

O conhecimento do professor em avaliação educacional e a proficiência do aluno

NEIMAR DA SILVA FERNANDES*
TUFI MACHADO SOARES**
ANDERSON CÓRDOVA PENA***
IAGO CARVALHO CUNHA****

RESUMO

Este artigo se propõe a mostrar a associação existente entre o desempenho dos alunos nos testes de proficiência aplicados pelo Proeb/Simave – Sistema Mineiro de Avaliação Escolar – e o conhecimento que os professores têm sobre o tema da avaliação em larga escala. Foi constatado que, mesmo controlados os efeitos das características socioeconômicas dos alunos, da escola e das variáveis referentes à motivação do professor e ao clima escolar, o conhecimento do docente sobre avaliação em larga escala associa-se a um maior desempenho dos alunos.

Palavras-chave: Rendimento escolar, Modelos de regressão hierárquicos, Avaliação da educação.

* Analista de Dados do Centro de Políticas Públicas e Avaliação da Educação (CAEd) da Universidade Federal de Juiz de Fora (neimar@caed.ufjf.br).

** Professor do Programa de Mestrado e Doutorado em Educação e Coordenador de Pesquisa do Centro de Políticas Públicas e Avaliação da Educação (CAEd) da Universidade Federal de Juiz de Fora (tufi@caed.ufjf.br).

*** Psicólogo do Centro de Políticas Públicas e Avaliação da Educação (CAEd) da Universidade Federal de Juiz de Fora (anderson@caed.ufjf.br).

**** Estatístico do Centro de Políticas Públicas e Avaliação da Educação (CAEd) da Universidade Federal de Juiz de Fora (iago@caed.ufjf.br).

RESUMEN

Este artículo se propone mostrar la asociación existente entre el desempeño de los alumnos en las evaluaciones aplicadas por el Proeb/Simave –Sistema de Evaluación Escolar del Estado de Minas, y el conocimiento que los profesores tienen sobre el tema de la evaluación en gran escala. Se constató que incluso controlando los efectos de las características socioeconómicas de los alumnos, de la escuela y de las variables que se refieren a la motivación del profesor y al ambiente escolar, el conocimiento del profesor sobre la evaluación en gran escala se asocia a un mayor desempeño de los alumnos.

Palabras clave: Rendimiento escolar, Modelos jerárquicos de regresión, Evaluación de la educación.

ABSTRACT

This paper aims to show the association between student performance, as measured by the Brazilian Proeb/Simave proficiency tests, and knowledge that teachers have about this large scale evaluation system. It was found that, even after controlling for socioeconomic characteristics of students and for school variables related to teacher motivation and school atmosphere, teachers' knowledge about large scale assessments is associated with students' higher performance.

Keywords: School effectiveness, Hierarchical linear models, Educational assessment.

1 INTRODUÇÃO

A avaliação educacional em larga escala tem-se disseminado no Brasil como uma importante ferramenta para o monitoramento da qualidade da educação básica. Empregada extensivamente pelo governo federal e por boa parte dos governos estaduais e municipais, nos últimos anos, os seus resultados têm tido substancial repercussão na sociedade brasileira e, cada vez mais, sido utilizados para orientar políticas públicas em educação. Contestada por parte de professores e sindicatos docentes, que veem a avaliação externa como “desnecessária”, “uma ingerência em suas atividades didáticas”, mas, principalmente, como uma “ameaça ao trabalho docente e à isonomia de tratamentos dentro da carreira docente”, tem-se revelado um importante instrumento de monitoração da qualidade do ensino.

A atitude de rejeição, entretanto, não é unânime. Alguns gestores e professores procuram conhecer melhor a forma como são avaliados e até mesmo traçar estratégias para a melhoria de seus resultados. Este trabalho se propõe a investigar se há associação entre o conhecimento dos docentes e o desempenho dos alunos nos testes. Para isso, são utilizadas informações oriundas dos questionários contextuais aplicados aos professores e alunos, bem como do desempenho dos alunos nos testes de proficiência no Proeb-2007.

O trabalho divide-se em quatro partes. Na primeira parte, são apresentadas as populações de alunos e professores estudadas, bem como suas características principais; na segunda, é construída, com base nos questionários aplicados, uma série de variáveis que descrevem as características dos docentes, dos estudantes e do meio escolar; na terceira, é apresentada uma medida do conhecimento do professor sobre as avaliações em larga escala; e, por fim, apresenta-se como esse conhecimento está relacionado ao desempenho dos alunos, mesmo quando se controla uma série de outras variáveis contextuais.

2 POPULAÇÃO ESTUDADA NAS ANÁLISES

As análises apresentadas neste estudo foram obtidas considerando as informações colhidas pelo Proeb/2007 (Programa de Avaliação da Educação Básica da Secretaria de Educação do Estado de Minas Gerais) em 4.885 escolas públicas mineiras, das quais 2.631 pertencem à rede estadual de ensino e 2.254 à rede municipal. Foram utilizados dados contextuais obtidos em questionários aplicados a 614.366 alunos do 5º e do 9º ano do ensino fundamental e do 3º ano do ensino médio. Além dos alunos, foram usadas informações de um questionário aplicado a 22.960 professores de Língua Portuguesa e Matemática que lecionam nos anos

avaliados. Os alunos, naturalmente, participaram de uma avaliação de habilidade cognitiva, em cada uma dessas disciplinas, por meio de teste de múltipla escolha. Os resultados foram aproveitados para a construção de uma medida de proficiência, utilizando-se modelos da Teoria da Resposta ao Item (TRI) (Lord, 1980). Na Tabela 1, são mostrados os totais de alunos, professores e escolas incluídos na pesquisa, em cada ano escolar avaliado.

