

A VALIDADE DO ENADE PARA AVALIAÇÃO DE CURSOS SUPERIORES: UMA ABORDAGEM MULTINÍVEL¹

THE VALIDITY OF ENADE TO ASSESS THE QUALITY OF THE
UNDERGRADUATE COURSES: A MULTILEVEL PERSPECTIVE

LA VALIDEZ DEL ENADE PARA LA EVALUACIÓN DE CURSOS SUPERIORES:
UN ENFOQUE MULTINIVEL

Ricardo Primi

Marjorie Cristina Rocha da Silva

Daniel Bartholomeu

RESUMO

O presente estudo se propõe a identificar qual parcela da variância da prova do Enade está associada ao que se pretende medir e, assim, identificar a sua capacidade de se tornar um indicador da qualidade dos cursos. Foram utilizados dados de 71.838 ingressantes e 58.428 concluintes do curso de Administração no Enade de 2006, além de informações sobre as instituições e o perfil socioeconômico dos respondentes. Os resultados demonstram que as notas do Enade decorrem de diferenças importantes de níveis de habilidade no momento de entrada no curso, e que a maior parte da variabilidade das notas do Enade está dentro dos cursos, explicada pelos perfis socioeconômicos.

Palavras-chave: aprendizagem; definição de pontos de corte; qualidade da escola; competências e habilidades.

¹ Este trabalho foi produzido com o financiamento do Observatório da Educação (Capes/Inep) e do CNPq.

ABSTRACT

The present study aims to identify which part of the variance of the Enade test is associated with what it intends to measure, that is, its capacity to become an indicator of the undergraduate course's quality. Data from 71,838 new freshmen and 58,428 undergraduates of the Administration course at Enade 2006 were used, as well as information about the institutions and the socioeconomic profile of the respondents. Results show that the Enade scores is related to differences in skill levels at the entrance, and that most of Enade's variance resides within courses, explained by socioeconomic profiles.

Keywords: learning; definition of cut-off points; quality of school; competencies and abilities.

RESUMEN

El presente estudio se propone a identificar qué parcela de la varianza de la prueba del Enade está asociada a lo que se pretende medir y así su capacidad para convertirse en un indicador de la calidad de los cursos. Se utilizaron datos de 71.838 ingresantes y 58.428 concluyentes del curso de Administración del Enade en 2006, además de informaciones sobre las instituciones de enseñanza y el perfil socioeconómico de los colaboradores. Los resultados exponen que los distintos conceptos implican diferencias importantes de niveles de habilidad en el momento de entrada en el curso. Y que la mayor parte de la variabilidad de las notas del Enade se refiere al período en que los estudiantes están realizando los cursos y se explica por sus perfiles socioeconómicos.

Palabras clave: aprendizaje; definición de puntos de corte; calidade de la escuela; habilidades y competencias.

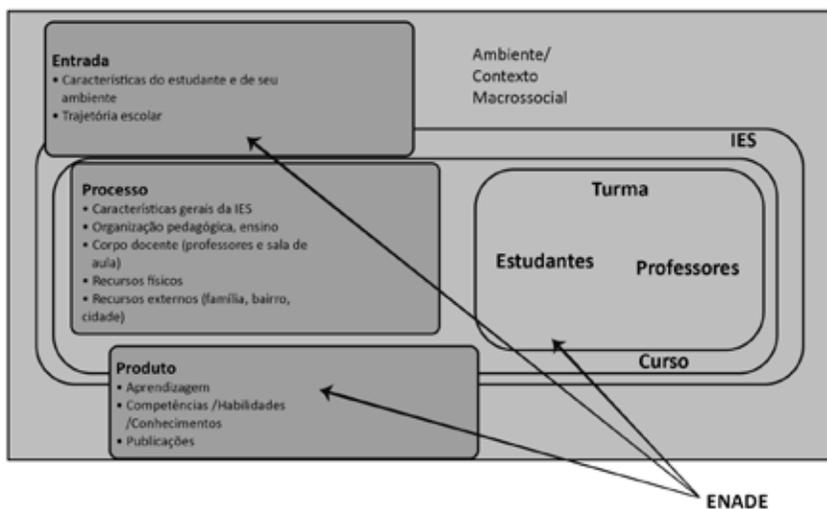
A organização multinível no Enade e sua validade

