

Influência do Professor e do Ambiente em Sala de Aula sobre a Proficiência Alcançada pelos Alunos Avaliados no Simave-2002

TUFI MACHADO SOARES¹

Professor Adjunto do Departamento de Estatística/UFJF
Pesquisador Associado ao Centro de Avaliação Educacional (CAED) da UFJF
Professor do Programa de Mestrado em Educação da UFJF
tufi@estatistica.ufjf.br

Resumo

Em geral os sistemas de avaliação educacional mantêm um mesmo objetivo prioritário de encontrar mecanismos para melhorar a qualidade do ensino oferecido à sociedade. Além de identificar resultados da aprendizagem dos alunos, avalia-se o conjunto do sistema educacional. Dentro desse contexto o *Sistema Mineiro de Avaliação da Educação Pública* (Simave) foi instituído pela Secretaria de Estado de Educação no ano de 2000. Os dados do contexto escolar mineiro que são disponibilizados pelo Simave-2002 estão agrupados hierarquicamente, ou seja, os alunos estão agrupados em turmas, que se agrupam em escolas. O presente estudo empregou modelos hierárquicos (multinível) com o objetivo de avaliar o impacto que características do professor e do ambiente em sala de aula apresentam sobre a proficiência dos alunos das 4^ª séries do Ensino Fundamental que fizeram os testes de Língua Portuguesa do Simave-2002. Buscando investigar os fatores que atuam na diferenciação entre as proficiências dos alunos foram considerados os modelos com os níveis de hierarquia: aluno (nível 1) e turma (nível 2). Foi encontrada explicação para uma grande proporção da variação dos escores alcançados pelos alunos quando aspectos relativos à sala de aula são considerados na modelagem.

Palavras-chave: avaliação educacional, modelo hierárquico, efeito da sala de aula.

Resúmen

En general, los sistemas de evaluación educativa tienen como objetivo primordial descubrir métodos para mejorar la calidad de la enseñanza ofrecida a la sociedad. Además de identificar los resultados del aprendizaje de los alumnos, se evalúa el conjunto del sistema educativo. Dentro de este contexto, el "*Sistema Mineiro de Avaliação da Educação Pública – SIMAVE*" (Sistema de Evaluación de la Enseñanza Pública de Minas Gerais) fue instituido por la Secretaria de Estado de la Educación en el año 2000. Los datos del contexto escolar del estado de Minas Gerais, disponibles en el SIMAVE-2002, están agrupados jerarquicamente, o sea, los alumnos están distribuidos en clases, que se agrupan en escuelas. El presente estudio utilizó modelos jerárquicos (multi-niveles) con el objetivo de valorar el impacto que tienen las características del profesor y del ambiente de la sala de clase sobre la calificación

¹ O autor agradece ao CAED pelos dados fornecidos e pelo apoio a este trabalho.

de los alumnos de la "4ª. série" (4º. Grado) de la enseñanza fundamental que hicieron los exámenes de lengua portuguesa del SIMAVE – 2002. Al investigar los factores que actúan en las diferencias entre las calificaciones de los alumnos se siguieron los modelos para dos niveles jerárquicos: alumno (categoría 1) y clase (categoría 2). Se encontró una explicación para una grande parte de la variación entre las puntuaciones alcanzadas por los alumnos cuando aspectos relativos a la clase fueron considerados en los modelos.

Palabras-clave: evaluación educacional, modelo jerárquico, efecto de la clase.

Abstract

In general, educational evaluation systems have a common objective of finding ways to efficiently improve the quality of education. Besides identifying results in the pupils' learning, the overall educational system is also evaluated. In this context, the Minas Gerais Evaluation System of Public Education (Sistema Mineiro de Avaliação da Educação Pública - SIMAVE) was introduced in 2000 by the State Secretary of Education. The data from the state of Minas Gerais regarding SIMAVE-2002 are hierarchically grouped, that is, pupils are grouped by classrooms and classrooms are grouped by schools. The present study had used multilevel models with the objective of measuring the impact of teacher and classroom aspects on the performance of 4th grade pupils that had taken the Portuguese language Simave-2002 tests. In order to investigate the factors that have a bearing on the different levels of student proficiency, multilevel models were used with the hierarchy levels: student (level 1) and classroom (level 2). Evidence was found to explain the great variation in students' scores when classroom variables were taken into account in the model.

Key words: Educational evaluation, hierarchical model, classroom effect.

INTRODUÇÃO

Os sistemas de avaliação da educação têm evoluído substancialmente, de tal forma que, a partir dos resultados obtidos, tem sido possível introduzir mudanças nos aspectos educacionais de interesse da sociedade. Atualmente um dos focos dessas avaliações é a escola. Para tanto, na composição dos sistemas de avaliação educacional mais recentes estão sendo avaliados não apenas o rendimento acadêmico dos alunos, mas também outros aspectos como os humanos e sociais. Nesse sentido, espera-se que a evolução dos instrumentos de avaliação educacional permita aos gestores e outros atores verificarem se as escolas estão enfrentando adequadamente os desafios das transformações econômicas e anseios da sociedade brasileira.

Atualmente, o principal interesse não é o de se comparar escolas apenas em relação ao rendimento de seus alunos. Mas, entre outros, o de procurar identificar características – de natureza humana, social, cultural, ética, metodológica e instrumental – que podem estar influenciando o desempenho escolar dos alunos. Além disso, tendo em vista a necessidade de se buscar para os sistemas educacionais maior eficiência e eficácia, sob qualquer que seja o enfoque dado a esses atributos, é importante medir o impacto e as interações entre essas características para cada sistema educacional específico. É preciso notar que os dados obtidos através dos sistemas de avaliação são estruturados hierarquicamente, isto é, alunos, turmas, escolas, constituem uma seqüência natural de agrupamentos aninhados, de tal forma, que as variáveis representativas das características nos diversos níveis podem interagir com outras variáveis dentro do mesmo nível hierárquico e, também, com variáveis de outro nível. Tendo em vista, principalmente, essa natureza dos dados, é freqüente a utilização de modelos de regressão hierárquicos (cf. Lee, 2001; Goldestein, 1995), que permitem investigar a influência das características de cada nível da hierarquia no desempenho escolar dos alunos e na diferenciação entre as escolas, por exemplo, e, ainda, separar a variabilidade nos resultados – associada às escolas – da variabilidade dentro de cada escola – associada aos alunos ou turmas de alunos. Embora comuns no exterior, estudos que empregam modelos de regressão hierárquicos são pouco comuns no Brasil. Entre esses poucos, menciona-se os seguintes.