Tabela 1 – População estudada para os anos escolares

Anos	5º EF*	9º EF*	3º EM*	Total
Escolas	4.194	2.946	1.364	4.885
Professores	10.643	8.171	4.146	22.960
Alunos	261.491	219.292	133.583	614.366

*Anos avaliados em Minas Gerais utilizando escala única.

Excluiu-se, de todas as análises, as escolas cujo número de alunos foi menor que 10 no ano escolar de ensino avaliado. Também, eliminou-se a escola em que não houve respostas dos professores aos questionários.

2.1 O instrumento contextual dos alunos

O questionário contextual aplicado aos alunos pode ser dividido em duas partes. Na primeira parte, apresentaram-se questões sociodemográficas, abrangendo cor, gênero e idade. Na segunda, com o objetivo de construir uma medida da condição socioeconômica dos alunos, solicitou-se a indicação da posse de bens de conforto, tais como geladeira, automóvel, freezer, entre outros, e, também, se a família do aluno recebe Bolsa Família.

Como um percentual expressivo dos alunos respondeu que recebia Bolsa Família, fato que pode ser constatado na tabela 2, em uma análise específica, essa variável mostrou-se melhor indicador da condição socioeconômica do que os indicadores múltiplos construídos com base na posse de bens materiais apresentados na segunda parte do questionário (Soares, 2005b). A possível explicação para o poder discriminante dessa variável é o fato de ela ser categórica, ou seja, estar menos sujeita a erros de coleta de dados.

A variável “defasagem” do aluno foi calculada pela diferença entre a sua idade e a recomendada para se cursar o ano escolar. As características dos estudantes avaliados estão descritas na tabela 2, a seguir.

Tabela 2 – Características gerais dos alunos

		Anos		
		5º EF	9º EF	3º EM
Número de alunos		261.491	219.292	133.583
Proficiência média	Língua Portuguesa	193,91	240,81	275,35
	Matemática	203,16	249,78	282,00
Gênero Masculino		50,48%	47,32%	41,61%
Cor	Branca	30,02%	31,82%	34,82%
	Parda	50,14%	47,98%	43,49%
	“Negra”	11,96%	12,72%	14,87%
	Outras	7,89%	7,48%	6,82%
Bolsa Família		48,02%	38,80%	23,13%
Defasagem	Não defasado	67,54%	61,25%	55,35%
	1 ano	19,49%	21,50%	23,12%
	2 anos	7,65%	9,84%	12,67%
	3 anos*	3,69%	4,89%	8,86%
	4 ou mais anos	1,62%	2,53%	-

*Os três anos de defasagem no 3º ano do ensino médio equivalem a três anos ou mais, uma vez que não foi possível calcular defasagens mais elevadas, em razão da formulação do questionário.

2.2 O instrumento contextual dos professores

O questionário do professor pode ser dividido em três partes. A primeira parte é formada por questões de gênero, formação, tempo de profissão e pela participação em programas de formação continuada. Na tabela 3, são apresentadas as distribuições dos docentes, segundo essas características.

Tabela 3 – Características gerais dos professores

		Anos			Total %
		5º EF %	9º EF %	3º EM %	
Gênero Feminino		97,1	80,0	72,6	87,0
Nível de escolaridade	Ensino médio	12,4	2,3	1,2	7,0
	Pedagogia	24,3	1,4	0,7	12,3
	Normal superior	32,7	1,8	1,1	16,6
	Licenciatura	19,6	79,3	81,1	50,8
	Outros	10,9	15,1	15,9	13,2
Pós-Graduação	Aperfeiçoamento (mínimo de 180 horas)	10,0	7,6	8,5	8,9
	Especialização (mínimo de 360 horas)	44,9	57,3	58,5	51,6
	Mestrado	0,3	1,7	1,7	1,0
	Doutorado	0,0	0,1	0,1	0,1
	Nenhum	44,9	33,2	31,2	38,4
Participação em atividade de formação continuada em 2007*		66,7	48,8	44,2	56,6

*Inclui atualização, treinamento, capacitação, entre outros.

Nota-se, na tabela 3, que grande parte dos professores entrevistados era do gênero feminino. No entanto, essa proporção decaiu nas séries mais avançadas. Como era esperado, um maior percentual de docentes com formação em Pedagogia e Normal Superior se concentra no 5º ano e, embora poucos possuam formação *stricto sensu*, a maioria tem alguma especialização ou, no mínimo, algum curso de aperfeiçoamento.

A segunda parte do questionário aborda os seguintes temas: relacionamento do professor com o diretor; relacionamento do professor com os colegas; motivação pessoal para o exercício do trabalho docente; relação com os pais dos alunos; indisciplina dos alunos; disponibilidade de materiais didáticos; instalações da escola; importância da avaliação em larga escala. Para avaliar a opinião dos professores, foi-lhes solicitado, no questionário, um conceito a respeito do conteúdo de determinadas expressões associadas aos diferentes temas. Por exemplo, a afirmativa “O projeto educacional desta escola é consequência da troca de ideias entre os professores” se relaciona com o tema “relacionamento do professor com os colegas”. Para esta, assim como para todas as demais expressões, solicitou-se o grau de concordância do professor, segundo a escala ordinal de Likert de cinco níveis: de “discordo plenamente” até “concordo plenamente”.

A terceira parte consiste em pequeno teste que procura medir o conhecimento do professor sobre a avaliação em larga escala. As variáveis construídas, com base nessas questões, serão apresentadas, em detalhes, a seguir.

2.2.1 Processo de construção de variáveis baseado no questionário do professor

Os temas abordados nos questionários foram escolhidos com base numa revisão da literatura e em opiniões de especialistas e gestores, sendo identificados como potencialmente relacionados ao desempenho dos alunos nos testes. As assertivas, em diversos casos, foram, propositalmente, construídas de maneira que estivessem ligadas a mais de um tema específico. Entretanto, associações inesperadas poderiam ocorrer. Assim, as respostas manifestadas para as diferentes assertivas apresentadas aos professores não são totalmente independentes, podendo estar relacionadas a um ou a mais temas.