As avaliações em larga escala, desenvolvidas por entidades públicas, têm um papel fundamental para a sociedade: buscam levantar conhecimentos sobre a eficiência e a qualidade das organizações que provêm bens públicos fundamentais à população, tais como saúde, educação e segurança. Essas

informações são importantes para a gestão eficaz dos recursos públicos, uma vez que clarifica virtudes e falhas do sistema, para que ações interventivas e regulatórias sejam criadas com o objetivo amplo de melhorar sua qualidade, otimizando a utilização dos recursos públicos.

Entre as avaliações adotadas no Brasil, encontra-se o Exame Nacional de Desempenho de Estudantes (Enade). Este é um exame aplicado em larga escala, composto por questões referentes a conteúdos gerais e específicos, elaboradas com o objetivo de aferir as habilidades acadêmicas e as competências profissionais desenvolvidas pelos estudantes dos cursos de graduação das instituições de educação superior (IES). Além disso, busca reunir informações relativas às características socioeconômicas dos estudantes (LIMANA; BRITO, 2005). Embora o Enade avalie o desempenho dos estudantes, sua finalidade é medir a qualidade dos cursos de graduação no que tange a sua contribuição para o desenvolvimento de competências, habilidades e conhecimentos por seus estudantes. A figura 1 ilustra a organização hierárquica alinhada entre IES, cursos, turmas e estudantes, que é inerente a esse sistema de avaliação.

Figura 1 Estrutura multinível inerente ao Enade



Fonte: autoria própria.

Nesse modelo de avaliação há uma suposição implícita de que o curso terá efeito nos alunos ao criar condições favoráveis à aprendizagem e, por meio dela, ao desenvolvimento de conhecimentos, habilidades, competências e atitudes. É importante conceber, de maneira clara, os níveis em que esses dois objetos (estudantes e cursos) da avaliação acontecem. O curso (nível 2), hierarquicamente superior, abriga estudantes (nível 1) de tal forma que qualquer processo pedagógico executado impactará em todos que estejam aninhados nele. Assim, ocorre uma segunda suposição: de que os cursos com diferentes níveis de qualidade terão impacto diferenciado, sendo que cursos de alta qualidade provocarão desenvolvimento em nível mais elevado do que cursos de baixa qualidade. Presume-se uma relação direta entre a qualidade do curso e o desempenho de seus estudantes. Sendo assim, o desempenho agregado dos estudantes de um curso poderia compor um índice ligado à qualidade do curso com base no produto que este gera (aprendizagem). As diferenças entre os índices de distintos cursos demonstrariam as variações de qualidade entre eles. Portanto, avalia-se o desempenho dos estudantes (nível 1) para se obter medidas de qualidade dos cursos (nível 2). Como resultado, esses índices agregados são transformados em cinco notas (1 a 5), consubstanciando os conceitos dos cursos.

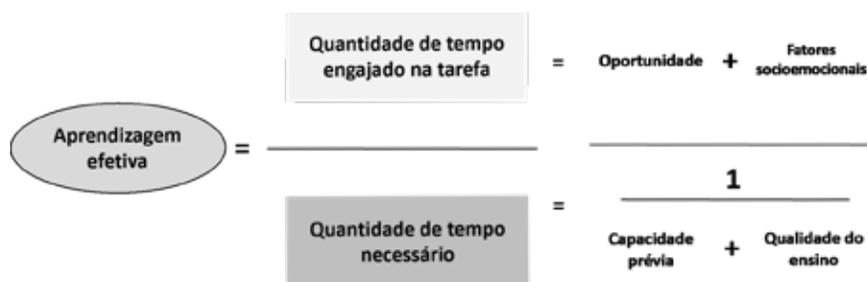
Na literatura psicométrica, essa questão é tratada dentro do conceito de validade. A validade de um sistema de avaliação diz respeito a se ele avalia o que se propõe e quão bem o faz (PRIMI; MUNIZ; NUNES, 2009; ANASTASI; URBINA, 2000; American Educational Research Association – AERA, American Psychological Association – APA, National Council on Measurement in Education – NCME, 1999; URBINA, 2004). Devido ao impacto social que tais sistemas possuem, é urgente que estudos sejam desenvolvidos para certificar a validade desses sistemas de avaliação.