Fletcher (1998) analisou em sua pesquisa educacional os efeitos das características do ambiente escolar e do ambiente familiar no rendimento dos alunos. Para o autor, as médias dos rendimentos dos alunos por escola, sem o ajuste das diferenças na composição social do alunado, distorcem os resultados das análises. Os dados utilizados por Fletcher (1998) foram os

dados do Saeb/95 colhidos através da prova de matemática e de questionário, aplicados aos alunos de 8ª série. Barbosa e Fernandes (2000), utilizaram em sua pesquisa dados do Saeb -97, colhidos através de testes e questionários com o objetivo de estabelecer uma relação entre as variáveis explicativas de dois níveis (alunos e escolas) e o rendimento escolar dos alunos da 8ª série. O objetivo do estudo de Soares *et al* (2001), foi o de conhecer o efeito das escolas de nível médio no vestibular da UFMG nos anos de 1998, 1999 e 2000 e ao mesmo tempo apresentar uma forma alternativa de avaliar os efeitos dessas escolas. Albernaz, Ferreira e Franco (2002) avaliaram o efeito de variáveis escolares, tais como a escolaridade do professor e a qualidade da infra-estrutura física no desempenho dos estudantes. Estes autores também utilizam os dados provenientes das avaliações do Saeb-97. Franco, Mandarino e Ortigão (2001) analisaram o efeito do projeto pedagógico sobre a qualidade e equidade das escolas.

Grosso modo, o que se pode observar nesses trabalhos mais recentes é que os níveis de hierarquia considerados são, quase sempre, aluno e escola. Os resultados dessas pesquisas indicam que existe sempre uma influência de agrupamento, presente no sistema escolar, no modelo de impacto das variáveis sobre o rendimento escolar, reforçando a idéia de que os alunos não estão distribuídos aleatoriamente pelas escolas.

No presente estudo, constrói-se um modelo de regressão hierárquico para a população composta pelos alunos da 4ª série que participaram da avaliação de português realizada no ano de 2002 pelo Simave/Proeb, num total de 2438 escolas, 6286 turmas e 161904 alunos. Do universo total estão excluídas aquelas escolas cujo número de alunos presentes ao teste foi inferior a 10. Este estudo foi planejado para servir de apoio e complementar, com informações de natureza qualitativa e quantitativa, as análises do relatório do Simave-2002 (cf. Relatório Técnico do Proeb/Simave-2002, 2003), visando obter um maior conhecimento do sistema educacional mineiro e fornecer subsídios para futuras medidas de governo destinadas a reverterem o quadro de baixo rendimento acadêmico dos alunos do sistema estadual de educação. Além disso, estudos posteriores, mais específicos, também podem ser construídos utilizando-se dos resultados aqui encontrados.

A proposta do presente trabalho é a de identificar a relação entre o desempenho escolar, medido pelo escore obtido no teste de Língua Portuguesa das provas do Simave/PROEB-2002, e as características da sala de aula, em articular aquelas referentes ao desempenho e dedicação do professor, controlada por variáveis associadas ao aluno, extremamente importantes na explicação da proficiência como, por exemplo, a condição

socioeconômica do aluno. Assim, o modelo proposto considera dois níveis hierárquicos, sendo o primeiro o aluno e o segundo a turma.

Na seção 2 apresenta-se a origem e a estruturação dos dados, discriminando-se as variáveis selecionadas para o estudo. Na seção 3 apresenta-se o modelo hierárquico final, com a devida interpretação e análise dos resultados obtidos de tal forma alcançar os objetivos deste estudo e a seção 4 traz uma introdução à formulação teórica para o modelo utilizado nas análises.

A ORIGEM E A ESTRUTURA DOS DADOS

O Proeb/Simave

Em fevereiro de 2000, o governo do Estado de Minas Gerais, através da Secretaria Estadual de Educação, instituiu o Sistema Mineiro de Avaliação da Educação Pública (Simave). O Simave se propõe a contribuir para uma nova cultura de avaliação educacional no estado, “compromissada com o sucesso escolar e com a educação de qualidade para todos” (cf. Relatório Técnico do Proeb/Simave-2000, 2001). Dentro dessa concepção está sendo implementado, a cada dois anos, o Programa de Avaliação da Rede Pública de Educação Básica (Proeb), que integra o Simave. Em 2002, o Proeb foi coordenado pelo Centro de Políticas Públicas e Avaliação da Educação (Caed) da Universidade Federal de Juiz de Fora (UFJF) – Minas Gerais. O Proeb é um programa de avaliação que tem por objetivo avaliar as escolas da rede estadual. O ciclo de avaliação se completa a cada dois anos. Em novembro de 2000, foram aplicados testes com o objetivo de avaliar as competências dos alunos em Língua Portuguesa e Matemática, em 2001 foram avaliadas as competências em Ciências Humanas e Naturais. Em 2002, foram aplicados novamente testes de português. Os testes são aplicados a todos os alunos da quarta e oitava séries do ensino fundamental e da terceira série do ensino médio da rede estadual. Além dos testes que avaliam as competências nessas disciplinas, o processo de avaliação inclui um questionário aplicado aos alunos, com o objetivo de obter dados sobre o perfil socioeconômico e trajetória escolar dos estudantes e, ainda, informações relevantes sobre o professor e características da turma. A metodologia empregada para avaliação pelo Proeb foi, em geral, a mesma utilizada pelo Sistema (Nacional) de Avaliação da Educação Básica (Saeb) implementado pelo Instituto Nacional de Estudos e Pesquisa (Inep) do Ministério da Educação, e é adequada para avaliação em larga escala, com precisão muito maior na avaliação das

