões acerca do ensino superior ficam ainda mais evidentes durante esse processo de crescimento acelerado. Entre elas, podem ser citadas:

- O nível socioeconômico do aluno é determinante de seu desempenho, ou seja, independentemente da instituição freqüentada, alunos mais ricos tendem a apresentar um melhor desempenho que os mais pobres?

- As características institucionais influenciam o processo de ensino e aprendizagem dos alunos de graduação? Em caso afirmativo, quais seriam elas?

- Empenho e assiduidade dos professores exercem alguma influência sobre o desempenho dos alunos?

- E qual o efetivo impacto da melhor qualificação e do regime de trabalho do corpo docente?

- Em que medida a participação em atividades de pesquisa e extensão por parte dos alunos contribui ao processo de ensino e aprendizagem no ensino superior?

Questões dessa natureza são identificadas com as linhas de pesquisa relativas à estimação da função de produção educacional e das análises da efetividade escolar. Porém, é justamente em relação ao ensino superior que se verifica, nessa área, uma enorme carência de trabalhos, mesmo em nível internacional.

O tema da efetividade dos recursos educacionais aplicados no ensino básico e médio encontra-se bastante desenvolvido tanto entre pesquisadores da área de economia como entre os educadores. Assim, principalmente em relação à educação básica, encontram-se muitas evidências, mesmo que contraditórias, sobre a importância dos recursos alocados para o setor educativo¹. Porém, relativamente ao ensino superior, as análises, mesmo em nível internacional, são escassas e bastante recentes. No Brasil, as informações disponíveis limitam-se a algumas poucas estatísticas básicas e avaliações bastante simples disponíveis no site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira(INEP).

¹ O tema, pela importância e pelas controvérsias existentes, foi objeto de grande atenção por parte de um conjunto de importantes economistas no *Symposium on School Quality and Educational Outcomes*, realizado na Kennedy School of Government da Harvard University em dezembro de 1994, cujas principais contribuições encontram-se reunidas no número de novembro de 1996 da prestigiosa *The Review of Economics and Statistics*. Para reforçar o argumento acerca da importância, cada vez maior, que o tema vem despertando entre os economistas, deve-se destacar que o número de fevereiro de 2003 do *Economic Journal* também foi dedicado à análise da relação entre recursos aplicados na área da educação e o desempenho escolar dos alunos.

Sendo assim, o objetivo deste trabalho é procurar contribuir ao preenchimento da lacuna existente nesse debate. Para tanto, pretende-se avaliar os aspectos individuais e, principalmente, institucionais que influenciaram o desempenho dos alunos no Provão realizado no ano de 2000, pelos formandos das áreas de Administração, Direito e Engenharia Civil. Essas três áreas foram selecionadas justamente por terem sido avaliadas desde o início do processo, ou seja, em 1996. Desse modo, pode-se considerar que em 2000 o processo já estava bastante consolidado. Com isso, pretende-se evitar ou, ao menos, reduzir a possibilidade de intervenção de fatores sobre o desempenho e/ou resposta dos alunos ou das instituições aos questionários, como decorrência do desconhecimento do processo de avaliação.

Na próxima seção é apresentada a metodologia dos modelos multinível, que vem sendo adotada na literatura recente sobre o tema da efetividade dos recursos educacionais. Na seção 3 encontram-se os resultados obtidos e a seção 4 contém as considerações finais acerca dos resultados e conclusões a que se chegou ao longo do trabalho.

2. Metodologia e Variáveis Utilizadas

2.1 Aspectos Teóricos

Em anos recentes, em muitas áreas das ciências sociais identificou-se a necessidade de incorporar nas análises os efeitos das conexões existentes entre os indivíduos e as características do ambiente onde os mesmos vivem e desenvolvem suas atividades. Como lembraram Rice e Jones(1997:561), uma grande atenção, principalmente na área de economia, tem sido dada à relação existente entre os indivíduos e as características das instituições às quais eles pertencem, *“most notably in assessing ‘effectiveness’ or ‘performance’ of public institutions”*.

Na área da saúde, por exemplo, esse tipo de análise apresentou um desenvolvimento notável, principalmente, nas pesquisas acerca dos efeitos de fatores ambientais sobre a susceptibilidade dos indivíduos a determinadas enfermidades. Na verdade, é evidente a necessidade de desenvolvimento, aprofundamento e, principalmente, de utilização de metodologias nas análises realizadas nas ciências sociais, que permitam uma identificação mais precisa da magnitude dos impactos resultantes de aspectos específicos do ambiente, seja ele físico ou institucional, ou mais genericamente ainda, do contexto em que o indivíduo está inserido em contraposição a suas características singulares.

Os modelos tradicionalmente utilizados, mais especificamente o Modelo Linear Geral estimado pelo Método dos Mínimos Quadrados Ordinários (MQO) apresentam limitações quando utilizados na análise de dados que apresentam uma estrutura hierárquica, como aqueles encontrados nas análises baseadas em pesquisas domiciliares, cuja estrutura é indivíduo/domicílio/ estado/região, ou mesmo, na análise de dados educacionais, com alunos agrupados em classes dentro de escolas.

Em termos genéricos, a existência de correlação entre os indivíduos pertencentes a um mesmo grupo, por exemplo, entre alunos de uma mesma escola, pacientes de um mesmo médico, etc. gera um viés negativo nas estimativas dos erros-padrão dos parâmetros do modelo estimado pelo Método de Mínimos Quadrados Ordinários conforme provado por Goldstein (1995).

Porém, o que se pretende a partir da metodologia associada à construção e estimação dos Modelos Multinível, não se resume simplesmente à obtenção de erros padrão “corretos” ou “corrigidos”, produzidos facilmente, por exemplo, por Mínimos Quadrados Generalizados, como ocorre nos modelos convencionais, quando se identifica a presença de autocorrelação ou heterocedasticidade do termo aleatório. A metodologia desenvolvida pretende e permite investigar simultaneamente efeitos individuais e efeitos contextuais, bem como, acomodar a correlação intra-grupo, e modelar estruturas complexas de variância.

É interessante notar que, justamente, em decorrência da formulação distinta da convencionalmente adotada, da ênfase dada ao processo de especificação e modelagem e à compreensão dos fatores determinantes do padrão de variância, freqüentemente, verifica-se que a análise realizada acabou por evidenciar impactos extremamente relevantes, principalmente, relacionados a aspectos contextuais, que normalmente acabariam ocultos.

2.1.1 Conceituação

Inicialmente, consideremos a relação entre uma variável explicativa, X_i , onde i representa cada indivíduo e uma variável dependente, Y_i . A variável X_i pode representar, por exemplo, a situação socioeconômica de cada aluno i enquanto a variável Y_i seria o resultado obtido em uma determinada avaliação, como, por exemplo, o Provão em um conjunto de J instituições. Formalmente, seguindo a notação utilizada por Snijders e Bosker(1999), um modelo com dois níveis, alunos (nível 1) agrupados em instituições de ensino (nível 2) representado por um intercepto aleatório² pode ser definido da seguinte maneira:

$$\text{Equação 1} \quad Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + R_{ij}$$

onde,

$$\text{Equação 2} \quad \begin{cases} \beta_{0j} = \gamma_{00} + U_{0j} \\ \beta_{1j} = \gamma_{10} \end{cases}$$

A substituição da Equação 2 na Equação 1 produz o seguinte modelo:

$$\text{Equação 3} \quad Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + U_{0j} + R_{ij}$$

onde U_{0j} e R_{ij} são variáveis aleatórias, mutuamente independentes com distribuição normal e média zero. Além disso, deve-se destacar que σ^2 é variância ao nível dos indivíduos ou dentro dos grupos e τ_0^2 é a variância ao nível do grupo, ou seja, entre grupos.

Neste caso, pode-se também obter a seguinte estatística:

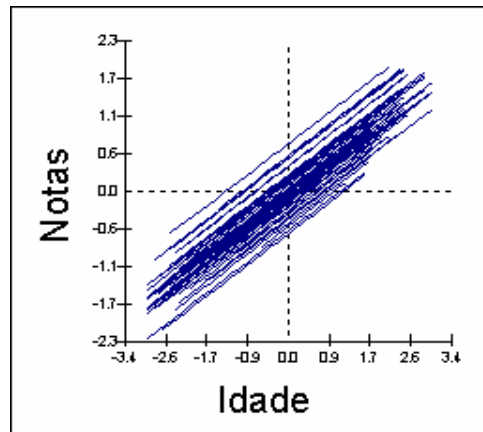
$$\text{Equação 4} \quad \rho(Y_{ij}, Y_{i'j}) = \frac{\tau_0^2}{\tau_0^2 + \sigma^2},$$

que corresponde à correlação intra-classe, conforme mostrado por Snijders and Bosker(1999:46). Esta é a correlação entre os valores da variável explicada(Y) referentes a dois indivíduos, aleatoriamente retirados de um determinado grupo, também aleatoriamente selecionado entre os diversos grupos existentes. Conforme Goldstein (1995:19) esta correlação também pode ser interpretada como a parcela da variância total devida aos grupos que definem o nível 2.