Uma questão central sobre a validade do Enade se refere à possibilidade de se avaliar a qualidade dos cursos por intermédio do desempenho dos estudantes (PRIMI, 2006a, 2006b), já que a literatura sobre o tema discute as evidências do impacto de instituições escolares (IES e cursos) na aprendizagem de seus alunos (por exemplo, SOARES, 2004; 2007). Porém, existem muitos questionamentos sobre se o impacto de cursos com diferenciados níveis de qualidade é substancialmente forte e nítido a tal ponto de ser possível avaliar a qualidade de um curso a partir da aprendizagem dos alunos.

Modelo de Carroll para a aprendizagem escolar

Carroll (1963) concebeu a aprendizagem como resultado da razão entre dois componentes: o tempo despendido no objeto a ser aprendido dividido pelo tempo necessário para aprender aquele objeto (figura 2). O estudante terá sucesso na aprendizagem se dedicar o tempo necessário à tarefa que precisa aprender. Mas há fatores que influenciam no tempo necessário e dedicado à aprendizagem. O tempo necessário depende, por um lado, das capacidades cognitivas como conhecimentos prévios e capacidade de aprender e, por outro, da qualidade do ensino. O ensino é o processo de organização dos materiais e da instrução que exibirá o objeto da aprendizagem em passos sistemáticos com vistas a facilitar a eficácia da aprendizagem. O tempo dedicado depende, por um lado, da oportunidade em si de aprender determinado objeto e, por outro, da motivação (persistência, determinação), entendida como um processo de aprendizagem autorregulada com investimento de tempo e esforço na tarefa.

Figura 2 Modelo de aprendizagem escolar de Carroll (1963)



Fonte: autoria própria.

Nesse modelo, capacidade, conhecimento prévio e fatores socioemocionais (motivação, por exemplo) são aspectos observados no processo de aprendizagem dos alunos (nível 1, destacado em cinza escuro), e oportunidade e qualidade do ensino são aspectos observados no que se refere ao curso (nível 2, destacado em cinza claro). O ponto central aqui é o reconhecimento da multideterminação da aprendizagem (SOARES, 2005; GOLDSTEIN; WOODHOUSE, 2008). Assim, embora os cursos tenham impacto na aprendizagem, o qual decorre do modo como é realizado o contato com o objeto de estudo e da criação de situações eficazes para promover a

aprendizagem – isto é, por meio da qualidade do ensino –, eles não são o único fator. Conhecimento prévio, capacidade e motivação do aluno são fatores bastante importantes que afetam simultaneamente a aprendizagem.

Problematização da validade do Enade

Em síntese, a aprendizagem é medida pelo Enade com base em informações de várias fontes: nível 2 – parte associada ao curso (qualidade e oportunidade); e nível 1 – parte associada aos estudantes e a seus recursos antes de entrarem no curso. Nesse cenário, a questão principal – “É possível avaliar a qualidade do curso a partir do desempenho dos alunos?” – é analisada em dois aspectos: i) é possível identificar e isolar a parcela relevante – de variância confiável de construto relevante –, que contém informações potencialmente ligadas à qualidade do curso; e ii) em caso afirmativo, a quantidade de parcela isolada será suficiente para distinguir cinco grupos de níveis diferenciados de qualidade?