A relação entre as respostas atribuídas às assertivas pode ser conceitualmente explicada pela existência de um *constructo* ou fator latente, não observado diretamente. Para identificar os *constructos*, ou ainda, confirmar a existência dos que foram construídos teoricamente, utiliza-se uma série de técnicas estatísticas denominadas, genericamente, de Análise Fatorial.

Na Análise Fatorial, a variância – medida de variabilidade observada para as diferentes respostas de uma assertiva – pode ser decomposta em uma parte comum, ou seja, compartilhada com as respostas de outras assertivas, e que é, teoricamente, explicada pela existência de fatores (uma possível representação do *constructo* latente), ou, em uma parte específica, aquela exclusiva à assertiva e não explicada pelos fatores (Johnson; Wichern, 1992).

Com base no modelo fatorial construído pode-se:

- (i) definir o número de fatores necessários para explicar, adequadamente, as correlações observadas para as assertivas e, conseqüentemente, o grau de explicação obtido com esse modelo; e
- (ii) observar quais são as assertivas mais correlacionadas a cada fator. Essas assertivas podem ser utilizadas com o intuito de se identificar e interpretar o fator como a representação para a construção teórica, isto é, para o *constructo*.

Na solução do primeiro problema (i), certas diretrizes podem ser utilizadas, algumas das quais foram empregadas neste trabalho, como identificar se o percentual da variância explicada pelo modelo com um determinado número de fatores é expressivamente maior que aquela explicada pelo modelo que inclui um número maior de fatores. Quanto ao segundo problema (ii), pode ser resolvido pela inspeção das chamadas cargas, estatísticas que medem a associação entre o fator e as assertivas. Uma assertiva que apresenta a sua carga, em um determinado fator, com valor absoluto baixo, provavelmente não está associada àquele fator e, portanto, é uma candidata à eliminação na explicação do fator, ou, até mesmo, do modelo. Uma solução útil, que facilita a análise e a interpretação dos fatores, é um modelo em que um conjunto de assertivas apresenta cargas elevadas para um único fator e cargas pouco expressivas para os demais. Para se alcançar uma configuração desse tipo, pode-se utilizar um procedimento muito comum, que consiste na transformação dos fatores originalmente construídos. Esse procedimento é a rotação ortogonal pelo método Varimax.

Os fatores são interpretados levando-se em consideração o conjunto das cargas de cada uma das assertivas: quanto maior for a carga, em valores absolutos, maior a influência do fator na variável (assertiva) considerada. Seu sinal indica se ambos aumentam na mesma direção ou na direção reversa. Neste trabalho, são usadas apenas assertivas com cargas superiores a 0,3, em valor absoluto, na interpretação do fator.

Finalmente, a medida do fator que será utilizada em análises posteriores, como, por exemplo, na explicação das proficiências dos alunos, é calculada empregando-se técnicas apropriadas (Johnson; Wichern, 1992).

2.2.2 Variáveis construídas com base no questionário do professor

Por meio da análise fatorial, concluiu-se que 7 fatores principais explicam 55,5% da variância total observada para as respostas às 44 assertivas (duas delas foram eliminadas por não se associarem a nenhum dos fatores). A interpretação de cada fator será apresentada a seguir e, após interpretado, o fator será considerado um *constructo*.

O *constructo* interpretado como a “Influência do diretor no trabalho do professor” está relacionado à percepção do professor sobre a ação do diretor para induzir ao ânimo, ao comprometimento e à motivação ao trabalho, às atividades inovadoras e ao bom funcionamento da escola. E está relacionado, também, à percepção sobre o respeito dispensado pelo diretor ao docente e ao respeito e à confiança que este tem pelo diretor; à capacidade do diretor em promover reuniões dinâmicas e em dar uma atenção especial à aprendizagem dos alunos.

O “Empenho da equipe” relaciona-se à percepção do professor sobre o empenho dos docentes, diretor e demais membros da equipe em suas atividades, tais como: melhorarem as aulas, serem receptivos às novas ideias; coordenarem o conteúdo entre diferentes séries; considerarem as ideias dos colegas; manterem alta a expectativa acerca do aprendizado dos alunos e se responsabilizarem por esse aprendizado; e, por fim, se o ensino da escola sofre influência do intercâmbio de ideias e experiências entre os membros da equipe.

A “Coesão e trabalho em equipe” se associa ao sentimento do professor quanto ao compartilhamento de ideias e experiências sobre o ensino e a aprendizagem na escola, as preocupações e frustrações profissionais; ao planejamento e à discussão do conteúdo programático entre as diferentes séries e da sua turma pela equipe.

O *constructo* “Importância das avaliações em larga escala” associa-se ao fato de o professor julgar serem ou não relevantes os resultados das avaliações em larga escala, de utilizar ou não esses resultados para analisar e rever suas práticas pedagógicas e seu trabalho em sala de aula. Esse fator também se correlaciona, negativamente, à assertiva: “Eu acho as avaliações desnecessárias, porque eu conheço bem os meus alunos”, fato já esperado.

A “Falta de colaboração dos pais” refere-se à percepção da participação familiar, abarcando a falta de apoio dos pais na aprendizagem dos filhos, a falta de disciplina dos alunos, a frequência irregular às aulas, bem como uma percepção, em geral, negativa acerca das famílias dos alunos: “Com as famílias que os alunos dessa escola têm, a possibilidade de aprendizado fica muito comprometida”.