Em termos gráficos, este modelo corresponde a um modelo de intercepto aleatório e declividade fixa, representando cada uma das retas um determinado grupo. Na figura abaixo se pode verificar um exemplo simulado. Os resultados apresentados referem-se a um modelo multinível mais elementar, ou seja, aquele em que o nível superior é apenas representado pelo intercepto aleatório. No exemplo, o modelo relaciona as notas de alunos com a variável idade centralizada, sendo que cada uma das retas representa uma instituição.

² Random intercept.

Figura 1: Representação Gráfica de um Modelo com Intercepto Aleatório



Essa estrutura, entretanto, pode ser generalizada para incorporar declividades aleatórias³, assim, como mais variáveis explicativas associadas a cada um dos diversos níveis. A seguir, é apresentada a generalização para o caso de p variáveis x , explicativas de nível 1 e q variáveis z , explicativas de nível 2. Além disso, incorporam-se intercepto e parâmetros de declividades (β_{hj}) aleatórios para todas as p variáveis x , sendo explicados pelas q variáveis z . Tem-se, então:

Equação 5
$$Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \dots + \beta_{pj}x_{pij} + R_{ij}$$

onde,

Equação 6
$$\beta_{hj} = \gamma_{h0} + \gamma_{h1}z_{1j} + \dots + \gamma_{hq}z_{qj} + U_{hj}$$

A substituição da Equação 6 na Equação 5, produz o seguinte modelo:

Equação 7
$$Y_{ij} = \gamma_{00} + \sum_{h=1}^p \gamma_{h0}x_{hij} + \sum_{k=1}^q \gamma_{0k}z_{kj} + \sum_{k=1}^q \sum_{h=1}^p \gamma_{hk}z_{kj}x_{hij} + U_{0j} + \sum_{h=1}^p U_{hj}x_{hij} + R_{ij}$$

Neste caso, também é importante lembrar dos parâmetros da matriz de variância-covariância do modelo. São eles:

- nível 1 dos indivíduos : σ^2 é variância entre indivíduos
- nível 2 do grupo:
 - $\text{var}(U_{hj}) = \tau_h^2$ ($h=1, \dots, p$);
 - $\text{cov}(U_{hj}, U_{kj}) = \tau_{hk}$ ($h, k=1, \dots, p$).

Esta será a especificação utilizada neste trabalho, porém, deve-se notar que a estrutura de modelagem permitiria generalizar ainda mais o modelo pela incorporação de níveis superiores, tornando-o aplicável, por exemplo, em um estudo em que se dispõem de informações que caracterizam alunos, as turmas ou o professor, e a instituição.

³ *Random slopes*

É essencial esclarecer que no caso do Exame Nacional de Cursos (*Provão*), verifica-se a presença de uma nítida estrutura hierárquica em que os alunos estão vinculados a universidades, ou seja, o nível 1 é representado pelos alunos e o nível 2, representado pelas instituições.

Antes de prosseguir, não se pode deixar de comentar a importância que as variáveis de contexto desempenham nesta estrutura de modelagem. O objetivo de introduzir variáveis de contexto, que entram no modelo normalmente como explicativas de nível 2, é explicitar os impactos que o ambiente familiar ou institucional tem sobre as ações ou resultados individuais. Estas variáveis podem, por exemplo, procurar refletir características institucionais como o percentual de mestres e doutores no corpo docente.

Pode ocorrer, também, que haja interesse específico sobre um aspecto contextual para o qual não existe uma variável disponível. Neste caso, pode-se construí-la, normalmente, pelo cálculo da média por instituição da respectiva variável existente ao nível dos indivíduos.

Para compreender melhor a função deste tipo de variáveis contextuais, pode-se considerar, como exemplo, a variável relativa à escolaridade paterna, que é reconhecida como importante determinante do padrão socioeconômico da família. Porém, é preciso lembrar que o impacto individual pode ser distinto do contextual. Assim, por exemplo, um determinado pai que tenha cursado o ensino superior pode dedicar-se muito ao trabalho, não tendo tempo para acompanhar o desempenho escolar dos filhos. Neste caso, individualmente, a variável escolaridade paterna pode gerar um impacto negativo, a princípio, inesperado. Obviamente, entretanto, isto não pode ser utilizado para afirmar que a maior escolaridade dos pais prejudica o rendimento dos filhos.

O que ocorre, normalmente, é que pais com escolaridade superior, tendem a criar um ambiente sociocultural mais dinâmico para o desenvolvimento dos filhos, com acesso a meios que acabam colaborando para maior progresso intelectual. Esta característica comum também resulta em uma preocupação em relação às instituições educacionais que o filho vai frequentar. Por esta via, espera-se que em instituições educacionais em que uma maior parcela de pais tenha escolaridade superior, os filhos tendam a apresentar um melhor desempenho.

Justamente, para testar efeitos dessa natureza, são introduzidas as variáveis contextuais.

Igualmente, ao verificado nos modelos lineares convencionais, a construção e realização de testes de hipóteses também constituem etapas fundamentais nesse tipo de análise.

Os testes aplicados podem se referir a hipóteses sobre os parâmetros fixos, ou seja, os γ e sobre os parâmetros de variância e covariância, ou seja, σ^2 , τ_h^2 e τ_{hk} . Também, são fundamentais os testes aplicados no processo de seleção de modelos, tais como o “*deviance test*” e os critérios de informação Akaike e Schwarz.

A hipótese nula dos testes relativos aos parâmetros fixos γ é a convencional, $H_0: \gamma_{hk} = 0$. A estatística do teste é a seguinte:

$$z = \frac{\hat{\gamma}_{hk}}{\text{erro-padrão}_{\hat{\gamma}_{hk}}}, \text{ onde } \hat{\gamma}_{hk} \text{ é a estimativa de máxima verossimilhança de } \gamma_{hk} \text{ e o}$$

erro-padrão _{$\hat{\gamma}_{hk}$} é o erro-padrão amostral estimado de $\hat{\gamma}_{hk}$. Segundo Raudenbush e Bryk (2002:58), formalmente, esta estatística tem distribuição assintoticamente normal.

Para testar a significância estatística das variâncias e covariâncias estimadas podem ser aplicados testes tipo “Wald”, cuja estrutura permite, inclusive, que, simultaneamente, várias hipóteses sejam testadas, desde que se disponha de uma amostra bastante grande. Assim, é possível considerar uma situação em que um vetor α contendo p parâmetros, por exemplo, a variância de $U_{1j} = \tau_1^2$, variância de $U_{2j} = \tau_2^2$ e a $\text{cov}(U_{1j}, U_{2j}) = \tau_{12}$, para o qual se deseja testar a hipótese de que todos sejam nulos.

Nesse caso, a hipótese nula desse teste será $H_0: \alpha = \mathbf{0}$, e a estatística do teste será: $\hat{\alpha}' \hat{\Sigma}_r^{-1} \hat{\alpha}$ onde $\hat{\Sigma}_r^{-1}$ é matriz de var/cov dos parâmetros testados. Sob a hipótese nula, esta estatística tem distribuição χ^2

com p graus de liberdade. Obviamente, é possível realizar o teste considerando uma restrição de cada vez, cuja estatística terá distribuição χ^2 com 1 grau de liberdade.

O “*deviance test*”, como um teste do tipo razão de verossimilhança⁴, também enquadra-se perfeitamente aos objetivos pretendidos, ao permitir que se comparem distintas especificações. A lógica do teste é simples: para testar a hipótese nula H_0 , modelo com m_0 parâmetros contra a hipótese alternativa H_1 , modelo com m_1 parâmetros, sendo $m_1 > m_0$, basta mensurar a seguinte estatística:

$$(D_0 - D_1) = -2 \log\text{likelihood}(\text{modelo da } H_0) - [-2 \log\text{likelihood}(\text{modelo da } H_1)]$$

Essa estatística tem uma distribuição χ^2 com $(m_1 - m_0)$ graus de liberdade. Assim, para $\text{Prob}(X^2 \leq \chi^2)$ ⁵ próximos de zero, pode-se rejeitar a hipótese nula.

Deve-se notar, finalmente, que os critérios de seleção de modelos Akaike(AIC) e Schwarz também são extremamente úteis no processo de avaliação e seleção dos modelos, tendo em vista que podem ser calculados a partir da estatística gerada pela estimação da função de verossimilhança ($-2 \log\text{likelihood}$). É interessante notar que, essencialmente, a diferença entre os critérios reside na magnitude da penalização imposta à introdução de variáveis explicativas no modelo. Sendo assim, o critério Schwarz penaliza mais severamente a perda de graus de liberdade, levando, normalmente à superioridade de modelos mais simples⁶.