Identificação da parcela relevante

O principal ponto a ser questionado é justamente a utilização do índice de desempenho médio dos concluintes de um curso, comparado relativamente ao índice de outros cursos, como indicador direto de sua qualidade. Essa questão é justificada pela desigualdade do perfil dos ingressantes em diferentes cursos. Dessa forma, as notas dos concluintes “carregam” consigo a influência das diferenças prévias ao efeito que o curso produz no aluno (GOLDSTEIN et al., 1993; SOARES; CASTRO; RIBEIRO, 2000; GOLDESTEIN, 2001; LANDEIRA-FERNANDEZ; PRIMI, 2002; PRIMI; FERNANDEZ; ZIVIANI, 2003; RAUDENBUSH, 2004).

Sabe-se que os cursos com maior demanda passam a selecionar um grupo de estudantes com características diferentes, o que resulta em grupos desiguais entre instituições já no início do percurso acadêmico. Assim, há grupos de alunos com um conjunto de competências, conteúdos e habilidades já adquiridos ao entrar no curso, enquanto há outros com um conjunto mais modesto de recursos. O primeiro grupo iniciará o curso a partir de um ponto mais avançado que o segundo, o que refletirá em sua aprendizagem. Portanto, as diferenças de desempenho (aprendizagem) que se observam ao final do curso não podem ser explicadas exclusivamente

pelo que o curso agregou ao aluno, pois podem refletir e ser confundidas com diferenças preexistentes no perfil do estudante ingressante. Isso tem sido referido como viés de seleção, variáveis omitidas ou endogeneidade na econometria (ANTONAKIS et al., 2010).

Assim, parte das diferenças nos resultados obtidos pelos concluintes, que é explicada por essa diversidade de estudantes no início do curso, não poderá ser atribuída ao que o curso agrega, uma vez que, nesse momento inicial, o aluno ainda não está sob influência do curso. Portanto, quando se observa uma correlação significativa entre o perfil do ingressante e as medidas de qualidade obtidas com os concluintes, já se sabe que uma parcela da variância do desempenho desses últimos não pode ser usada como indicador da qualidade do curso. Desse modo, torna-se, então, necessário aplicar algum procedimento de “correção” da nota do aluno concluinte, na tentativa de expurgar diferenças associadas ao perfil inicial, com vistas a poder produzir indicadores mais depurados daquilo que, por hipótese, poderia estar associado ao efeito do curso (RAUDENBUSH, 2004a, 2004b; RUBIN, 1974).

Uma possibilidade de realizar esse controle é por meio da medida de valor agregado. Assim, como afirma Raudenbush:

[...] um tema comum na literatura é que nós devemos avaliar escolas e professores por intermédio de comparações de seus ‘valores agregados’ à aprendizagem do aluno, ao invés de níveis de desempenho não ajustados ou, como na prática corrente, a proporção de alunos em uma escola ou classe identificada como proficientes (2004b, p. 121).

O valor agregado é definido como a taxa de mudança ocorrida em um determinado tempo em que o aluno foi exposto a um ambiente educacional (FERRÃO, 2003; RAUDENBUSH, 2004b). Esse indicador pressupõe uma avaliação inicial e final, a partir da qual se pode estimar a mudança ocorrida.

Embora o Enade não adote um levantamento longitudinal avaliando o mesmo estudante quando ingressa e conclui um curso, ele avaliou, em um conjunto de edições, estudantes ingressantes e concluintes no mesmo ano. Assim, produziu dados mais próximos do que era possível com o Exame

Nacional de Cursos (ENC), para se tentar chegar a uma medida aproximada desse valor agregado. Assim, se for razoável assumir que o nível de desempenho dos alunos concluintes era, na época em que ingressaram, similar ao dos alunos ingressantes no ano em que se coletou os dados, então as diferenças médias ingressantes-concluintes, ainda que obtidas no mesmo ano, poderiam ser tomadas como indicativos de mudança típica (média). Essa é a lógica que fundamenta o Índice de Diferença de Desempenho (IDD). Com esse índice, calcula-se a média esperada do desempenho dos concluintes do curso considerando alguns preditores baseados no perfil do ingressante: desempenho médio no Enade ou no Exame Nacional do Ensino Médio (Enem) e proporção de estudantes ingressantes cujos pais tenham nível de escolaridade superior. A diferença entre o desempenho esperado, a partir do perfil prévio dos alunos que ingressam no curso, e o valor, de fato, observado entre os concluintes é o IDD.