A “Indisciplina dos alunos” relaciona-se à percepção dos professores quanto à indisciplina em diversos níveis: se os alunos da escola intimidam os colegas; se desrespeitam o professor e se são mais propensos ao uso de álcool e drogas. Vale destacar que os professores cuja percepção é mais aguçada em relação à indisciplina dos alunos também tendem a identificar a insensibilidade dos outros docentes diante das necessidades individuais dos estudantes.

Por último, a “Falta de materiais didáticos e instalações” está relacionada, na percepção do professor, à carência de recursos materiais e estruturais da escola, e, também, à indiferença de alguns docentes diante das necessidades individuais dos alunos e aos descasos, em geral, com a educação no país.

2.2.3 A medida do conhecimento do professor sobre a avaliação educacional em larga escala

No questionário contextual aplicado aos professores, foi inserido um pequeno instrumento a fim de se mensurar o nível de conhecimento dos docentes sobre a avaliação em larga escala. O instrumento é constituído por quatro questões (57 a 60). Cabe ressaltar que embora o número de questões seja pequeno o de respondentes é consideravelmente grande – mais de dez mil – o que possibilita o uso de modelos da TRI. Por outro lado, o erro obtido para as médias, em cada escola, é substancialmente menos importante que o erro individual de medida.

Foi utilizado o modelo de dois parâmetros da TRI para os itens, exceto para o item 57, para o qual foi usado o modelo de respostas graduadas de Samejima (1969), tendo em vista que uma outra opção, que não a correta, mostrou-se importante na escolha dos professores. Para os cálculos, foi utilizado o *software* Parscale[®] 4.1 (Muraki; Bock, 2003).

Na tabela 4, são apresentados os itens com seus respectivos percentuais de respostas e as correlações polisseriais que, semelhantes às cargas fatoriais, representam uma medida da correlação entre as respostas dadas ao item e o escore total. Quanto maior é o valor da correlação polisserial maior é a capacidade de discriminação do item.

Tabela 4 – Estatísticas dos itens

Questões	Percentual de escolha por alternativa					Não respondido %	Correlação polisserial
	1ª %	2ª %	3ª %	4ª %	5ª %		
Q.57	50,2	4,3	38,3	0,5	2,8	3,9	0,654
Q.58	66,0	0,9	3,8	20,8	5,7	2,8	0,661
Q.59	12,8	3,5	59,0	3,1	16,9	4,7	0,735
Q.60	3,1	22,1	46,9	2,7	19,3	5,9	0,603

*O percentual da resposta correta é apresentado em negrito.

Percebe-se, pelos parâmetros **b** dos itens apresentados na tabela 5, que o teste foi formado por questões com níveis de dificuldade diferenciados e por parâmetros de discriminação (**a**) razoavelmente bons, de no mínimo 0,76. Ou seja, os itens são bem informativos ao longo da escala de medida que foi produzida.

Tabela 5 – Parâmetros dos itens

Questões	Parâmetros		
	a	b₀	b₁*
Q.57	0,86	0,22	1,28
Q.58	0,88	0,16	-
Q.59	1,09	0,13	-
Q.60	0,76	0,00	-

*O modelo de crédito parcial foi ajustado somente para o item Q.57.

É apresentada, a seguir, uma interpretação do conteúdo dos itens, levando-se em consideração a intenção de seus elaboradores.

Figura 1 – Imagem do item 1

57. NO ÂMBITO DO SIMAVE, O PROEB TEM COMO FOCO DE ANÁLISE O DESEMPENHO

do aluno.

do professor.

da escola.

da superintendência.

Não sei.

Fonte: Questionário aplicado aos professores no Proeb (2007).

O item 1 examina o foco da análise da avaliação externa. O teste do Proeb é aplicado aos alunos que cursam as séries finais de um ciclo, ou seja, àqueles do 5º e 9º anos do ensino fundamental e do 3º ano do ensino médio. Apesar de a coleta de informação de desempenho ter por fonte o aluno, o seu resultado no teste não se reflete em alguma forma de progressão ou admissão, como acontece nas provas para avaliação da aprendizagem no decorrer do ano letivo, ou em exames de concursos como o Enem e os vestibulares. No caso da avaliação do Simave/Proeb, o resultado do teste aplicada o conhecimento relativo ao que a escola ofereceu aos alunos em termos de competências e habilidades. Em outras palavras, o que se tenta extrair dos testes são os resultados gerais dos alunos de cada escola. Assim, o foco da análise é, essencialmente, a escola. Essa questão, todavia, gera dúvidas nos professores, que, na maioria, acreditam que é o aluno o foco da análise das avaliações em larga escala. Não é de estranhar esse resultado, uma vez que a concepção de “aplicação de provas” remete ao universo de sala aula, em que o elemento avaliado é a aprendizagem do aluno em determinado conteúdo. Nesse último caso, a unidade correspondente é o próprio indivíduo em sua capacidade de compreensão dos conteúdos tratados pelo professor, fato que difere da avaliação externa. Pode-se supor que essa percepção aproxima a noção de desempenho aos aspectos do indivíduo, ao mesmo tempo em que a afasta do plano da instituição.

Figura 2 – Imagem do item 2

<p>58. A MATRIZ DE REFERÊNCIA PARA AVALIAÇÃO DO PROEB É COMPOSTA POR DESCRITORES QUE EXPRESSAM O CRUZAMENTO ENTRE</p> <ul style="list-style-type: none"><input type="radio"/> conteúdo curricular e habilidades/competências.<input type="radio"/> conteúdo curricular e esquemas mentais de memorização.<input type="radio"/> conhecimento e resolução de problemas.<input type="radio"/> conhecimento e aprendizagem significativa.<input type="radio"/> Não sei.

Fonte: Questionário aplicado aos professores no Proeb (2007).