Finalmente, os métodos de estimação de modelos multinível merecem um comentário adicional. Para estimar os parâmetros de um modelo – o conjunto de parâmetros fixos, γ_{hk} , e as variâncias e covariâncias, σ^2 , τ_h^2 e τ_{hk} – considerando a hipótese de normalidade dos termos U_{ij} e R_{ij} , existem algumas alternativas, como o método de máxima verossimilhança, de máxima verossimilhança restrita e o método Bayesiano.

Como bem destacam Raudenbush e Bryk(2002:437), escolhido o método de estimação, passa-se à etapa seguinte de escolha do algoritmo computacional que permitirá a implementação do referido método. Assim, por exemplo, se a opção for pelo método de máxima verossimilhança será necessário maximizar a função de verossimilhança, o que pode implicar tarefa bastante difícil, tendo em vista a complexidade da expressão envolvida. É justamente o que ocorre com os modelos multinível em que se utiliza o Método de Mínimos Quadrados Generalizados Iterativo.

Essencialmente, o método inicia-se pela estimativa dos parâmetros fixos, γ_{hk} , pelo Método de Mínimos Quadrados Ordinários. A partir desses resultados é possível obter estimativas dos resíduos que permitem que se estime a matriz de variâncias e covariâncias do modelo (σ^2, τ_h^2 e τ_{hk}). Isso possibilita uma re-estimação dos parâmetros fixos, pela aplicação do Método de Mínimos Quadrados Generalizados, que, por sua vez, serão utilizados para obter uma re-estimativa da matriz de variâncias e covariâncias. Estas etapas irão se sucedendo até que se verifique uma convergência dos resultados obtidos. Justamente por esta razão, o método é denominado iterativo⁷.

2.2 Fonte de Dados e Definição das Variáveis

Os microdados relativos aos graduandos das Carreiras de Administração, Direito e Engenharia Civil que realizaram a avaliação no ano de 2000 foram fornecidos pelo INEP em dois bancos de dados – um referente ao cadastro dos cursos e o outro, dos alunos. As informações relacionadas ao aluno são originárias do questionário preenchido na ocasião de realização das provas e os dados cadastrais das instituições são provenientes dos questionários preenchidos pelos coordenadores dos cursos. Também foi fornecida a nota geral que cada um dos alunos obteve no exame. A base contém informações acerca de 92.594 indivíduos. No quadro abaixo encontram-se descritas as variáveis utilizadas na análise.

⁴ Likelihood ratio test.

⁵ Marginal significance level ou *p-values*.

⁶ Detalhes acerca dos critérios podem ser encontrados em Kennedy(1998:103).

⁷ Detalhes técnicos podem ser encontrados em Raudenbush e Bryk(2002) e Goldstein(1995).

Quadro 1: Descrição das Variáveis dos Modelos

<i>Variável</i>	<i>Descrição</i>	<i>Nível</i>
nota-ger (explicada)	Nota do aluno obtida na prova geral (escala de 0 a 100)	--
cons	Constante – vetor de 1	--
renda_1	Variável Binária (1= 1ª faixa de renda familiar – até R\$ 453,00 = 3 salários mínimos, 0=caso contrário)	Indivíduo
renda_2	Variável Binária (1= 2ª faixa de renda familiar – entre R\$ 454,00 a R\$ 1.510,00 = 3 a 10 sal. Mínimos, 0=caso contrário)	Indivíduo
renda_3	Variável Binária (1= 3ª faixa de renda familiar – entre R\$ 1.511,00 a R\$ 3.020,00 = 10 a 20 sal. Mínimos, 0=caso contrário)	Indivíduo
renda_4	Variável Binária (1= 4ª faixa de renda familiar 4 – entre R\$ 3.021,00 a R\$ 7.550,00 = 20 a 50 sal. Mínimos, 0=caso contrário)	Indivíduo
renda_5	Variável Binária (1= 5ª faixa de renda familiar – mais de R\$ 7.550,00 = mais de 50 sal. Mínimos, 0=caso contrário)	Indivíduo
dum_sex_corr	Variável Binária (0=sexo masculino, 1=sexo feminino)	Indivíduo
idade_corr	Idade Centralizada (Idade- Média Idade da Instituição onde o aluno estudou)	Indivíduo
dumpai	Variável Binária (1=pai com formação superior, 0=caso contrário)	Indivíduo
dumpai_context	Variável contextual – percentual de pais com educação superior	Instituição
dummae	Variável Binária (1=mãe com formação superior, 0=caso contrário)	Indivíduo
dumspriv	Variável Binária (1= estudou em escola privada no segundo grau, 0=caso contrário)	Indivíduo
dumicro	Variável Binária (1=possui e utiliza micro em sua residência, 0=caso contrário)	Indivíduo
direito	Variável Binária (1=carreira direito, 0=caso contrário)	Instituição
engenharia	Variável Binária (1=carreira engenharia civil, 0=caso contrário)	Instituição
dumext	Variável Binária referente à participação em atividades de extensão (1= sim, 0=caso contrário)	Indivíduo
dumpesq	Variável Binária referente à realização de iniciação científica ou tecnológica, monitoria, projetos de pesquisa (1= sim, pelo menos uma das atividades, 0=caso contrário)	Indivíduo
dumequil	Variável Binária (1=há equilíbrio entre quantidade de conteúdo e carga horária das disciplinas do seu curso, 0=caso contrário)	Indivíduo
dumpart	Variável Binária (0= predominância de aulas expositivas, 1= a maioria dos docentes utilizou técnicas de ensino que envolvem mais práticas e participação)	Indivíduo

<i>Variável</i>	<i>Descrição</i>	<i>Nível</i>
dumpes52	Variável Binária referente à realização de atividades de pesquisas como estratégia de aprendizagem (1=sim, na maior parte ou em todas as disciplinas, 0=caso contrário)	Indivíduo
dumpes52_context	Variável contextual – percentual de alunos que mencionaram ter desenvolvido atividades de pesquisa na maior parte das disciplinas	Instituição
dumempen	Variável Binária (1=todos ou a maior parte dos professores demonstrou empenho, assiduidade e pontualidade, 0=caso contrário)	Indivíduo
dumempme	Variável Binária (1=metade dos professores demonstrou empenho, assiduidade e pontualidade, 0=caso contrário)	Indivíduo
dumexima	Variável Binária (1= curso deveria ter exigido mais de mim, 0=caso contrário)	Indivíduo
dumexime	Variável Binária (1=curso deveria ter exigido menos de mim, 0=caso contrário)	Indivíduo
norte	Variável Binária (1= região Norte, 0=caso contrário)	Instituição
nordeste	Variável Binária (1= região Nordeste, 0=caso contrário)	Instituição
sul	Variável Binária (1= região Sul, 0=caso contrário)	Instituição
centro-oeste	Variável Binária (1= região Centro-Oeste, 0=caso contrário)	Instituição
sudeste	Variável Binária (1= região Sudeste, 0=caso contrário)	Instituição
ndout	Percentual de doutores	Instituição
nmest	Percentual de mestres	Instituição
dout_mest	Percentual de doutores ou mestres	Instituição
jmais40	Percentual de docentes com jornada de 40 horas semanais	Instituição
exmenos8	Percentual de docentes com exercício em sala de aula inferior a 8 horas semanais	Instituição
ndocen	Número de docentes informado pela instituição de ensino	Instituição
ndocen_corr	ndocen (Centralizada)	Instituição
privada	Variável Binária (1= instituição privada, 0=caso contrário)	Instituição

As demais variáveis representam efeitos cruzados e resultam da multiplicação entre as variáveis originais.

É interessante notar que, no quadro acima, há uma terceira coluna que procura indicar as variáveis pelo nível hierárquico. É importante destacar que a apresentação de algumas variáveis representativas de características de docentes não implica a associação das mesmas com um nível hierárquico adicional. Isso poderia ocorrer, mas não é esse o caso, tendo em vista que as referidas variáveis refletem a visão dos alunos. Assim, um mesmo professor pode ter sido avaliado de maneira distinta por seus diversos alunos. A existência do nível intermediário entre os alunos e as instituições dependeria da existência de informações que caracterizariam turmas, ou mesmo os docentes, como por exemplo, escolaridade ou sexo do docente.

3. Resultados

3.1 Análise Preliminar

Encontram-se na Tabela 1 algumas informações acerca das variáveis contínuas presentes em algum dos modelos selecionados, cujos resultados embasarão a análise apresentada na próxima seção. São elas: o número de observações com informações disponíveis(N), média e desvio-padrão.