proficiências médias de grupos de alunos do que, propriamente, na avaliação individual. Os itens (no caso, questões de múltipla escolha) desses testes foram construídos por especialistas tendo como base a proposta curricular de Minas Gerais e as matrizes de competências utilizadas pelo Saeb. Utiliza-se uma metodologia de construção de testes que é denominada de Blocos Incompletos Balanceados (BIB). Com base em um conjunto inicial de cerca de 500 itens para cada série avaliada, foram selecionados de acordo com suas características estatísticas um total de 169 itens. Esse total de itens é disposto em 13 blocos com 13 itens cada. Segundo uma combinação apropriada dos blocos, forma-se 26 cadernos de teste constituídos de 3 blocos cada. Cada aluno responde a um caderno e, conseqüentemente, a 39 itens. A forma de construção dos cadernos faz com que haja blocos comuns entre eles produzindo, assim, o que se denomina de *equalização horizontal* dos escores estimados pela calibração simultânea dos 169 itens – isto é, a estimação dos parâmetros dos modelos que são impostos aos itens. O modelo que, geralmente, vem sendo utilizado na produção das proficiências é um modelo logístico de três parâmetros cuja construção é baseada na Teoria da Resposta ao Item (TRI) (cf. Hambleton, Swaminathan, Rogers, 1991; Andrade, Tavares, Valle, 2001; Lord, 1980). Esse tipo de modelo permite produzir “escores” de proficiências que sejam independentes dos testes aplicados e da população de examinandos (*ibidem*), estas são as principais razões pelas quais seu emprego vem se universalizando em avaliações educacionais. Assim, após análises estatísticas para verificar propriedades dos itens e dos testes, tais como, *unidimensionalidade* do teste aplicado, ausência de comportamento diferencial do item entre as principais regiões do estado e as séries avaliadas (no caso dos itens em comum), ajuste dos dados empíricos ao modelo do item proposto – o que às vezes conduzem a eliminação de alguns itens, produz-se a chamada calibração definitiva dos itens (isto é, a estimação dos parâmetros dos modelos dos itens) e a estimação dos escores da **proficiência** (ou habilidade) de cada aluno. Para o presente estudo, as proficiências foram fornecidas pelo Caed/UFJF.

As variáveis utilizadas no estudo

Como o principal enfoque deste trabalho é o de avaliar o impacto de variáveis de turma sobre a proficiência, decidiu-se utilizar apenas variáveis referentes aos alunos que sejam de extrema importância na explicação dessa proficiência na construção do modelo. Nesse sentido, optou-se por manter 4 variáveis que são reconhecidamente importantes, quais sejam: uma variável indicadora da condição socioeconômica do aluno (E-SOCIO), uma variável

indicadora do sexo masculino contrastando com o feminino (S-MASC), uma variável que mede, em anos, a defasagem de idade do aluno (DEFAS) calculada pela diferença entre a idade do aluno e a idade correta para a série considerada (no caso da 4ª série, 10 anos de idade, admitindo-se uma tolerância para aqueles que completam 11 anos no segundo semestre) e uma variável indicadora de se o aluno é de raça negra (R-NEGRO) contrastando com ser de outra raça. Enfatizando, essas variáveis são apontadas em estudos anteriores como importantes na explicação da proficiência e, além disso, na análise de impacto direto foram as mais explicativas dentre aquelas avaliadas nos questionários do PROEB-2002.

O índice que representa a condição socioeconômica dos alunos, de impacto reconhecidamente importante no rendimento escolar, foi obtido através de um escore construído também com auxílio da Teoria da Resposta ao Item, com base em uma série de variáveis indicadoras da condição socioeconômica, como por exemplo: “Se a residência da criança tem água encanada ou não”, “se tem televisor ou não”, “se tem automóvel ou não”, e assim por diante. As seguintes variáveis foram obtidas a partir da *dicotomização*² das variáveis introduzidas no questionário apresentado aos alunos para possibilitar a construção de um indicador do nível socioeconômico:

Tabela 1 - Variáveis indicadoras da condição socioeconômica

Questão	Descrição
Q06	• Você trabalha?
Q07	• Onde você mora existe água encanada?
Q08	• Onde você mora existe eletricidade?
Q09	• A rua em que você mora tem calçamento ou asfalto?
Q10	• A rua em que você mora tem coleta de lixo?
Q11	• A casa em que você mora tem cozinha?
Q12	• A casa em que você mora tem banheiro?
Q13	• A casa em que você mora tem sala?
Q14	• A casa em que você mora tem quarto?
Q15	• A casa em que você mora tem televisão a cores?
Q16	• A casa em que você mora tem geladeira?
Q17	• A casa em que você mora tem máquina de lavar roupa?
Q18	• A casa em que você mora tem aparelho de som?
Q19	• A casa em que você mora tem aparelho de videocassete?
Q20	• A casa em que você mora tem freezer?
Q21	• A casa em que você mora tem telefone?
Q22	• A casa em que você mora tem telefone celular?
Q23	• A casa em que você mora tem computador?
Q24	• A casa em que você mora tem automóvel?

² Isto é, da transformação em variáveis binárias, para aquelas que já não o eram.

O primeiro passo na construção do índice foi o da análise da correlação bisserial (cf. Lord, 1980) para cada uma das variáveis acima. A correlação bisserial é uma medida clássica da capacidade de discriminação do item, representando o grau de correlação que cada item binário tem com o escore bruto calculado com todos os itens. Foi utilizado o *software testfact*[®], e todas as variáveis apresentaram coeficiente estimado para a correlação bisserial superior a 0,50, exceto a variável Q06 cujo valor de correlação bisserial foi de 0,04 indicando que essa variável não é adequada para a produção do índice pretendido. Portanto, ela foi excluída das análises posteriores.

Em seguida, uma análise da dimensionalidade associada às demais questões foi produzida utilizando-se o método de análise fatorial *Full Bayesian Factorial Analysis* (cf. Wilson, Wood, Gibbons, 1998) em amostra de 1000 alunos. Através do critério G^2 (*ibidem*), constatou-se a presença de dois fatores dominantes, e em razão desse resultado as variáveis Q07, Q08, Q10, Q11 e Q14, foram excluídas, com o que se alcançou um único fator dominante para as demais.