O item 2 trata de informações sobre a composição da Matriz de Referência para Avaliação. A Matriz de Referência para a avaliação em larga escala é apenas uma amostra representativa da Matriz Curricular do Sistema de Ensino, no caso do Proeb, uma amostra do Conteúdo Básico Comum (CBC) de Língua Portuguesa e Matemática. A Matriz de Referência para Avaliação agrupa um conjunto de descritores de habilidades e competências que dão origem aos itens dos testes de proficiência.

Esse assunto tem sido abordado nos Boletins Pedagógicos com bastante frequência. Além disso, termos “competências”, “habilidades” e “aprendizagem significativa” que já fazem parte do discurso dos coordenadores, supervisores e orientadores pedagógicos há algum tempo. Talvez, por essa razão, a frequência de respostas tenha sido maior para a alternativa 1, seguida da alternativa 2. Assim, a alternativa correta “conteúdo curricular e habilidades/competências” é a que apresenta os conceitos mais comumente incorporados às políticas de capacitação de professores e diretrizes pedagógicas. Logo, espera-se que, mesmo aqueles que possuem um menor conhecimento acerca da Matriz de Referência para Avaliação, possam ter optado por essa alternativa, em virtude da familiaridade com os termos, o que contribuiu para o elevado índice de acerto desse item (66%).

Figura 3 – Imagem do item 3

<p>59. A MÉDIA DE DESEMPENHO ALCANÇADA PELA ESCOLA GANHOU SIGNIFICADO PEDAGÓGICO AO SER INTERPRETADA POR MEIO DA</p> <p><input type="radio"/> Matriz de referência.</p> <p><input type="radio"/> Matriz curricular.</p> <p><input type="radio"/> Escala de proficiência.</p> <p><input type="radio"/> Escala de conteúdo.</p> <p><input type="radio"/> Não sei.</p>

Fonte: Questionário aplicado aos professores no Proeb (2007).

O item 3 é basicamente uma questão que avalia se o professor reconhece uma escala de proficiência mostrada exaustivamente nos Boletins Pedagógicos. A escala de proficiência é uma representação do desempenho do aluno por meio de um esquema gráfico em que se agrupam as habilidades e competências avaliadas. Essa questão teve alto índice de acerto (59%). Esse alto índice de respostas à alternativa correta pode, também, estar vinculado à forma de divulgação dos resultados que as escolas recebem. As políticas de avaliação em larga escala, tanto em nível nacional (Prova Brasil) quanto estadual (Simave, Proalfa), apresentam, de diversas formas, os resultados dos testes, dentre eles, a média geral do desempenho da escola, bem como é a forma mais utilizada pela imprensa para elaboração de *ranking* das escolas. Assim, das alternativas apresentadas, a única que está relacionada à “média de desempenho” é a alternativa “escala de proficiência”, pois esses conceitos são apresentados de maneira vinculada nos instrumentos de divulgação dos resultados, como os Boletins Pedagógicos. Assim, mesmo aqueles que não

possuem conhecimento do assunto podem ter assinalado a alternativa correta em virtude da familiaridade com os termos “Média de Desempenho” e “Escala de Proficiência”.

Figura 4 – Imagem do item 4

60. NA ESCALA DE PROFICIÊNCIA DO SIMAVE/PROEB, AS HABILIDADES SÃO APRESENTADAS EM TRÊS CORES: BRANCO, AMARELO E AZUL. ESSAS CORES DESIGNAM A PROBABILIDADE DOS ALUNOS EM ACERTAREM AO ITEM. A COR AMARELA EXPRESSA QUE A HABILIDADE ESTÁ EM PROCESSO DE DESENVOLVIMENTO, E A PROBABILIDADE DOS ALUNOS DAQUELE NÍVEL ACERTAR O ITEM É DE

- 0% a 30%.
- 31% a 49%.
- 50% a 80%.
- 81% a 90%.
- Não sei.

Fonte: Questionário aplicado aos professores no Proeb (2007).

Em um primeiro momento, só seria possível resolver corretamente o item 4, memorizando-se os percentuais de acerto associados a cada uma das cores empregadas na interpretação da escala do Proeb e apresentados nos Boletins de Resultados. Entretanto, é possível chegar à solução correta (“50% a 80%”) por meio do bom senso e da eliminação dos distratores.

3 MODELOS LINEARES HIERÁRQUICOS DE REGRESSÃO

Os modelos de regressão tradicionais admitem quatro pressupostos básicos para as características dos dados: linearidade, normalidade, homocedasticidade e independência entre os elementos amostrais. Para uma exposição mais detalhada de tais pressupostos, pode-se consultar Raudenbush e Bryk (2002). Em geral, os três primeiros são razoavelmente admissíveis nos dados educacionais, ou contornados com base na utilização de grandes amostras. Por outro lado, não é admissível a independência dos elementos amostrais em dados de pesquisas educacionais, uma vez que a população de alunos está organizada em turmas e, estas, em escolas. Alunos de uma mesma escola tendem a apresentar características sociodemográficas mais homogêneas do que a população em geral, além de serem submetidos a processos pedagógicos que diferem dos de outras escolas. Logo, a estrutura dos dados na po-

pulação é naturalmente correlacionada e hierárquica. Por outro lado, estudos empíricos indicam que essa correlação é muito importante.

Essa dificuldade é contornada pelo uso dos modelos lineares hierárquicos, cujo objetivo principal é o de considerar a correlação das medidas em função da estrutura hierárquica dos dados. Esse modelo tem a possibilidade, ainda, de explicar, mais apropriadamente, a influência das variáveis independentes, nos diferentes níveis, sobre as proficiências dos alunos nos testes. Ele possibilita, inclusive, que algumas variáveis, como a condição socioeconômica do aluno, possam ser consideradas em diferentes níveis: no nível de aluno, como a influência da sua condição socioeconômica individual, ou no nível de escola, como a influência da condição socioeconômica de todo o grupo de alunos.