Tabela 1: Estatísticas Básicas de Variáveis Contínuas

	<i>N</i>	<i>Média</i>	<i>Desvio_padrão</i>
ndocen (instituições)	774	42,4	31,1
ndout (instituições)	774	10,6	14,4
nmest (instituições)	774	30,4	16,4
jmais40(instituições)	774	29,0	29,3
Exmenos8(instituições)	774	40,4	28,6
nota_ger (alunos)	92.594	36,54	13,83
idade (alunos)	92.584	27,31	6,57

É interessante notar a maior dispersão das instituições quanto à manutenção de doutores em seus quadros vinculados aos cursos de direito, administração e engenharia civil relativamente ao padrão da contratação de mestres. Assim, enquanto o coeficiente de variação(desvio-padrão/média) é de 1,36 no caso da variável percentual de doutores, a estatística correspondente no caso da variável percentual de mestres é de 0,54.

Na tabela abaixo se verifica que, entre as carreiras analisadas, a maior parcela de formandos é da área de Direito, responsável por aproximadamente 49% dos alunos, seguida de perto pela, de Administração, com pouco mais de 44% dos alunos. Bem distante, encontra-se a área de Engenharia Civil, com pouco menos de 6,5% de participação.

Tabela 2: Número de Alunos por Carreira

<i>Carreira</i>	<i>Alunos</i>	<i>%</i>
1- Administração	41.050	44,3
2 – Direito	45.552	49,2
3 – Engenharia Civil	5.992	6,5
TOTAL	92.594	100

Na Tabela 3 verifica-se que a grande maioria dos alunos formou-se em instituições situadas na Região Sudeste, responsável pela enorme parcela de 63,6% dos formandos.

Tabela 3: Número de Alunos por Região

<i>Região</i>	<i>Alunos</i>	<i>%</i>
1-Norte	2820	3,0
2-Nordeste	8863	9,6
3-Sudeste	58853	63,6
4-Sul	15487	16,7
5-Centro-Oeste	6571	7,1
TOTAL	92594	100

Na Tabela 4 é possível notar que a maior parte, ou seja, 62,2% dos alunos formandos em alguma das três carreiras analisadas, estudou em universidades. Merece destaque também o fato de que aproximadamente 21% dos alunos estudaram em estabelecimentos isolados.

Tabela 4: Número de Alunos por Natureza da Instituição

<i>Natureza</i>	<i>Alunos</i>	<i>%</i>
1-Universidade	57.608	62,2
2-Centro Universitário	8.804	9,5
3-Faculdades Integradas	6.556	7,1
4-Estabelecimento Isolado	19.626	21,2
TOTAL	92.594	100

É interessante verificar na Tabela 5 que as instituições particulares já são responsáveis por 76% dos formandos das carreiras analisadas⁸.

Tabela 5: Número de Alunos por Dependência

<i>Dependência</i>	<i>Alunos</i>	<i>%</i>
1-Federal	8.763	9,5
2-Estadual	6.335	6,8
3-Municipal	7.166	7,7
4-Particular	70.330	76,0
TOTAL	92.594	100

Finalmente, nota-se na Tabela 6 que os indivíduos do sexo masculino ainda são maioria entre os formandos, a exceção da carreira de Direito. Merece destaque, entretanto, a reduzida participação (26,4%) de mulheres na carreira de Engenharia Civil.

Tabela 6: Número de Alunos por Carreira e Sexo

<i>Carreira</i>		<i>Masculino</i>	<i>Feminino</i>	<i>TOTAL</i>
1-Administração	N	21.500	19.550	41.050
% do Sexo na Carreira	%	52,4	47,6	100,0
2-Direito	N	21.725	23.750	45.475
% do Sexo na Carreira	%	47,8	52,2	100,0
3-Engenharia	N	4.410	1.582	5.992
% do Sexo na Carreira	%	73,6	26,4	100,0
TOTAL		47.635	44.882	92.517
% do Sexo no Total de Formandos	%	51,5	48,5	100,0

3.2 Os Modelos

Inicialmente, deve-se destacar que todas as estimações foram produzidas pelo software MLwinN versão 1.10.0007 de Rasbash et al(2001). É importante lembrar, ainda, que este é um modelo cuja estrutura hierárquica possui dois níveis: alunos(nível 1) agrupados dentro das instituições(nível 2). Os resultados dos modelos estimados encontram-se na Tabela 7.

Antes de passar aos comentários dos resultados é importante notar que a terminologia utilizada na literatura referente aos modelos hierárquicos ou multinível difere daquela adotada na literatura econométrica relativa aos modelos de dados em painel. Assim, os efeitos fixos referem-se aos coeficientes das variáveis explicativas, independentemente de se referirem ao nível 1, 2 ou superiores, caso existam. Os efeitos aleatórios referem-se às estimativas das variâncias e covariâncias. Assim, será adotada aqui a aceção utilizada na literatura dos modelos multinível.

⁸ Dos formandos em geral, ou seja, considerando todas as carreiras, as instituições particulares respondiam por aproximadamente 67% em 2001.

O primeiro modelo estimado, dentro da lógica dos modelos hierárquicos, é um modelo incondicional, comumente denominado Modelo “Vazio”⁹ justamente por não incorporar qualquer variável explicativa. Na verdade, os resultados produzidos por este modelo correspondem àqueles originados por uma análise de decomposição da variância com efeitos aleatórios. Nesse caso, a variância encontra-se decomposta entre os dois níveis considerados e estes resultados básicos serão tomados como referência para a análise do modelo final.

O modelo “vazio” tem a seguinte especificação:

$$\text{Equação 8: } Y_{ij} = \beta_{0j} + R_{ij}$$

$$\text{Equação 9: } \beta_{0j} = \gamma_{00} + U_{0j}$$

Verifica-se que o coeficiente β_0 deve ser interpretado como a média esperada para um aluno selecionado aleatoriamente em uma instituição também selecionada aleatoriamente. De qualquer modo, a estimativa obtida foi muito baixa situando-se próxima de 35,3^{10,11}.

Nos modelos mais complexos o parâmetro β_0 deverá ser interpretado como a média esperada para um estudante cujas características são representadas por todas as explicativas contidas no modelo iguais a zero. Por esta razão, é importante a centralização de algumas das referidas explicativas, como idade, por exemplo. Com esta centralização, o aluno representativo terá a idade média do grupo em que está inserido, ou seja, a instituição em que se formou, e não zero, o que tornaria a interpretação consideravelmente mais complexa. Para as variáveis binárias, a situação representada pelo valor zero normalmente já tem um significado preciso.

Prosseguindo a análise dos resultados do Modelo “Vazio”, verificou-se que a estimativa da variância entre alunos, ou seja, variância no nível 1 (σ^2) foi de 142,4 enquanto a variância entre instituições, ou a variância no nível 2 (τ_0^2) foi de 57,76. Estes resultados permitem que se calcule a correlação intra-classe, conforme definição apresentada na Equação 4. O valor obtido foi

$$\rho(Y_{ij}, Y_{i'j}) = \frac{\tau_0^2}{\tau_0^2 + \sigma^2} = \frac{57,76}{57,76 + 142,4} = 28,86\%, \text{ o que indica que a maior parte da variância deve-se à}$$

variação intra-instituição e não à variação entre-instituições. Este resultado coincide com aqueles encontrados na literatura, principalmente relacionada a análises da educação básica e/ou secundária. De acordo com Snijders e Bosker(1999:46), normalmente os valores referentes a este indicador situam-se no intervalo 0,05 e 0,20. É interessante notar, no entanto, que em relação ao ensino superior, tanto em Belfield e Fielding(2001) como em Van der Hulst e Jansen(2002) a medida de correlação intra-classe obtida a partir do modelo “vazio” foi consideravelmente inferior, não chegando ao valor de 0,08.

Para confirmar a significância estatística do efeito aleatório incluído realizaram-se dois testes: teste tipo “Wald”.

A estatística encontrada no teste tipo “Wald” foi de 375,24. Considerando que esta estatística tem distribuição qui-quadrado com 1 grau de liberdade, conclui-se pela rejeição da hipótese nula de que a variância do parâmetro β_0 é zero.

Assim, pode-se concluir que no caso brasileiro existe um considerável grau de semelhança no desempenho de alunos do ensino superior de uma mesma instituição, bastante maior que o verificado até o momento no exterior, evidenciando a relevância de se incorporar adequadamente os fatores institucionais que possam ser responsáveis por esta correlação.

Sendo assim, a seguir, será apresentado o Modelo 1 que é composto apenas pelas variáveis que se mostraram significativas ao longo do processo de especificação.

⁹ “Empty” model

¹⁰ Lembrar que a escala de notas vai de 0 a 100.

¹¹ Detalhes técnicos acerca do método de estimação podem ser encontrados em Snijders e Bosker(1999:56-63) e Raudenbush e Bryk (2002:38-65).