O índice final foi construído (usando-se o software BILOG-MG[®]) estimando-se para cada variável restante os parâmetros do modelo logístico (de dois parâmetros) da Teoria da Resposta ao Item (cf. Hambleton, Swaminathan, Rogers, 1991; Andrade, Tavares, Valle, 2001; Lord, 1980) pelo método de verossimilhança marginal máxima e, obtendo-se o escore socioeconômico por meio do método de verossimilhança máxima (cf. Andrade, Tavares, Valle, 2001).

Existem outros critérios para obtenção do índice socioeconômico, por exemplo, através da “renda familiar” ou do “nível educacional” dos pais. Porém, um grande número dos entrevistados – no caso presente alunos da 4ª série do ensino fundamental – desconhece esses dados familiares e, além desse aspecto, a abordagem direta de se inquirir sobre a renda familiar é provavelmente menos precisa. Além disso, a construção de um índice com base em vários indicadores produz, muito provavelmente, maior quantidade de informação do que apenas uma variável indicadora.

As variáveis associadas ao nível de turma que foram obtidas a partir do questionário apresentado aos alunos totalizaram 9 variáveis e são descritas na seqüência. A primeira é o número aproximado de faltas do professor (F-PROF) durante o ano. No questionário, essa questão foi avaliada a partir de uma escala ordinal – (A) *Nunca* (B) *Raramente* (C) *Algumas vezes* (D) *Muitas vezes* – e para efeito de análise neste trabalho foi aproximada para uma escala intervalar tendo em vista o número de dias letivos do ano e agregada, através de sua média, para cada turma. Uma segunda variável é a que mede se o professor passa ou não deveres de casa (DEVER), portanto,

sendo originalmente dicotômica, torna-se quando ela é agregada ao nível de turma um índice que indica o percentual de alunos da turma que afirmam que o seu professor passa deveres de casa. O turno de funcionamento da turma, se manhã (TURNO_M), tarde (TURNO_T) ou noite (TURNO_N), todas dicotômicas, foram também considerados para análise do modelo construído. No questionário respondido pelos alunos, 10 questões foram introduzidas para avaliar aspectos como os de dedicação e empenho do professor e, ainda, aspectos do ambiente escolar. Essas questões, enumeradas de Q61 a Q70, foram apresentadas da seguinte forma: admitindo-se uma escala ordinal – (A) *Em todas as aulas* (B) *Na maioria das aulas* (C) *Em algumas aulas* (D) *Nunca* – os alunos foram perguntados sobre a frequência com que os seguintes fatos ocorrem na sala de aula:

Tabela 2 - Questões sobre o ambiente em sala de aula

Questão	Descrição
Q61	• O(A) Professor(a) tem que esperar muito tempo até que os alunos façam silêncio?
Q62	• O(A) Professor(a) exige que os alunos estudem e prestem atenção nas aulas?
Q63	• O(A) Professor(a) exige que os trabalhos dos alunos sejam bem feitos?
Q64	• O(A) Professor(a) mostra interesse no aprendizado de todos os alunos?
Q65	• O(A) Professor(a) dá oportunidade de os alunos expressarem suas opiniões?
Q66	• O(A) Professor(a) está disponível para esclarecer as dúvidas dos alunos?
Q67	• O(A) Professor(a) continua a explicar até que todos os alunos entendam a matéria?
Q68	• O(A) Professor(a) se esforça para ajudar os alunos?
Q69	• Os alunos não prestam atenção ao que o professor fala?
Q70	• Há barulho e desordem na sala de aula?

Novamente, a escala ordinal foi aproximada para uma escala intervalar e, nesse caso, também, adotando-se um critério de aproximação baseado no número de dias do ano letivo. Assim, à opção (A) *Em todas as aulas* foi atribuído o escore 200 (aproximadamente 200 dias letivos), à opção (B) *Na maioria das aulas* o escore 160, à opção (C) *Em algumas aulas* o escore 40 e, finalmente à opção (D) *Nunca* o escore 0. Produziu-se, dessa forma, 10 variáveis intervalares (Q61R – Q70R) a partir das quais as análises subsequentes foram elaboradas. Após a construção dessas 10 novas variáveis, procedeu-se ao estudo das dimensões latentes a elas associadas através da análise fatorial (cf. Timm, 2002), produzida através do software SPSS[®]. Utilizou-se o método de extração por componentes principais a partir da matriz de correlação das variáveis, produzindo em seguida uma solução

através do método de rotação VARIMAX, ainda que, pelos resultados, esse procedimento de rotação seja praticamente desnecessário tendo em vista que a solução obtida sem rotação já seria suficientemente conclusiva. Para a melhor solução encontrada, três dimensões latentes foram identificadas, e juntas explicam cerca de 71% da variância total dos dados. Para uma verificação da adequação da análise fatorial foi utilizada a medida KMO (Kaiser-Meyer-Olkin), que apresentou um valor estimado de 0.85 indicando uma adequação muito satisfatória. A tabela 3, que apresenta as chamadas *cargas* de cada variável na explicação de cada uma das dimensões, permite que se faça uma análise e interpretação dessas dimensões.

Foram reforçados em negrito na tabela os valores das maiores cargas encontradas na explicação de cada dimensão. As variáveis que estão associadas a essas maiores cargas são usadas para a interpretação de cada fator. Assim, nota-se que a primeira dimensão é aquela que, num certo sentido, representa e mede o grau de interesse e dedicação do professor. A segunda dimensão representa, num certo sentido, o grau de motivação do aluno e, por conseqüência, o ambiente mais ou menos comportado em sala de aula. A terceira dimensão mede o grau de exigência do professor com relação a trabalhos e atenção às aulas. Os escores produzidos a partir da análise fatorial para essas três dimensões foram agregados, em valores médios, para cada turma e, posteriormente, padronizados, de tal forma que seus efeitos sobre a proficiência possam ser diretamente comparados através dos coeficientes de regressão do modelo construído. Portanto, com esse procedimento foram geradas três novas variáveis associadas ao nível de turma, representadas por (FAC1_T), (FAC2_T) e (FAC3_T), cada uma delas medindo uma das dimensões latentes associadas a sala de aula e interpretadas conforme o resultado da análise fatorial.