3.1 Metodologia de construção e a análise dos modelos

O processo básico mais utilizado na construção de um modelo hierárquico é o *bottom-up*, isto é, parte-se do modelo nulo – no qual somente se ajustam os termos constantes relativos a cada nível representado; logo após, são incluídas variáveis seguindo uma heurística definida pelo especialista (Soares, 2005a). Todos os cálculos foram realizados utilizando o *software* HLM⁵ 5.04 (Raudenbush; Bryk; Congdon, 2001).

Procurou-se, também, a existência de efeitos randômicos nos coeficientes estimados. O parâmetro foi aceito como fixo quando não apresentou significância ao nível de 0,001.

No Proeb-2007, só é possível construir modelos de três níveis para o 5º ano escolar, uma vez que para o 9º ano do ensino fundamental e 3º ano do ensino médio não é possível associar o professor às suas turmas.

De fato, há uma diferença conceitual importante entre os modelos de três e de dois níveis. Pode-se afirmar que ambos contribuem para o entendimento da realidade, sob um enfoque diferente, não sendo, portanto, o modelo de dois níveis apenas uma visão parcial, inferior ou menos adequada daquela obtida com o modelo de três níveis. Inicialmente, note-se que um modelo estruturado em três níveis hierárquicos só pode ser construído para escolas com pelo menos duas turmas e, portanto, para parte da população envolvida. Na estrutura de dois níveis (alunos e escola), determinados alunos são analisados comparativamente aos demais estudantes da escola, enquanto no modelo de três níveis, eles são analisados comparando-os aos colegas de turma. Por outro lado, no modelo de três níveis, as turmas são analisadas, comparativamente, dentro das escolas. Nesse modelo, por exemplo, é possível identificar

melhor o efeito do tipo de política de distribuição de alunos pelas turmas adotada pela escola, entretanto, políticas que afetam a escola como um todo, como as estudadas em questão, são melhores analisadas por modelos de dois níveis.

3.2 Modelo nulo

Pelo ajuste somente de constantes nos modelos, pôde-se estimar a proporção da variabilidade (variância) atribuída a cada nível, calculada a partir das variâncias estimadas para os erros entre o nível dos alunos e o nível de escolas (Soares, 2005a).

O resultado apresentado na tabela 6 mostra que do percentual de variabilidade total das proficiências dos alunos as variáveis de segundo nível – o nível de escola – explicam entre 11,4% e 19,0%.

Tabela 6 – Decomposição da variância entre os diferentes níveis para cada ano e disciplina

Anos	5º EF		9º EF		3º EM	
	Disciplina					
Nível	LP*	MAT*	LP*	MAT*	LP*	MAT*
	%	%	%	%	%	%
1º	83,7	81,0	88,1	85,1	88,6	84,9
2º	16,3	19,0	11,9	14,9	11,4	15,1
Total	100,0	100,0	100,0	100,0	100,0	100,0

*LP: Língua Portuguesa; MAT: Matemática.

3.3 Processo de agregação das variáveis

Para se obter as variáveis do nível de escola, calcularam-se os valores dos escores de cada *constructo* para cada professor e, posteriormente, a sua média nas respectivas escolas, obtendo-se, assim, as medidas para as variáveis “Coesão e trabalho em equipe” e “Conhecimento sobre avaliação em larga escala”, entre outras. A média desses escores constitui-se em um indicador mais fidedigno dos perfis das escolas do que apenas a medida da percepção individual. Por fim, todas essas variáveis foram padronizadas para serem utilizadas adequadamente nas análises fatoriais. Calculou-se, também, o percentual de alunos “negros”, o de meninos, o de alunos que recebem Bolsa Família e a defasagem média dos que são representados pelas variáveis “Percentual de alunos de cor ‘negra’”, “Percentual de alunos do gênero masculino”, “Percentual de alunos que recebem Bolsa Família” e “Defasagem média dos alunos”, respectivamente.

3.4 Variáveis empregadas na construção dos modelos

As variáveis utilizadas neste estudo podem ser sintetizadas no quadro 1, apresentado a seguir.

Quadro 1 – Variáveis utilizadas nos modelos de dois níveis

Nível da variável	Descrição das variáveis	Variável
Aluno	Variáveis referentes à condição socioeconômica do aluno	Gênero masculino Cor "negra" Bolsa Família
	Variáveis referentes à situação escolar	Defasagem
Escola	Variáveis referentes às questões acerca da escola, na percepção dos professores	Conhecimento do professor acerca da avaliação em larga escala Influência do diretor no trabalho do professor Empenho da equipe Coesão e trabalho em equipe Importância das avaliações em larga escala Falta de colaboração dos pais Indisciplina dos alunos Falta de materiais didáticos e instalações
	Variáveis referentes à condição socioeconômica da escola	Percentual de alunos do gênero masculino Percentual de alunos de cor "negra" Percentual de alunos que recebem Bolsa Família Defasagem média dos alunos

3.5 Modelos finais

Os coeficientes dos modelos construídos pelos procedimentos descritos anteriormente, para cada ano e disciplina estudados, estão apresentados nas tabelas 7 e 8.