Equação 10:

$$Y_{ij} = \beta_{0j} + \beta_{1j} * renda_1 + \beta_{2j} * renda_2 + \beta_{3j} * renda_4 + \\ \beta_{4j} * dum_sex_corr + \beta_{5j} * idade_corr + \beta_{6j} * dumpai + \\ + \beta_{7j} * dumspriv + \beta_{8j} * dumicro + \beta_{9j} * dumpesq + \\ \beta_{10j} * dumext + \beta_{11j} * dumequil + \beta_{12j} * dumpart + \beta_{13j} * dumpes52 + \\ \beta_{14j} * dumempen + \beta_{15j} * dumempeme + \beta_{16j} * dumexima + \\ \beta_{17j} * dumexime + \beta_{18j} * renda_1.dumext + \beta_{19j} * renda_1.dumempen + \\ \beta_{20j} * renda_2.dumext + \beta_{21j} * renda_2.exmenos8 + R_{ij}$$

Equação 11:

$$\left\{ \begin{array}{l} \beta_{0j} = \gamma_{00} + \gamma_{01} * direito + \gamma_{02} * engenharia + \gamma_{03} * nordeste + \\ \gamma_{04} * sudeste + \gamma_{05} * sul + \gamma_{06} * centroeste + \gamma_{07} * dout_mest + \\ \gamma_{08} * jmais40 + \gamma_{09} * exmenos8 + \gamma_{010} * privada + \\ \gamma_{011} * dumpai_context + \gamma_{012} * dumpes52_context + U_{0j} \\ \beta_{1j} = \gamma_{10}, \beta_{2j} = \gamma_{20}, \beta_{4j} = \gamma_{40}, \beta_{5j} = \gamma_{50}, \beta_{6j} = \gamma_{60}, \\ \beta_{7j} = \gamma_{70}, \beta_{9j} = \gamma_{90}, \beta_{10j} = \gamma_{100}, \beta_{11j} = \gamma_{110}, \beta_{12j} = \gamma_{120}, \\ \beta_{13j} = \gamma_{130}, \beta_{14j} = \gamma_{140}, \beta_{15j} = \gamma_{150}, \beta_{16j} = \gamma_{160}, \beta_{17j} = \gamma_{170}, \\ \beta_{18j} = \gamma_{180}, \beta_{19j} = \gamma_{190}, \beta_{20j} = \gamma_{200}, \beta_{21j} = \gamma_{210} \end{array} \right.$$

Inicialmente, deve-se lembrar que foi encontrado um padrão não-linear no impacto da renda familiar sobre o rendimento dos alunos, pois os alunos das duas primeiras faixas de renda, apresentaram um desempenho esperado inferior àquele do grupo de referência, ou seja, o dos alunos com renda familiar superior a 50 salários mínimos, ou R\$ 7.550,00 em 2000. Os graduandos da terceira faixa não apresentaram diferenças significativas em relação ao grupo de referência, enquanto os formandos com renda familiar entre 20 e 50 salários mínimos obtiveram um resultado superior aos dos alunos mais ricos.

Os coeficientes negativos de **dum_sex_corr** e **idade_corr** indicaram que as mulheres, controlados todos os demais fatores individuais, contextuais, institucionais, etc. tenderam a apresentar piores resultados do que os homens e que o desempenho esperado de alunos mais velhos é inferior ao de indivíduos mais jovens.

Os coeficientes das variáveis **dumpai**, **dumspriv** e **dumicro** devem ser analisados conjuntamente com os resultados obtidos para a variável contextual **dumpai_context**, já que todas elas pretendem captar os efeitos do ambiente socioeconômico e cultural em que o formando está inserido. A conclusão a que se chegou é que o maior impacto sobre o rendimento dos alunos foi captado pela variável **dumpai_context**, cujo resultado indica que os alunos que estão se formando em instituições em que é maior o percentual de pais que possuem ensino superior tendem a apresentar um desempenho consideravelmente superior àqueles que estudaram em escolas em que a maioria dos pais dos alunos não tem formação superior. O coeficiente negativo de **dumpai** indica que um aluno cujo pai não tem ensino superior, mas estuda em uma escola em que é grande o percentual de alunos com pais com formação superior tende a ser mais favorecido por este contexto do que alunos cujos pais têm formação superior. Por outro lado, alunos cujos pais tenham escolaridade superior mas que estudem em escolas em que é baixo o percentual de alunos com escolaridade superior, serão mais prejudicados por este contexto desfavorável do que os demais. É

preciso notar, no entanto, que a magnitude deste efeito é bem pequena, não chegando a 0,3 ponto, o mesmo ocorrendo com **dumspri**. Nesse caso, o resultado negativo indica que a frequência à escola de segundo grau privada não é garantia de bom desempenho ao término do curso superior. Assim, os efeitos positivos, caso existam, parecem esgotar-se no ingresso ou mesmo ao longo da formação superior. Finalmente, o acesso a microcomputador na residência produziu efeitos positivos sobre a aprendizagem dos alunos.

Relativamente às práticas didáticas dos docentes, verificou-se que a utilização de atividades de pesquisa como estratégia de ensino/aprendizagem produz resultados positivos no desempenho dos alunos, representados pelo coeficiente positivo da variável **dumpes52**. Este resultado foi ainda mais reforçado pelo coeficiente da variável **dumpes52_context**, que indica que em instituições em que é alto o percentual de alunos que tiveram que desenvolver atividades de pesquisa nas disciplinas freqüentadas, os formandos tendem a apresentar um rendimento superior àqueles que estudaram em escolas em que essa prática era menos freqüente. Assim, um aluno que desenvolveu muitas atividades de pesquisa ao longo de sua formação e que freqüentou instituição em que estas atividades eram comumente desenvolvidas pode chegar a obter um diferencial de mais de 6 pontos em relação ao aluno que não desenvolveu qualquer pesquisa e estudou em instituição em que nenhum docente utiliza esta estratégia pedagógica.

De forma coerente com os resultados comentados acima, o engajamento dos alunos em atividades de extensão (variável **dumext**) ou atividades de pesquisa, como iniciação científica (variável **dumpesq**), por exemplo, também produz efeitos positivos sobre a aprendizagem dos alunos. É interessante destacar, inclusive, que a participação em atividades de extensão gera ainda um efeito atenuador sobre o impacto negativo resultante da condição socioeconômica desfavorável do aluno. Assim, no caso da variável **renda_1.dumext** verifica-se que se o aluno tinha uma renda familiar até R\$ 454,00 seu desempenho, em média, tenderia a ser 3,559 pontos inferior ao de alunos com renda familiar superior a R\$ 7.550,00. Porém, se ele participar de atividades de extensão, essa diferença se reduzirá para aproximadamente 2 pontos e se os todos os docentes mostrarem empenho e dedicação a diferença se situará na faixa de 1 ponto. Ademais, deve-se lembrar que a demonstração de empenho, dedicação e pontualidade dos docentes (variáveis **dumempme** e **dumempen**) produz impactos positivos comparativamente à situação em que nenhum ou poucos docentes mostraram-se empenhados em suas atividades docentes.

É interessante destacar, entretanto, que os parâmetros das variáveis **dumequil** e **dumpart** foram negativos. Relativamente a **dumequil**, esse resultado indicaria que cursos com maior equilíbrio entre quantidade de conteúdo e carga horária das disciplinas gerariam um impacto negativo sobre o desempenho dos alunos. Verificou-se, inicialmente, que o percentual de alunos que consideraram o curso equilibrado foi relativamente baixo, atingindo 34%. Assim, além de poucos alunos considerarem que o curso freqüentado era equilibrado, ainda estes apresentaram um desempenho pior do que aqueles que consideraram o curso desequilibrado. Uma explicação possível para este resultado decorreria do fato desta variável ter sido construída a partir das respostas dos alunos. Assim, poderia ter capturado o efeito de respostas “prudentes” dos alunos mais desinteressados e com pior desempenho. Outra explicação poderia ser a de que os alunos com melhor desempenho tendem a ser mais exigentes, ou mais realistas, considerando que os cursos que freqüentaram eram desequilibrados. Por outro lado, alunos com pior desempenho, ou por pouca atenção, consciência ou preocupação com a estrutura do curso freqüentado, tenderam a considerá-lo mais equilibrado. De qualquer modo, os resultados sinalizam que esta variável pode não ter captado, efetivamente, o efeito pretendido.

No caso de **dumpart** verificou-se que mais de 70% dos alunos mencionaram que a maioria dos seus docentes utilizou, predominantemente, alguma técnica de ensino mais prática ou participativa, como por exemplo, trabalhos de grupo desenvolvidos em sala de aula, e não apenas aulas expositivas. O resultado negativo para o coeficiente obtido indica que estas atividades podem não estar sendo conduzidas de maneira didaticamente adequada.