Tabela 3 - Matriz de componentes rotacionadas

Variável	Componentes		
	1	2	3
Q61R	-0.108	0.726	0.146
Q62R	0.253	0.077	0.804
Q63R	0.251	-0.017	0.805
Q64R	0.672	-0.012	0.405
Q65R	0.742	0.001	0.096
Q66R	0.792	-0.016	0.125
Q67R	0.790	0.013	0.183
Q68R	0.793	0.021	0.229
Q69R	0.175	0.646	-0.094
Q70R	-0.066	0.814	0.014

Além das 8 variáveis apresentadas acima, foi utilizado, ainda, o escore socioeconômico médio dos alunos por turma (E-SOCIOM) como uma variável associada ao nível de turma. Essa variável tem sido freqüentemente utilizada em estudos correlatos, e tem se mostrado de grande importância na construção dos modelos (ver, por exemplo, Fletcher, 1998), pelo menos no âmbito da escola. Espera-se que o mesmo ocorra no âmbito da turma.

MODELOS HIERÁRQUICOS

Os sistemas escolares são um exemplo típico de estrutura hierárquica, pois os alunos estão agrupados em turmas, as turmas agrupadas em escolas, as escolas em uma determinada localidade, e assim por diante.

O modelo multinível (cf. Goldstein, 1995) também chamado de modelo hierárquico (cf. Bryk, Raudenbush, 1992), leva em consideração a estrutura de agrupamento dos dados. Concretamente isso se reflete na especificação do modelo multinível, por exemplo, da seguinte forma: para o modelo de regressão clássico o intercepto e o coeficiente de inclinação são parâmetros fixos enquanto que para o modelo multinível o intercepto e o coeficiente de inclinação são considerados parâmetros aleatórios, dependentes da influência do nível hierárquico mais alto.

As análises que consideram em seus modelos a estrutura de agrupamento dos dados têm várias vantagens: (i) baseiam-se em modelos mais flexíveis e estruturados que utilizam melhor a informação presente na amostra, e fornecem, ainda, uma equação para cada escola ou turma, conforme o caso por exemplo, o que permite análises individuais para cada grupo; (ii) o uso da informação do agrupamento dos dados possibilita formular e testar hipóteses relativas a efeitos entre os níveis; (iii) permite a partição da variabilidade da variável resposta entre a explicação que é devida aos diversos níveis.

Os modelos de regressão multinível têm por objetivo descrever, através de um modelo matemático, a relação entre variáveis explicativas e independentes, representadas genericamente por x , e uma variável dependente y (ou mais de uma, no caso de modelos multivariados). Para a especificação dos modelos multinível, considere uma amostra aleatória de dados coletados numa estrutura com dois níveis, de tal forma que os valores das variáveis independentes sejam determinadas pelas características de cada um dos níveis. Neste trabalho só são considerados modelos com dois níveis hierárquicos. Assim, considere-se que uma amostra aleatória de dados tenha sido coletada a partir de uma estrutura em dois níveis, estando as

unidades do 1º nível agrupadas segundo as unidades do 2º nível. Cada unidade do 1º nível é representada pelo índice i , e o índice j representa cada unidade do 2º nível. Suponha que x represente uma variável do 1º nível e w uma variável do 2º nível. O modelo multinível então, terá a seguinte expressão geral:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij} \quad (1)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} w_j + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} w_j + u_{1j} \quad (3)$$

Substituindo (2) e (3) em (1) obtém-se:

$$y_{ij} = \gamma_{00} + \gamma_{01} w_j + \gamma_{10} x_{ij} + \gamma_{11} w_j x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{ij} \quad (4)$$

No modelo acima, os coeficientes apresentam a seguinte interpretação:

β_{0j} é o intercepto geral do modelo, sendo definido como variável aleatória;

β_{1j} é o coeficiente de inclinação associado à variável x . Representa o impacto da variável explicativa na variável resposta. Também é definido, *a priori*, como variável aleatória;

γ_{00} , γ_{01} , γ_{10} , γ_{11} , são parâmetros fixos a serem estimados;

u_{0j} é o denominado de efeito individual da escola, que é a componente de erro aleatório do nível 2 associada ao intercepto. Pressupõe-se ter distribuição normal com média zero e variância σ_{u0}^2 ;

u_{1j} é a componente de erro aleatório do nível 2 associada ao coeficiente de inclinação. Pressupõe-se ter distribuição normal com média zero e variância σ_{u1}^2 ;

e_{ij} é a componente de erro aleatório associado ao nível 1. Representa o resíduo da medida do rendimento do aluno não explicado pelo modelo. Pressupõe-se ter distribuição normal com média zero e variância σ_e^2 ;

σ_{u0}^2 , σ_{u1}^2 e σ_e^2 são denominados de componentes de variância do modelo.

Por hipótese, admite-se que o erro e , de 1º nível, seja independente dos erros de 2º nível. Note-se, ainda que $\beta_{01}w_j$ representa o impacto da variável explicativa w_j de nível 2 no rendimento escolar e, $\beta_{11}w_j x_{ij}$ é o termo de interação entre as duas variáveis explicativas (nível 1 e nível 2).

Na equação de regressão anterior poderão ser incluídas outras variáveis explicativas de nível 1 e também de nível 2. A estrutura para o modelo resultante é análoga àquela apresentada através das equações de (1) a (4). A extensão do modelo multinível para outras variáveis permite obter o impacto das novas variáveis no rendimento escolar bem como obter outros

termos de interação, alcançando uma maior diminuição da variabilidade total e conseqüente aumento da capacidade de explicação da variável dependente pelo modelo resultante. Além disso, é possível analisar como as diversas variáveis se interagem e como seus impactos sobre a variável dependente se comportam na presença das outras variáveis.