O controle estatístico do IDD foi um avanço importante em relação ao ENC, pois tornou os índices mais consistentes e menos enviesados por variáveis irrelevantes à avaliação do curso (diferenças iniciais entre os estudantes). O IDD consiste em uma tentativa de solucionar o primeiro problema de validade, visto que, por meio do controle estatístico, isola-se a variância confiável do construto relevante de outros fatores irrelevantes à avaliação do curso (MESSICK, 1995).

Distinção dos cinco níveis diferenciados de qualidade

A segunda questão, sobre se há um montante suficiente para se criar uma distinção em cinco níveis, ainda não foi testada. Nos documentos técnicos acerca da proposição desses níveis, não há dados estatísticos que isolem os fatores relevantes e irrelevantes reportando a quantidade de variância explicada por cada um deles.

Neste trabalho foi adotada uma abordagem de análise de regressão hierárquica multinível, de forma a acomodar uma estrutura dos dois níveis já mencionados (estudantes e cursos), buscando-se identificar qual parcela de variância está associada a fatores irrelevantes e, conseqüentemente, qual parcela efetiva poderia ser usada na criação de um índice de qualidade dos cursos superiores. Com isso, pretende-se verificar se essa medida é suficiente para criar uma distinção de cinco níveis, como o atual sistema propõe.

Resumidamente, em um sistema válido de avaliação de cursos, espera-se que haja uma parte suficiente de variância associada às diferenças que são construídas durante o percurso acadêmico do aluno na IES. Esse será o propósito principal deste estudo: identificar qual percentual de variância da prova do Enade seria potencialmente útil na avaliação dos cursos, pois este está associado ao que se pretende medir. Portanto, este estudo pretende identificar as potencialidades e as limitações da prova do Enade em detectar competências, habilidades e conhecimentos que são aprendidos no contexto específico do ensino superior e, por conseguinte, a capacidade destes de se tornarem um indicador da qualidade dos cursos.

Método

Participantes

Com o objetivo de testar o modelo de análise, selecionou-se o banco de dados que contém os resultados dos estudantes do curso de Administração no Enade de 2006, além das informações do questionário socioeconômico. Esses dados representam uma amostra de 130.266 estudantes de 1.121 cursos de Administração no país, 71.838 ingressantes e 58.428 concluintes, o que corresponde aproximadamente à totalidade de cursos oferecidos na época da avaliação. Do total de participantes, 52% são do sexo feminino e têm idade média de 26 anos. O desvio-padrão foi de sete anos. Para este estudo, foram selecionados os estudantes que responderam pelo menos 70% das questões da parte específica da prova.

Em relação à amostra de IES, aproximadamente a metade está localizada no Sudeste (49%), seguida do Sul (22%) e do Nordeste (14%). A grande maioria das IES analisadas fazem parte da rede particular de ensino (90,2%).

Materiais

Foi utilizada a prova do Enade de 2006, a qual apresenta 40 questões que avaliam conhecimentos, competências profissionais e habilidades acadêmicas de estudantes do ensino superior. Dessas questões, 10 (8 objetivas e 2 dissertativas) são relativas ao componente de formação geral (FG) e 30 (24 objetivas e 6 dissertativas) são específicas da área do curso em avaliação – componente específico (CE).