Merece destaque também o sinal negativo da variável **dumexime**. Este resultado indica que os alunos que consideraram que o curso deveria ter sido menos exigente, apresentaram um desempenho pior comparativamente aos alunos que mencionaram que o nível de exigência foi adequado. Isto parece indicar

que os alunos com maiores dificuldades ao longo do curso não conseguem superá-las ao término do mesmo, apresentando um pior desempenho nas avaliações.

A variável **dumexima**, de maneira análoga a **dumexime**, tem por objetivo captar o efeito da insatisfação dos alunos em relação ao nível de exigência do curso. No caso de **dumexima**, procura-se mensurar o impacto da insatisfação com o baixo grau de exigência do curso. Teoricamente, espera-se que esta variável tenha sinal positivo, na medida em que se considera que normalmente os melhores alunos tendem a demonstrar este tipo de insatisfação. Foi exatamente este o resultado encontrado.

É importante destacar que as variáveis caracterizadoras das instituições, ou seja, **dout_mest**, **jmais40**, **exmenos8**, **ndocen_corr**, **privada** e as binárias relativas às regiões de localização também são consideradas como contextuais, na medida em que refletem aspectos comuns a cada instituição e que podem, portanto, ser responsáveis por um certo grau de homogeneidade no desempenho dos alunos. Apesar de pequena magnitude, porém, é interessante notar que os coeficientes das variáveis **dout_mest**, **jmais40** e **exmenos8** se mantiveram positivos em todos os modelos estimados, indicando que esses aspectos têm impactos efetivos sobre o rendimento dos alunos. A magnitude desses coeficientes também sinaliza que, quando se controlam os demais fatores, isoladamente o impacto dessas características é relativamente pequeno, o que era esperado na medida em que a maior qualificação do quadro docente e as melhores condições de trabalho representadas por **jmais40** e **exmenos8** permitem a aplicação de práticas acadêmicas e didáticas mais adequadas e estas, sim, seriam as responsáveis pelo melhor rendimento dos alunos.

Em relação às regiões verificou-se que, em média e com todos os demais fatores individuais e institucionais controlados, os formandos das regiões nordeste, sudeste, sul e centro-oeste tiveram desempenho superior ao obtido pelos alunos da região norte.

Finalmente, merece destaque o resultado relativo ao desempenho de alunos de instituições privadas, que, mesmo controlados todos os demais aspectos, tende a ser pior do que o de alunos de instituições públicas.

É interessante notar ainda que a variância entre as escolas (nível 2) reduziu-se consideravelmente no Modelo 1 (21,92) comparativamente ao Modelo “Vazio”(57,76). Assim, controlados aspectos individuais e institucionais, a correlação existente entre dois alunos de mesma instituição, ou seja, a correlação intra-classe reduziu-se para 14,25%. No Modelo “Vazio” a correlação intra-classe era 28,86%.

Outra estatística auxiliar produzida foi a variância explicada no Nível 2, no Modelo 1, que atingiu a faixa de 62,05%. Este indicador reflete a redução na proporção da variância existente no nível 2, comparativamente ao Modelo “Vazio”, ou seja, as variáveis incluídas no modelo explicam 62,05% da variância existente entre as instituições.

É importante destacar que ambos os critérios de seleção de modelos – Akaike e Schwarz – indicam que o Modelo 1 é superior a todas as demais especificações testadas. Os resultados encontram-se destacados na parte inferior da Tabela 7 na respectiva coluna.

Finalmente, o Modelo 2 contém o mesmo conjunto de variáveis incluídas no Modelo 1, mas foi estimado pelo Método de Mínimos Quadrados convencional. Assim, antes das comparações acerca dos resultados produzidos convém apresentar a especificação do Modelo 2:

Equação 12:

$$\begin{aligned}
Y_{ij} = & \beta_{0j} + \beta_{1j} * renda_1 + \beta_{2j} * renda_2 + \beta_{3j} * renda_4 + \\
& \beta_{4j} * dum_sex_corr + \beta_{5j} * idade_corr + \beta_{6j} * dumpai + \\
& \beta_{7j} * dumspriv + \beta_{8j} * dumicro + \beta_{9j} * direito + \beta_{10j} * engenharia \\
& \beta_{11j} * dumpesq + \beta_{12j} * dumext + \beta_{13j} * dumequil + \beta_{14j} * dumpart + \beta_{15j} * dumpes52 + \\
& \beta_{16j} * dumempen + \beta_{17j} * dumempeme + \beta_{18j} * dumexima + \beta_{19j} * dumexime + \\
& \beta_{20j} * nordeste + \beta_{21j} * sudeste + \beta_{22j} * sul + \beta_{23j} * centroeste + \\
& \beta_{24j} * dout_mest + \beta_{25j} * jmais40 + \beta_{26j} * exmenos8 + \beta_{27j} * privada + \\
& \beta_{28j} * renda_1.dumext + \beta_{29j} * renda_1.dumempen + \beta_{30j} * renda_2.dumext + \\
& \beta_{31j} * renda_2.exmenos8 + \beta_{32j} * dumpai_context + \beta_{33j} * dumpes52_context + R_{ij}
\end{aligned}$$

Equação 13:

$$\left\{ \begin{aligned}
& \beta_{0j} = \gamma_{00}, \beta_{1j} = \gamma_{10}, \beta_{2j} = \gamma_{20}, \beta_{3j} = \gamma_{30}, \beta_{4j} = \gamma_{40}, \beta_{5j} = \gamma_{50}, \\
& \beta_{6j} = \gamma_{60}, \beta_{7j} = \gamma_{70}, \beta_{8j} = \gamma_{80}, \beta_{9j} = \gamma_{90}, \beta_{10j} = \gamma_{100}, \beta_{11j} = \gamma_{110}, \\
& \beta_{12j} = \gamma_{120}, \beta_{13j} = \gamma_{130}, \beta_{14j} = \gamma_{140}, \beta_{15j} = \gamma_{150}, \beta_{16j} = \gamma_{160}, \\
& \beta_{17j} = \gamma_{170}, \beta_{18j} = \gamma_{180}, \beta_{19j} = \gamma_{190}, \beta_{20j} = \gamma_{200}, \beta_{21j} = \gamma_{210}, \\
& \beta_{22j} = \gamma_{220}, \beta_{23j} = \gamma_{230}, \beta_{24j} = \gamma_{240}, \beta_{25j} = \gamma_{250}, \beta_{26j} = \gamma_{260}, \\
& \beta_{27j} = \gamma_{270}, \beta_{28j} = \gamma_{280}, \beta_{29j} = \gamma_{290}, \beta_{30j} = \gamma_{300}, \beta_{31j} = \gamma_{310}, \\
& \beta_{32j} = \gamma_{320}, \beta_{33j} = \gamma_{330}
\end{aligned} \right.$$

Genericamente, pôde-se observar que os sinais dos coeficientes estimados são coincidentes em todas as variáveis. Relativamente à magnitude dos coeficientes obtidos a partir dos Modelos 1 e 2, verificou-se, também, que a maior parte (mais de 60%) delas apresenta diferenças entre os coeficientes inferiores a 1 erro-padrão.

Porém, é fundamental destacar que os critérios de seleção de modelos – Akaike e Schwarz – indicam que o Modelo 2 é inferior a todas as demais especificações testadas.

Tabela 7: Resultados das Estimações dos Modelos com Distintas Especificações

<i>Variáveis</i>	<i>Modelo Vazio</i>	<i>Modelo 1</i>	<i>Modelo 2</i>
EFEITOS FIXOS	<i>Coefficientes</i>	<i>Coefficientes</i>	<i>Coefficientes</i>
cons	35.2600 (*)	25.0600 (*)	27.8700 (*)
renda_1		-3.5590 (*)	-3.5920 (*)
renda_2		-0.9128 (*)	-0.8055 (*)
renda_3		---	---
renda_4		0.8529 (*)	0.8248 (*)
dum_sex_corr		-2.7080 (*)	-2.7330 (*)
idade_corr		-0.1106 (*)	-0.1176 (*)
dumpai		-0.2945 (*)	-0.2752 (*)
dummae		---	---
dumspriv		-0.5948 (*)	-0.6872 (*)
dumicro		1.7660 (*)	1.8440 (*)
direito		3.0370 (*)	2.3210 (*)
engenharia		-13.4300 (*)	-13.8000 (*)
dumpesq		0.5191 (*)	0.5406 (*)
dumext		0.9043 (*)	1.1130 (*)
dumequil		-0.4308 (*)	-0.5132 (*)