Conforme Goldstein (1995), um indicador do grau de agrupamento da população em estudo é o coeficiente de “intracorrelação”. Para o modelo multinível especificado o que se mede é a proporção da variância do resultado dos alunos que é devida às características do segundo nível (turma ou escola, conforme o caso). Para obter o coeficiente de intracorrelação primeiramente constrói-se um modelo multinível sem variáveis explicativas, também chamado de modelo nulo, que tem apenas três termos: β_0 , u_{0j} e e_{ij} , de tal forma que a variância total para a variável dependente nesse modelo é dada apenas por $\sigma_{u0}^2 + \sigma_e^2$. O coeficiente de intra-correlação é definido pela fórmula:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_e^2 + \sigma_{u0}^2} \quad (5)$$

Esse coeficiente toma valores no intervalo [0,1] e quanto maior o seu valor maior a proporção da variância que é devida ao segundo nível. Seu cálculo é usado para justificar o emprego de um modelo multinível ao invés de um modelo de regressão clássico.

Neste trabalho, a estimação dos coeficientes fixos é realizada através do método de mínimos quadrados generalizados (cf. Bryk, Raudenbush, 1992) e a estimação das componentes de variância é realizada através dos métodos de máxima verossimilhança plena e máxima verossimilhança restrita (*ibidem*). Para tanto, foi utilizado o *software HLM5*[®] (Raudenbush *et al.*, 2000). A medida de ajuste do modelo utilizada foi a chamada estatística de *deviance*, definida por:

$$D = -2 \text{ LOG } (L),$$

Onde L é o valor da função de verossimilhança avaliada em seu valor máximo (cf. Bryk, Raudenbush, 1992).

A CONSTRUÇÃO E A ANÁLISE DO MODELO

O processo mais utilizado na construção de um modelo hierárquico é do tipo “Botton-up”, isto é, parte-se do modelo nulo e vão-se incluindo as variáveis segundo uma heurística específica, que se baseia na verificação da significância dos coeficientes (parâmetros fixos e aleatórios) para cada modelo. Pode-se utilizar a estatística *deviance*, que é uma medida do grau de ajustamento dos dados ao modelo construído, para produzir um critério de escolha entre dois modelos aninhados. Isto é, utiliza-se a *deviance* para comparar um modelo mais simples com um modelo mais geral. Normalmente, os modelos com a *deviance* mais baixa são melhores, no entanto, para testar se a diferença entre as *deviances* de dois modelos é significativa, ou não, emprega-se o teste de significância χ^2 com o número de graus de liberdade igual à diferença de parâmetros entre os dois modelos. Outro critério empregado para inclusão ou não de variáveis no modelo foi o critério AIC (Akaike, 1974, *apud* Hox, 2001). Este critério é utilizado para comparar modelos diferentes (normalmente aninhados) e é calculado a partir do valor da *deviance* adicionado a um fator que penaliza o número de parâmetros estimados. Segundo a sugestão de Hox, 2001), este critério pode ser empregado para se decidir entre dois modelos hierárquicos. O critério AIC é dado por:

$$AIC = d + 2q \quad (6)$$

sendo que d é a *deviance* e q é o número de parâmetros estimados no modelo.

A vantagem do AIC em relação ao teste do χ^2 , para testar a diferença entre as *deviances*, é que no critério AIC não é necessário estipular um nível de significância, o que é sempre um critério subjetivo. Nesse trabalho, ambos os métodos foram empregados e não houve conflito entre os resultados.

Normalmente, recomenda-se (cf. Hox, 2001) verificar as interações entre as variáveis após a última variável ter sido incluída no modelo.

Essa heurística utilizada visa a construção de um melhor modelo multinível. No entanto, essa abordagem não é a única possível, mas pretende-se que com o seu emprego possa se alcançar um modelo suficientemente informativo para os fins. O processo é, portanto, iterativo e em cada passo uma variável (do nível de turma ou escola) é introduzida, ou não, no modelo de acordo com os critérios descritos acima. Eventualmente, uma variável anteriormente introduzida pode ser excluída a partir do momento que outra “mais explicativa” seja introduzida. As possíveis

interações entre as variáveis são testadas ao final do processo a partir daquelas variáveis que estão no modelo e outras que não se encontram no modelo.

No presente estudo o primeiro modelo analisado é o modelo nulo, representado pela equação:

$$y_{ij} = \gamma_{00} + u_{0j} + e_{ij}.$$

Para esse modelo, obteve-se uma estimativa $\gamma_{00} = 196,09$ para o parâmetro fixo, e as estimativas $\sigma_e^2 = 1363,95$, $\sigma_{u0}^2 = 788,89$, para os parâmetros aleatórios. A estatística de *deviance* final deste modelo foi de 202952. A partir da equação (5) obtém-se, portanto, um coeficiente de intracorelação de 0,366, indicando que é de grande importância na explicação da proficiência do aluno a variabilidade entre as turmas, o que justifica fortemente o emprego de modelos hierárquicos.

A partir do modelo nulo, construiu-se um modelo básico com apenas a variáveis associadas aos alunos, que servirão de controle no modelo final, no qual serão introduzidas as variáveis associadas às características do professor e da sala de aula. Esse modelo básico é representado pela equação:

$$y_{ij} = \gamma_{00} + \gamma_{10} \text{ DEFAS} + \gamma_{20} \text{ E_SOCIO} + \gamma_{30} \text{ S_MASC} + \gamma_{40} \text{ R_NEGRO} + u_{1j} \text{ DEFAS} + u_{2j} \text{ E_SOCIO} + u_{4j} \text{ R_NEGRO} + u_{0j} + e_{ij} \quad (7)$$

Tendo em vista que a população em estudo é constituída de cerca de 6000 turmas e 156000 alunos, foi necessário construir a análise em amostras desse universo de tal forma a não comprometer a interpretação dos resultados, principalmente quanto à significância das correlações encontradas. Foram utilizadas duas amostras básicas, uma com 30% das turmas avaliadas (cerca de 1800 turmas e 47000 alunos) e outra com 10% das turmas (cerca de 600 turmas e 15500 alunos). Todos os coeficientes apresentados foram estimados a partir da amostra de 30%. A significância desses coeficientes foi testada em ambas as amostras. Os resultados foram extremamente robustos em outras amostras de 30% das turmas e, razoavelmente robusta, em outras amostras de 10%. As estimativas obtidas para os parâmetros fixos e aleatórios são apresentadas na tabela 4.