Foi analisado, também, o questionário socioeconômico respondido pelos estudantes, composto por 114 questões sobre: perfil socioeconômico, dedicação às atividades acadêmicas, hábitos de leitura, conhecimento de línguas, história acadêmica, uso do computador, acesso à internet e percepção de elementos da IES e da formação por ela oferecida. Por intermédio desse questionário, é possível extrair informações importantes a respeito do perfil dos estudantes, além de obter dados contextuais característicos dos diferentes cursos. O questionário foi preenchido pelos alunos em casa e entregue por eles no dia da realização da prova. Verificou-se que 80% dos estudantes responderam a ele.²

Especificamente em relação às variáveis socioeconômicas do questionário do Enade, partiu-se da análise realizada por Silva et al. (2010), com base no Enade de 2005, para congregar grupos de itens que se esperavam estar associados ao desempenho do estudante no exame. Esses autores se basearam na análise fatorial do questionário, nas correlações dos itens com o desempenho no componente específico da prova e na análise qualitativa das questões a fim de agrupá-las e assim utilizá-las de forma mais sintetizada em estudos sobre a influência das instituições na qualidade da educação.

O quadro 1 descreve as diferentes variáveis inseridas nos modelos estatísticos construídos. A variável “escolaridade dos pais” foi inserida durante o processo de modelagem, mas descartada em etapas posteriores por não mais apresentar um efeito significativo.

2 Para uma análise detalhada do questionário socioeconômico, ver Silva et al. (2010).

Quadro 1 Escalas das variáveis explicativas do modelo de regressão multinível

NÍVEL	VARIÁVEL	DESCRIÇÃO	CODIFICAÇÃO
1	Tipo_instit	Tipo de escola	1 – Pública 2 – Mista 3 – Privada
	escol_pais	Escolaridade dos pais	Variação de 1 a 5 (nenhuma escolaridade – ensino superior)
	renda	Renda familiar	Variação de 1 a 7 (até 3 salários mínimos – mais de 30 salários mínimos)
	Sus_trab	Formas de sustento e trabalho	Variação de 1 a 5 (não trabalha – principal responsável pelo sustento)
2	curso	Variável randômica associada ao código de cada curso	1.121 valores correspondentes aos cursos
	ce.measure_mean.0	Pontuação no CE ³	- 3,0 a 3,0
	nt_fg_mean.0	Pontuação na FG ⁴	0 a 100

Fonte: autoria própria.

³ Desempenho no CE na escala theta de habilidade (ingressantes).

⁴ Desempenho na FG (ingressantes).

Procedimento e modelagem estatística

Toda variância sistemática associada ao período em que os estudantes estavam sob influência do curso pode potencialmente se caracterizar como variância confiável de construto relevante (MESSICK, 1995). Após a separação desse último componente e, portanto, depois de expurgados os fatores irrelevantes, buscou-se investigar a quantidade de variância restante e que potencialmente poderia ser usada para avaliar os cursos e refazer as médias corrigidas, comparando-se os resultados dos cursos com e sem o controle de variáveis irrelevantes.

Na modelagem estatística foi usada a Análise de Regressão Hierárquica ou Modelos Multiníveis (RAUDENBUSH; BRYK, 2002; SNIJDERS; BOSKER, 2003; FERRÃO, 2003), empregando-se o programa de computador MLwiN.⁵ Adotou-se um modelo de dois níveis: os estudantes no nível 1 e os diferentes cursos no nível 2.

Empregou-se como variável dependente o resultado do Enade dos estudantes que concluíram a parte específica da prova. Como fatores explicativos do perfil prévio dos estudantes (nível 1), selecionaram-se as variáveis indicadoras de níveis socioeconômicos (se os estudantes trabalhavam para conseguir se sustentar, escolaridade dos pais e renda familiar). Essas variáveis foram incluídas no modelo como fatores fixos centralizados na média geral, que expurgam parte da variância associada a diferenças prévias de perfil socioeconômico dos estudantes; certamente importantes, mas irrelevantes para a avaliação da qualidade dos cursos. No nível 2, a variável “curso” configurou-se na especificação de agrupamentos de alunos de cada curso, configurando-se em um termo constante e randômico, correspondente à expectativa de desempenho médio dos alunos de cada curso quando todas as variáveis de controle do nível 1 fossem iguais a zero. Assim, no termo constante, há a representação da variabilidade do desempenho médio dos estudantes de cada curso expurgando-se as diferenças prévias decorrentes dos fatores socioeconômicos.