Tabela 7 – Modelos em Língua Portuguesa e Matemática para cada ano escolar

Anos	5º EF		9º EF		3º EM	
	LP**	MAT**	LP**	MAT**	LP**	MAT**
Variáveis do primeiro nível	Coefficientes					
Intercepto	187,29*	190,55*	234,97*	240,82*	268,31*	275,12*
Gênero masculino	-10,06*	0,85*	-11,49*	6,89*	-6,43*	14,32*
Cor "negra"	-7,91*	-10,13*	-3,54*	-6,74*	-3,13*	-7,11*
Defasagem	-10,34*	-11,75*	-8,42*	-10,69*	-10,62*	-12,03*
Bolsa Família	-11,24*	-11,84*	-7,38*	-8,22*	-4,64*	-4,82*
Variáveis do segundo nível	Coefficientes					
Percentual de alunos do gênero masculino	-8,38	-6,06	-15,62*	-18,88*	-19,85*	-13,27*
Percentual de alunos de cor "negra"	-20,83*	-32,77*	-18,15*	-38,66*	-11,57*	-27,33*
Defasagem média dos alunos	1,59	5,71*	1,39	4,16*	-4,03*	-4,69*
Percentual de alunos que recebem Bolsa Família	-35,87*	-36,41*	-34,17*	-32,53*	-34,91*	-26,71*
Conhecimento do professor sobre a avaliação em larga escala	2,28*	2,23*	1,84*	2,37*	1,59*	2,32*
Influência do diretor no trabalho do professor	0,68*	0,80*	0,49	0,73*	0,20	0,32
Empenho da equipe	0,93*	1,15*	0,15	0,55	-0,64	-0,39
Coesão e trabalho em equipe	0,85*	1,17*	-0,79*	-0,98*	-0,79	-1,03
Importância das avaliações em larga escala	0,71*	0,72*	0,80*	1,25*	1,43*	1,67*
Falta de colaboração dos pais	-1,78*	-1,86*	-1,29*	-1,97*	-1,10*	-1,00
Indisciplina dos alunos	-1,07*	-1,53*	-1,33*	-1,72*	-1,46*	-2,30*
Falta de materiais didáticos e instalações	-2,06*	-2,24*	-1,22*	-1,37*	-0,84	-1,14*

*Coefficiente significativo ao nível de 0,01.

**LP: Língua Portuguesa; MAT: Matemática.

Tabela 8 – Efeitos randômicos em Língua Portuguesa e Matemática para cada ano considerado

	5° EF		9° EF		3° EM	
	Desvio Padrão					
	LP*	MAT*	LP*	MAT*	LP*	MAT*
Intercepto	14,03	17,00	11,83	14,69	12,09	15,95
Gênero masculino	3,87	4,37	4,29	5,84	3,89	6,09
Defasagem	4,24	4,61	2,79	3,31	3,20	3,94
Bolsa Família	4,38	5,22	-	-	-	-

*LP: Língua Portuguesa; MAT: Matemática.

4 ANÁLISE DOS RESULTADOS DOS MODELOS

4.1 Associação das variáveis do aluno com o desempenho

As variáveis do aluno associadas com o desempenho podem ser divididas em dois grupos: as características referentes às suas condições socioeconômicas e as relacionadas à sua trajetória escolar.

No primeiro grupo, estão inseridas três variáveis: o gênero, a cor/raça “negra” e “Bolsa Família”. Quanto ao gênero, os modelos indicaram associação positiva do gênero masculino com a proficiência em Matemática e negativa em Língua Portuguesa, resultado tradicionalmente citado em muitos outros estudos (Carvalho, 2003). À baixa condição socioeconômica dos alunos, representada pela variável “Bolsa Família”, associa-se uma menor proficiência, resultado muito discutido na literatura específica (Luz, 2006). Por outro lado, que a variável cor “negra” esteja associada aos piores desempenhos é um efeito também já descrito e discutido anteriormente por muitos outros autores (Roscigno; Ainsworth-Darnell, 1999); o que surpreende (Soares; Mendonça, 2003) é que se observa um efeito líquido, pois o modelo já inclui controle pela condição socioeconômica.

A “Defasagem”, uma variável característica da trajetória escolar do aluno, como vem sendo observadas neste tipo de estudo, tem influência negativa, ou seja, quanto mais defasado é o aluno menor é a sua proficiência.

Novamente, fica claro que, apesar de esses efeitos serem muito conhecidos, o que se deseja é controlar o efeito das variáveis de interesse por essas variáveis socioeconômicas. Além disso, permitem testar a robustez dos modelos e dos efeitos mencionados.

4.2 Associação das variáveis da escola com as proficiências

Inicialmente, observam-se os efeitos já esperados e relatados pela literatura; em estudos anteriores ver, por exemplo, Soares (2005a). Variáveis indicadoras das condições socioeconômicas da escola, como “Percentual de alunos que recebem Bolsa Família”, “Percentual de alunos de cor ‘negra’” e “Percentual de alunos do gênero masculino”, inseridas no segundo nível, apresentaram-se, em geral, significantes mesmo tendo sido consideradas no nível do aluno.

Pode-se, assim, interpretar esses resultados da seguinte maneira: alunos, de qualquer condição socioeconômica, gênero, ou raça/cor, apresentam menores proficiências quando inseridos em escolas onde é maior a concentração de alunos negros, de meninos e de alunos de baixa condição socioeconômica.

Não foi significativa a influência das variáveis, agregadas para o nível de escola, que descrevem o relacionamento entre os professores e o do professor com o diretor, como a “Influência do diretor no trabalho do professor”, “Empenho da equipe” e “Coesão e trabalho em equipe”. Por outro lado, as variáveis “Indisciplina dos alunos”, a “Falta de materiais didáticos e instalações” e a “Falta de colaboração dos pais” foram estatisticamente significantes, apresentando efeito negativo. O fator “Importância das avaliações em larga escala”, que mede a importância atribuída pelos professores às avaliações em larga escala, apresentou um efeito positivo e significativo sobre o desempenho dos alunos.