Tabela 4 - Modelo básico* (proficiências de português)

Variável	Parâmetros	Parâmetros Fixos		Parâmetros Aleatórios
		Valores Estimados	Erro Padrão	Valores Estimados
INTERCEPTO	(β_{0j})			
Intercepto	(γ_{00})	206,65	0,739	
DEFAS	(β_{1j})			
Intercepto	(γ_{10})	-6,035	0,315	
E-SOCIO	(β_{2j})			
Intercepto	(γ_{20})	5,369	0,394	
S-MASC	(β_{3j})			
Intercepto	(γ_{30})	-8,895	0,564	
R-NEGRO	(β_{4j})			
Intercepto	(γ_{40})	-10,467	0,932	
	σ_e^2			1265,24
	σ_{u0}^2 (intercepto)			642,74
	σ_{u1}^2 (E_SOCIO)			18,44
	σ_{u2}^2 (DEFAS)			25,89
	σ_{u3}^2 (R-NEGRO)			109,22**
	σ_{u4}^2 (S-MASC)			*** **
DEVIANCE	201683			

* Apenas as variáveis associadas ao nível do aluno são incluídas.

** Valores não significativos na amostra de 10% ($p > 0.05$).

*** Valores não significativos na amostra de 30% ($p > 0.005$).

Como se pode notar pelo modelo estimado acima, o escore socioeconômico apresenta um efeito positivo sobre a proficiência, enquanto as demais variáveis apresentam efeito negativo, o que já era esperado tendo em vista todos os estudos anteriormente realizados. Note-se, ainda, que há evidentemente uma redução do valor da *deviance* em relação àquela encontrada para o modelo nulo, indicando grande aumento na explicação da proficiência. Esse fato é ainda reforçado pela queda na estimativa da variância do erro de nível 1, de cerca de 1363, no caso do modelo nulo, para 1265 no caso desse modelo básico.

Note-se que cada ano de defasagem escolar produz uma queda, em média, de cerca de 6 pontos na proficiência (a proficiência média da população em estudo está em torno de 194 com desvio padrão de 47), cada unidade de aumento no escore socioeconômico produz um aumento de cerca de 5 pontos na proficiência, o fato do aluno ser do sexo masculino produz uma queda esperada de cerca de 9 pontos e ser de raça negra uma queda de 10 pontos na proficiência de português. Ainda, conforme o modelo, esses

valores são diferentes para cada turma, o que indica efeitos relativos à turma, ou escola, sobre o impacto dessas variáveis.

No modelo final são introduzidas as variáveis associadas ao nível da turma conforme a heurística descrita anteriormente, incluindo a verificação de interações entre as variáveis de níveis diferentes. O modelo final construído é representado pela seguinte equação:

$$y_{ij} = \gamma_{00} + \gamma_{01} E_SOCIOM + \gamma_{02} F_PROF + \gamma_{03} DEVER + \gamma_{04} FAC1_T + \gamma_{05} FAC2_T + \gamma_{06} FAC3_T + (\gamma_{10} + \gamma_{11} E_SOCIOM) DEFAS + \gamma_{20} E_SOCIO + \gamma_{30} S_MASC + (\gamma_{40} + \gamma_{41} E_SOCIOM) R_NEGRO + u_{1j} DEFAS + u_{2j} E_SOCIO + u_{4j} R_NEGRO + u_{0j} + e_{ij} \quad (8)$$

Os valores estimados para os respectivos coeficientes são apresentados na tabela 5:

Tabela 5 - Modelo final (proficiências de português)

		Parâmetros Fixos		Parâmetros Aleatórios
Variável	Parâmetros	Valores Estimados	Erro Padrão	Valores Estimados
INTERCEPTO	(β_{0j})			
Intercepto	(γ_{00})	164,15	6,58	
E_SOCIOM	(γ_{01})	20,60	1,37	
F-PROF	(γ_{02})	-0,85	0,22	
DEVER	(γ_{03})	47,94	6,70	
FAC1_T	(γ_{04})	4,70	0,61	
FAC2_T	(γ_{05})	-6,86	0,57	
FAC3_T	(γ_{06})	4,78	0,61	
DEFAS	(β_{1j})			
Intercepto	(γ_{10})	-5,78	0,33	
E_SOCIOM	(γ_{11})	-1,95	0,64**	
E-SOCIO	(β_{2j})			
Intercepto	(γ_{20})	3,67	0,40	
S-MASC	(β_{3j})			
Intercepto	(γ_{30})	-8,87	0,56	
R-NEGRO	(β_{4j})			
Intercepto	(γ_{40})	-10,09	0,94	
E-SOCIOM	(γ_{41})	-5,83	1,99**	
	σ_e^2			1265,96
	σ_{u0}^2 (intercepto)			300,28
	σ_{u1}^2 (E_SOCIO)			17,56
	σ_{u2}^2 (DEFAS)			24,76
	σ_{u3}^2 (R-NEGRO)			86,26**
	σ_{u4}^2 (S-MASC)			*** **
DEVIANCE	200703			

** Valores não significativos na amostra de 10% ($p > 0.05$).

*** Valores não significativos na amostra de 30% ($p > 0.005$).

Note-se a queda obtida para a estatística *deviance* desse modelo, em relação àquelas obtidas para os modelos apresentados anteriormente, assim como a queda nas variâncias dos erros estimadas para o nível de turma, indicando que as variáveis associadas à sala de aula produziram um aumento substancial na explicação da proficiência através do modelo. De fato, a redução alcançada na estatística de *deviance* desse modelo para o modelo básico (apenas com variáveis do nível de aluno), cerca de 980, foi comparável àquela alcançada do modelo básico para o modelo nulo, cerca de 1270. Os resultados são coerentes com o esperado. De fato, o aumento na frequência de faltas do professor de uma turma (F-PROF) produz impacto negativo sobre a proficiência do aluno e, da mesma forma, um aumento na falta de motivação dos alunos, refletida no comportamento da turma (FAC2_T), produz queda na proficiência. Por outro lado, quanto maior dedicação e a disponibilidade por parte do professor (FAC1_T) maior é a proficiência do aluno. Professores mais exigentes (FAC3_T) e que passam deveres (DEVER) produzem, também, maior rendimento por parte dos alunos. Confirmando os resultados já observados para o efeito das escolas note-se que um aumento no escore socioeconômico médio da turma (E-SOCIOM) produz maior proficiência do aluno, o que sempre é inquietante tendo em vista, sobretudo, que o impacto da condição socioeconômica associada diretamente ao aluno também está sendo considerado no modelo. Na análise das interações só foram encontradas interações significativas entre o escore socioeconômico médio da turma com a variável que mede a defasagem escolar, e com aquela que indica se o aluno é de raça negra (na amostra de 30%, mas não amostra de 10%). Em ambos os casos, a tendência é de um maior escore socioeconômico médio por turma potencializar o efeito negativo dessas variáveis.