dumpart		-1.7360 (*)	-2.3270 (*)
dumpes52		1.3430 (*)	1.2600 (*)
dumempen		1.8050 (*)	2.2820 (*)
dumempme		1.3870 (*)	1.6010 (*)
dumexima		1.7510 (*)	1.1370 (*)
dumexime		-1.9780 (*)	-1.1540 (*)
nordeste		1.6120 (**)	0.9640 (*)
sudeste		3.6000 (*)	2.5090 (*)
sul		4.1590 (*)	2.4990 (*)
centroeste		2.7930 (*)	3.1290 (*)
dout_mest		0.0385 (*)	0.0486 (*)
jmais40		0.0442 (*)	0.0094 (*)
exmenos8		0.0242 (*)	0.0150 (*)
ndocen_corr		---	---
Privada		-4.1720 (*)	-5.8830 (*)
renda_1.dumpesq		---	---
renda_1.dumext		1.6630 (*)	1.6640 (*)
renda_1.dumpart		---	---
renda_1.dumpes52		---	---
renda_1.dumempen		0.9066 (*)	1.3730 (*)
renda_1.dout_mest		---	---
renda_1.jmais40		---	---
renda_1.exmenos8		---	---
renda_2.dumpesq		---	---
renda_2.dumext		0.5782 (*)	0.5983 (*)
renda_2.dumpart		---	---
renda_2.dumpes52		---	---
renda_2.dumempen		---	---
renda_2.dout_mest		---	---
renda_2.jmais40		---	---
renda_2.exmenos8		-0.0065 (*)	-0.0063 (**)
dumpai_context		16.9000 (*)	17.9100 (*)
dumspriv_context		---	---
dumicro_context		---	---
dumpes52_context		4.7810 (*)	4.5560 (*)
Efeitos Aleatórios			
<i>Nível 2</i>			
variância $U_{0j}=\tau_0^2$	57.7600	21.9200	---
variância $U_{1j}=\tau_1^2$		---	---
variância $U_{2j}=\tau_2^2$		---	---
covar(U_{0j},U_{1j})= τ_{01}		---	---
covar(U_{0j},U_{2j})= τ_{02}		---	---
covar(U_{1j},U_{2j})= τ_{12}		---	---
<i>Nível 1</i>			
variância R_{ij}	142.4000	131.9000	150.0000
-2 loglikelihood	724,719.0	573,362.0	581,084.0
Número de parâmetros	3	36	35
Teste em Relação ao		Modelo 2	---
m ₁ -m ₀		1	---
D ₀ -D ₁		7,722.0	---

Prob($X^2 \leq \chi^2$) – Deviance Test		0.00000	---
Estatísticas Auxiliares:			
Variância explicada no Nível 2		62.05%	---
Variância explicada no Nível 1		7.37%	---
obs	92,594	74,035	74,035
AIC	7.8269	7.7454	7.8497
Schwarz	7.8272	7.7499	7.8541
correlação intra-classe	28.86%	14.25%	---

(*) estatisticamente significativo ao nível de 5% (**) estatisticamente significativa ao nível de 10%.

4. Considerações Finais

O principal objetivo deste trabalho foi analisar o impacto de características individuais e institucionais sobre o desempenho dos alunos no Exame Nacional de Cursos (Provão), por meio da estimação de modelos multinível.

Relativamente aos aspectos individuais, verificou-se que a condição econômica dos alunos é muito importante como determinante do rendimento dos alunos. O fato que merece destaque, entretanto, é que os resultados para as carreiras analisadas evidenciaram a existência de um padrão não-linear, na medida em que os alunos das duas primeiras faixas de renda apresentaram um desempenho esperado inferior àquele do grupo de referência, ou seja, o dos alunos mais ricos, cuja renda familiar era superior a 50 salários mínimos ou R\$ 7.550,00 em 2000. Já os graduandos da terceira faixa não apresentaram diferenças significativas em relação ao grupo de referência, enquanto os formandos com renda familiar entre 20 e 50 salários mínimos obtiveram um resultado superior aos dos alunos mais ricos. Esses resultados mostraram que os formandos pertencentes à classe média ou média alta parecem compor uma elite acadêmica no país. Seria interessante investigar qual a trajetória desse grupo, relativamente à continuidade dos seus estudos em nível de pós-graduação, por exemplo. De qualquer modo, verificou-se que a contribuição de níveis superiores de renda sobre o desempenho dos alunos possui um limite, a partir do qual, observam-se, inclusive, impactos negativos.

Prosseguindo na análise dos impactos das características socioeconômicas dos indivíduos sobre o rendimento no *Provão*, foi possível verificar que o ambiente em que o indivíduo se formou contribui de maneira significativa para o seu desempenho. Esse efeito foi captado pela variável contextual relativa ao percentual de pais com escolaridade superior existente na instituição em que o aluno está se formando.

Conforme mencionado anteriormente, para compreender melhor a função desse tipo de variáveis contextuais, pode-se considerar, em especial, a variável relativa à escolaridade paterna, que é normalmente reconhecida como importante determinante do padrão socioeconômico da família. Porém, é preciso lembrar que o impacto individual pode ser distinto do contextual. Assim, por exemplo, um determinado pai que tenha cursado o ensino superior pode dedicar-se muito ao trabalho, não tendo tempo para acompanhar o desempenho escolar dos filhos. Nesse caso, individualmente, a variável escolaridade paterna pode, inclusive, apresentar impactos negativos inesperados. Foi justamente o que ocorreu na análise realizada. Obviamente, esse resultado, justamente por ser incompleto, não pode ser utilizado para afirmar que a maior escolaridade dos pais prejudica o rendimento dos filhos.

O que ocorre, normalmente, é que pais com escolaridade superior tendem a criar um ambiente sociocultural mais dinâmico para o desenvolvimento dos filhos, com acesso a meios que acabam colaborando para maior progresso intelectual. A melhor formação dos pais também resulta em uma preocupação em relação às instituições educacionais que o filho vai frequentar. Por essa via, espera-se que em instituições educacionais em que uma maior parcela de pais tenha escolaridade superior, os filhos tendam a apresentar um melhor desempenho.

Essas evidências em relação aos impactos de variáveis socioeconômicas sobre o rendimento dos alunos foram complementadas pela confirmação de que o acesso a microcomputador na residência também contribui positivamente sobre a aprendizagem dos graduandos.

Os resultados indicaram, ainda, que nas carreiras analisadas, indivíduos mais velhos tendem a apresentar um pior desempenho do que os mais jovens e que as mulheres, controlados os demais fatores, tenderam a apresentar piores resultados do que os homens.

Relativamente aos fatores institucionais, foram confirmadas algumas expectativas existentes quanto a alguns aspectos.

Assim, encontraram-se efeitos positivos, apesar de sua pequena magnitude, provenientes da maior qualificação e das melhores condições de trabalho para o corpo docente. Verificou-se que maiores percentuais de mestres e doutores no quadro docente, com jornada de trabalho de 40 horas semanais e com exercício em sala de aula inferior a 8 horas semanais produzem impactos efetivos sobre o rendimento dos alunos. Conforme já destacado anteriormente, a magnitude destes coeficientes também sinaliza que, quando se controlam os demais fatores, isoladamente o efeito dessas variáveis é relativamente pequeno. Obviamente, esse não é um resultado surpreendente, na medida em que se espera que essas características, na verdade, permitam a aplicação de práticas acadêmicas e didáticas mais adequadas e estas, sim, seriam as responsáveis pelo melhor rendimento dos alunos. A investigação dos efeitos destas práticas também corroborou algumas hipóteses existentes.

Relativamente aos docentes, os resultados indicaram que a demonstração de empenho, dedicação e pontualidade por parte da maioria do corpo docente produz impactos positivos sobre o desempenho dos alunos, comparativamente à situação em que nenhum ou poucos docentes mostraram-se empenhados e dedicados às suas atividades educativas.

Em relação às práticas didáticas e acadêmicas, foi interessante comprovar o efeito positivo proveniente da utilização de atividades de pesquisa como estratégia de ensino/aprendizagem. Mais importante, ainda, foi verificar que em instituições em que é alto o percentual de alunos que tiveram que desenvolver atividades de pesquisa nas disciplinas freqüentadas, os formandos tendem a apresentar um rendimento superior àqueles que estudaram em escolas em que essa prática era menos freqüente. Assim, um aluno que desenvolveu muitas atividades de pesquisa ao longo de sua formação e que freqüentou instituição em que estas atividades eram comumente desenvolvidas pode chegar a obter um diferencial de mais de 6 pontos em relação ao aluno que não desenvolveu qualquer pesquisa e estudou em instituição em que nenhum docente utiliza essa estratégia pedagógica.