Uma possível explicação para a não significância de boa parte das variáveis originadas nas respostas dadas pelos professores aos instrumentos contextuais pode ser atribuída à subjetividade e à falta de neutralidade, que dificultaria um diagnóstico mais objetivo, por parte do docente, do conjunto da escola. Assim, a avaliação dos diretores e de outras características relativas à instituição, como “coesão” e “entrosamento”, ficariam comprometidas. Eysenck (1964), em suas pesquisas específicas, já informava sobre o caráter julgador que alguns questionários podem assumir. Seria, então, de esperar um diagnóstico mais acurado dos professores sobre a indisciplina dos alunos, a colaboração dos pais e os materiais e instalações na escola do que aspectos que avaliam diretamente as suas atividades ou as de seus colegas.

Entretanto, a maior importância da inclusão dessas variáveis no modelo está no fato de que elas podem se constituir em possíveis variáveis de “confundimento” para a influência da variável foco no estudo, que é o “Conhecimento do professor sobre avaliação em larga escala”. Por exemplo, presume-se que quem dá mais importância a esse tema também o conhece mais (e vice-versa) e, por outro lado, escolas em que a coesão interna é maior, maior deverá ser o conhecimento sobre a avaliação.

A variável “Conhecimento do professor sobre avaliação em larga escala” foi estatisticamente significativa em todos os modelos construídos. No entanto, é notável que as demais variáveis relativas à percepção dos professores sobre a escola parecem ter uma importância menor quando se analisa os coeficientes dos modelos. No entanto, não é possível concluir se o “Conhecimento do professor sobre avaliação em larga escala” é de fato mais importante, ou se a medida é mais precisa. Deve-se lembrar que o instrumento destinado a medir o conhecimento do professor em avaliação é mais objetivo que os demais.

CONCLUSÃO

Ficou comprovado, então, neste estudo, que o “Conhecimento do professor sobre avaliação em larga escala” está associado ao desempenho do aluno nos testes de avaliação.

Nota-se, ainda, que há uma variação expressiva de atitudes quando o tema tratado é a avaliação em larga escala. Entretanto, nesse leque de variações, é possível constatar um eixo basal, no qual tais atitudes podem ser projetadas: (i) em um extremo, estariam os mais receptivos à aplicação dos testes, os mais informados e que mais discutem e utilizam os seus resultados e, também, os mais sintonizados com os objetivos e formas de tais avaliações; (ii) no outro extremo, estariam os mais resistentes à aplicação de testes, os mais inconformados e críticos dos resultados e, possivelmente, os menos conhecedores dos objetivos de tais avaliações. Poderia se cogitar se o primeiro grupo estaria procurando ajustar suas práticas escolares às avaliações em larga escala. Assim, independentemente de suas demais características, este grupo se empreenderia em conhecer melhor o funcionamento do processo de avaliação em larga escala e, conseqüentemente, prepararia melhor os seus alunos. Por outro lado, pode-se cogitar que os professores com melhor conhecimento sobre o processo de avaliação educacional são aqueles que priorizam a formação dos alunos.

Outras hipóteses para explicar essa associação, como também as duas apresentadas, precisam ser levantadas e analisadas em estudos futuros.

REFERÊNCIAS BIBLIOGRÁFICAS

BROOKE, N.; SOARES, J. F. (Org.). *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte: UFMG, 2008. 552 p.

CARVALHO, M. P. Sucesso e fracasso escolar:

uma questão de gênero. *Educação e Pesquisa*. São Paulo, v. 29, n. 1, p. 185-193, jan./jun. 2003.

EYSENCK, H. J. *Usos e abusos da psicologia*. 2.ed. São Paulo: IBRASA, 1964.

- FREIRE, P. *Educação como prática da liberdade*. 30.ed. São Paulo: Paz e Terra, 2007.
- HOX, J. Multilevel analysis of regression and structural equation models. In: HOX, J. (Ed.). *Hierarchical Models of Survey Data*. 54th ed. Michigan: Summer Institute, 2001.
- JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 3.ed. New Jersey: Prentice Hall, 1992.
- 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. *Applications of item response theory to practical testing problems*. New York: Lawrence Erlbaum Associates, 1980.
- LUZ, L. S. Os Determinantes do desempenho escolar: a estratificação educacional e o efeito valor adicionado. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS DA ABEP, 15. 2006, Caxambu. *Anais do...* Caxambu, ABEP, 2006.
- MURAKI, E.; BOCK, R. D. *Parscale 4.01 for Windows*. Chicago: Scientific Software International, 2003.
- RAUDENBUSH, S. W.; BRYK, S. A. *Hierarchical linear models*. Califórnia: Sage, 2002.
- RAUDENBUSH, S. W.; BRYK, S. A.; CONGDON, R. *HLM 5.04 - Hierarchical linear and nonlinear modeling*. Illinois: Scientific Software International, 2001.
- ROSCIGNO, V. J.; AINSWORTH-DARNELL, J. W. Race, cultural capital, and educational resources: persistent inequalities and achievement returns. *Sociology of Education*, v. 72, p. 158-178, jul. 1999.
- SAMEJIMA, F. *Estimation of latent ability using a response pattern of graded scores*. Michigan: Psychometric Society, 1969.
- SOARES, T. M. Modelo de três níveis hierárquicos para a proficiência dos alunos de 4^a série avaliados no teste de Língua Portuguesa do Simave/Proeb-2002. *Revista Brasileira de Educação*, São Paulo, v. 29, p. 73-87, 2005a.
- _____. Utilização da Teoria da Resposta ao Item na produção de indicadores socioeconômicos. *Pesquisa Operacional*, Rio de Janeiro, v. 25, n. 1, p. 83-112, jan./abr. 2005b.
- SOARES, T. M.; MENDONÇA, M. C. M. Construção de modelo de regressão hierárquico para os dados do Simave-2000. *Pesquisa Operacional*, Rio de Janeiro, v. 23, n. 3, p. 421-441, 2003.

Recebido em: setembro 2009

Aprovado para publicação em: outubro 2010