Finalmente, não foram encontradas evidências de que o horário da turma, se de manhã, tarde ou noite, tenha impacto sobre a proficiência em português do aluno da 4ª série.

CONCLUSÕES

Estudos, como o aqui apresentado, buscam obter conhecimento para subsidiar as soluções para problemas que são constantes preocupações de todos aqueles que estão envolvidos com sistemas educacionais. A utilização da turma como unidade de 2º nível mostrou que existe uma percentagem da variação total do desempenho que é devida a diferenças entre as turmas e outra percentagem da variação total que é devido a diferenças entre os alunos. Comprovou-se que as características do professor e do ambiente em sala de aula afetam decisivamente o

rendimento dos alunos e explicam substancial parcela da variabilidade observada para a proficiência do aluno que é devida ao efeito da sala de aula.

Outro estudo, aprofundando o que aqui foi iniciado, poderia ser o da análise mais acurada do efeito diferenciado entre as turmas para as variáveis que medem a defasagem e aquela que mede se o aluno é de raça negra, principalmente, buscando encontrar grupos de escolas, ou características do professor, para as quais esse efeito é minimizado. Em outras palavras, buscar características e atitudes adotadas por parte do professor que promovam maior “eqüidade” nos resultados das proficiências dos alunos, estudando suas práticas pedagógicas. Se essas práticas, além de produzirem essa maior “eqüidade” também promoverem maiores níveis médios de proficiência para os alunos independentes de suas características socioeconômicas, poderão ser consideradas como eficientes e servirem de modelos dentro de um sistema educacional. Como o estudo realizado neste trabalho teve como enfoque na 4ª série do ensino básico, seria interessante que estudos desse tipo também fossem realizados para a 8ª série e a 3ª série do 2º grau. De tal forma que seja possível verificar se os resultados aqui encontrados se mantêm para essas séries ou se há especificidades que os diferenciam.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBERNAZ, A.; FERREIRA, F. H. G.; FRANCO, C. Qualidade e Equidade no Ensino Fundamental Brasileiro. *Pesquisa e Planejamento Econômico*, v. 32, n. 3, 2002.

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. *Teoria da Resposta ao Item: Conceitos e Aplicações*. São Paulo: ABE – Associação Brasileira de Estatística, 2000.

BARBOSA, M. E.; FERNANDES, C. Modelo multinível: uma aplicação a dados de avaliação educacional. *Estudos em Avaliação Educacional*, Fundação Carlos Chagas, n. 22, p. 135-153, 2000.

BARBOSA, M. E. *Modelo Multinível para Dados Discretos Longitudinais*. Tese de Doutorado. Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 1999.

BRYK, S. A.; RAUDENBUSH, W. *Hierarchical Linear Models*. Newbury Park, Califórnia : Sage Publications, Inc., 1992.

FLETCHER, P. R. *À Procura do Ensino Eficaz*. Rio de Janeiro: Ministério da Educação e Cultura, Departamento da Avaliação da Educação Básica, 1998. (mimeo)

FRANCO, C.; MANDARINO, M.; ORTIGÃO, M. I. Projeto Pedagógico de Escola Promove Qualidade e Equidade em Educação? *Revista UNDIME-RJ*, VII, n.2, p. 30-46, 2001.

GOLDSTEIN, H. *Multilevel Statistical Models*. New York: John Wiley & Sons, 2 ed., 1995.

HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. *Fundamentals of Item Response Theory*. Newbury Park, Califórnia: Sage Publications, Inc., 1991.

HOX, J. Multilevel Analysis of regression and Structural equation models. In: _____. *Hierarchical models of Survey Data*. Michigan: 54th Summer Institute. Michigan, 2001.

KLEIN, R. Indicadores educacionais: disparidades regionais e sócio-econômicas no Brasil. In: BOMENY, H. *Avaliação e determinação de padrões na educação Latino-Americana*. Realidades e desafios. Rio de Janeiro: Fundação Getulio Vargas, 1997.

LEE, V. L. What are Multilevel Questions, and How might we explore them with quantitative methods? *Estudos em Avaliação Educacional*, n. 24, p. 31-68, 2001.

LORD, F. M. *Applications of item response theory to practical testing problems*. Hillsdale, New York: Lawrence Erlbaum, 1980.

RAUDENBUSH, S.; BRYK, A.; CHEONH, Y. F.; CONGDON, R. *HLM5 - Hierarchical Linear and Nonlinear Modeling*. Illinois: Scientific Software International, Inc. Lincolnwood, 2000.

RELATÓRIO TÉCNICO DO PROEB/SIMAVE-2000. *Minas Gerais: Avaliação da educação*. Juiz de Fora: Secretaria Estadual de Educação, Governo do Estado de Minas Gerais, 2001.

SOARES, J. F. *et al.* O efeito de 248 escolas de nível médio no vestibular da UFMG nos anos de 1998, 1999 e 2000. *Estudos em Avaliação Educacional*. São Paulo: Fundação Carlos Chagas, n. 24, p. 69-117, 2001.

TIMM, N. H. *Applied Multivariate Analysis*. New York: Springer Text in Statistics, Springer Verlag, 2002.

WILSON, D. T.; WOOD, R.; GIBBONS, R. *TESTFACT: Test Scoring, Item Statistics, and Item Factor Analysis*. Chicago: Scientific Software International, Inc., 1998.

Recebido em: junho 2003 Aprovado para publicação: setembro 2003