Outra hipótese comumente levantada também foi comprovada: a de que o engajamento dos alunos em atividades de extensão ou atividades de pesquisa, como iniciação científica, por exemplo, também produz efeitos positivos sobre a aprendizagem dos alunos. É interessante destacar, inclusive, que a participação em atividades de extensão gera ainda um efeito atenuador sobre o impacto negativo resultante da condição socioeconômica desfavorável do aluno. Assim, verificou-se que se o aluno tinha uma renda familiar até R\$ 454,00 seu desempenho, em média, tenderia a ser 3,6 pontos inferior ao de alunos com renda familiar superior a R\$ 7.550,00. Porém, se ele participar de atividades de extensão essa diferença se reduzirá para aproximadamente 2 pontos e, se os todos os docentes mostrarem empenho e dedicação, a diferença se situará na faixa de 1 ponto.

Deve-se notar, ainda, que alguns resultados evidenciaram que é preciso ter um cuidado especial com a elaboração do questionário de avaliação do curso, a ser respondido pelos alunos. Caso contrário, existe o risco das questões não captarem as informações que se pretendia obter. Esse problema pode ter ocorrido com a questão que procurava verificar os efeitos da existência de um maior equilíbrio entre quantidade de conteúdo e carga horária das disciplinas no curso freqüentado. Verificou-se, inicialmente, que o percentual de alunos que consideraram o curso equilibrado foi relativamente baixo, atingindo 34%. Assim, além de poucos alunos considerarem que o curso freqüentado era equilibrado, ainda, estes apresentaram um desempenho inferior àqueles que consideraram o curso desequilibrado. Uma explicação possível poderia estar no fato de que esta questão acabou capturando o efeito de respostas “prudentes”

dos alunos mais desinteressados e com pior desempenho. Outra explicação poderia ser a de que os alunos com melhor desempenho tendem a ser mais exigentes, ou mais realistas, considerando que os cursos que freqüentaram eram desequilibrados. Por outro lado, alunos com pior desempenho, ou por pouca atenção, consciência ou preocupação com a estrutura do curso freqüentado, tenderam a considerá-lo mais equilibrado.

O coeficiente negativo obtido para a variável relativa à adoção, por parte da maioria dos docentes, de alguma técnica de ensino mais prática ou participativa, como, por exemplo, trabalhos de grupo desenvolvidos em sala de aula, em contraposição à utilização apenas de aulas expositivas, parece indicar que essas atividades podem não estar sendo conduzidas de maneira didaticamente adequada.

Verificou-se, ainda, que os alunos que consideraram que o curso deveria ter sido menos exigente apresentaram um desempenho pior comparativamente aos alunos que mencionaram que o nível de exigência foi adequado. Isto parece indicar que os alunos com maiores dificuldades ao longo do curso não conseguem superá-las ao término do mesmo, apresentando um pior desempenho nas avaliações. A identificação precoce desses casos permite que a instituição possa criar mecanismos de apoio a esses alunos, o que pode ser decisivo na superação das dificuldades enfrentadas, contribuindo para um melhor aproveitamento e, conseqüente, formação desse grupo de alunos.

Não foi possível comprovar a existência de qualquer impacto como decorrência do número de docentes da instituição, ou seja, as evidências indicaram não haver relação entre o desempenho dos alunos e o fato de estarem se graduando em instituições menores ou maiores.

Em relação às regiões verificou-se que, em média e com todos os demais fatores individuais e institucionais controlados, os formandos das regiões nordeste, sudeste, sul e centro-oeste tiveram desempenho superior ao obtido pelos alunos da região norte.

Finalmente, merece destaque o resultado relativo ao desempenho de alunos de instituições privadas, que, mesmo controlados todos os demais aspectos, tende a ser pior do que o de alunos de instituições públicas.

Neste ponto, abre-se a perspectiva de criação de uma nova linha de investigação, na medida em que se pode questionar se parte desse resultado negativo não poderia desaparecer caso fosse possível construir um modelo distinto, em que se incorporasse uma medida de desempenho antes do ingresso no nível superior, como por exemplo, o desempenho no Exame Nacional do Ensino Médio (ENEM). Nesse caso, o modelo teria outra natureza, já que pretenderia avaliar algo como o “valor adicionado” ao conhecimento de cada indivíduo decorrente da freqüência a um curso superior. Obviamente, este objetivo foge ao escopo da análise aqui realizada.

Neste trabalho procurou-se explorar as potencialidades e riqueza das informações geradas pelo Sistema Nacional de Avaliação do Ensino Superior e colaborar para a identificação de meios de formar e capacitar, cada vez melhor, os indivíduos. Esta é a questão crucial que inquieta todos os policy-makers da área educacional tanto no nível institucional, como municipal, estadual e nacional, neste momento de intensa expansão do ensino superior. Expansão essa que decorre, principalmente, da crescente consciência da população acerca da importância de maiores níveis educacionais para a inserção no mercado de trabalho, já globalizado e altamente competitivo, e para o desenvolvimento econômico e social de qualquer nação.

5. Bibliografia

- ALBERNAZ A, FERREIRA, FHG, FRANCO C. (2002). “Qualidade e Equidade na Educação Fundamental Brasileira”. *Texto para Discussão*, n. 455. Departamento de Economia PUC-Rio.
- BARBOSA MEF, FERNANDES C. (2001). “A Escola Brasileira faz Diferença? Uma investigação dos efeitos dos efeitos da escola na proficiência em Matemática dos alunos da 4ª série”. *Programa de Seminários Acadêmicos*, n. 09/2001. Instituto de Pesquisas Econômicas/USP.

- BELFIELD CR, FIELDING A. (2001). "Measuring the relationship between resources and outcomes in higher education in the UK". *Economics of Education Review*, vol. 20: 589-602.
- BIRDSALL N, SABOT RH. (1996). (eds) *Opportunity Foregone: Education in Brazil*. Washington, D.C.: Inter-American Development Bank, distributed by The Johns Hopkins University Press.
- CASE A, DEATON A. (1999). "School inputs and educational outcomes in South Africa". *Quarterly Journal of Economics*, vol. 114, n.3 (August): 1047-1084.
- DIAZ, MDM. (1999). "Extended Stay at University: an application of multinomial logit and duration models". *Applied Economics*, vol. 31, n. 11: 1411-1422.
- DUNCAN C, JONES K, MOON G. (1998). "Context, Composition and Heterogeneity: using multilevel models in health research". *Social Science and Medicine*, vol. 46, n.1:97-117.
- DUSTMANN C. (2003). "The Class Size Debate and Educational Mechanisms: Editorial". *Economic Journal*, vol.113 (February), pp. F1-F2.
- GOLDSTEIN H, WOODHOUSE G. (2000). "School Effectiveness Research and Educational Policy". *Oxford Review of Education*, vol. 26, n. 3 e 4: 353-363.
- GOLDSTEIN H, YANG M, OMAR R, TURNER R, THOMPSON S. (2000). "Meta-analysis using multilevel models with an application to the study of class size effects". *Appl. Statist.*, vol. 49, Part.3: 399-412.
- GOLDSTEIN H. (1995). *Multilevel Statistical Models*. 2nd edition. Kendall's Library of Statistics 3. Edward Arnold.
- GREENE WH. (2000). *Econometric Analysis*, 4th edition. New Jersey: Prentice-Hall, Inc.
- HANUSHEK EA, RIVKIN SG, TAYLOR LL. (1996). "Aggregation and the Estimated Effects of School Resources". *The Review of Economics and Statistics*, vol. 78, n.4:611-627.
- HANUSHEK, EA. (2002). "The Failure of Input-Based Schooling Policies". *Economic Journal*, vol.113 (February), pp. F64-F98.
- INEP website (1999) – http://www.inep.gov.br/noticias/news_213.htm
- KENNEDY P. (1998). *A Guide to Econometrics*. 4th edition. Cambridge: The MIT Press.
- KRUEGER, AB. (2003). "Economic Consideration and Class Size". *Economic Journal*, vol.113 (February), pp. F34-F63.
- LADD HF, WALSH RP. (2002). "Implementing value-added measures of school effectiveness: getting the incentives right". *Economics of Education Review*, vol. 21:1-17.
- LONGFORD NT. (1993). *Random Coefficient Models*. Oxford Statistical Science Series n.11. Clarendon Press.
- RASBASH J, BROWNE W, HEALY M, CAMERON B, CHARLTON C. (2001). MLwiN version 1.10.0007. Multilevel Models Project, Institute of Education, Londres.
- RAUDENBUSH SW, BRYK AS. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edition. Sage Publications.
- RICE N, JONES A. (1997). "Multilevel Models and Health Economics". *Health Economics*, vol.6: 561-575.
- SNIJDERS TAB, BOSKER RJ. (1999). *Multilevel Analysis: an introduction to basic and advanced multilevel modeling*. Sage Publications.
- VAN der HULST M, JANSEN E. (2002). "Effects of curriculum organisation on study progress in engineering studies". *Higher Education*, vol. 43: 489-506.