⁵ Disponível em: <<http://www.bristol.ac.uk/cmm/software/mlwin/>>.

Ainda no nível 2, buscou-se explicar as variabilidades dessas médias com base nos resultados médios no Enade dos ingressantes de cada IES. Assim, subtrai-se da variabilidade das médias dos cursos aquelas diferenças do desempenho médio dos ingressantes. Os resíduos desse termo constante no nível 2 consistem na parcela de variância que poderia ser atribuída às eventuais diferenças de qualidade entre cursos. Com isso, buscou-se verificar a proporção de variância relevante disponível no desempenho dos estudantes. Assim, é possível decompor a variância do desempenho do estudante concluinte no Enade separando-se a parcela mais relevante na construção de indicadores de qualidade dos cursos. A última etapa dessa análise consistiu em estimar os resíduos no nível 2 do modelo, que foram constituídos pelas notas mais refinadas dos cursos, eliminando-se parcelas irrelevantes da variância. Tendo essas notas, foi feito um estudo correlacional com os conceitos dos cursos calculados pelo método tradicional, a fim de se discutir os aspectos referentes à validade do uso do Enade para verificar a qualidade dos cursos.

Resultados

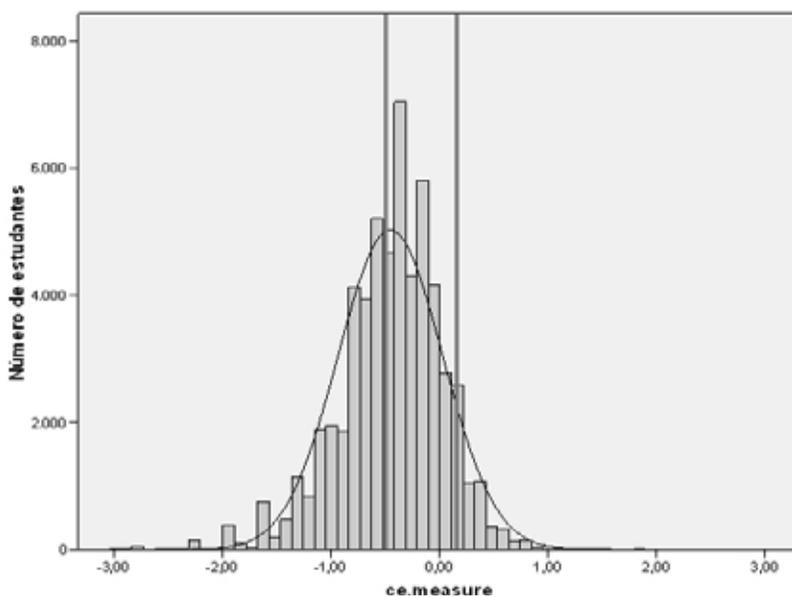
Para a análise dos dados, foi considerado como variável dependente o desempenho dos alunos concluintes da parte relativa ao CE da prova. Primeiramente foram eliminados da análise os estudantes que responderam menos de 23 questões, a fim de selecionar apenas aqueles que responderam 75% dos itens de CE. Utilizou-se o resultado do Enade dos estudantes concluintes do CE estimado com a Teoria de Resposta ao Item (TRI), cujos detalhes são apresentados em Silva et al. (2011). Essa escala é similar à escala Z, com valores típicos de -4 a +4, com média 0. Entretanto, em razão do procedimento de calibração empregado, a média 0 foi centrada na escala de dificuldades dos itens. Portanto, o valor 0 corresponde à parte dos itens com dificuldade média na prova. A escala de habilidade dos estudantes é relativa a esse ponto, o que faz com que a média não seja 0 porque a média de acertos é menor do que 50%. Na tabela 1 e no gráfico 1, são apresentadas as estatísticas descritivas e a distribuição das notas dos alunos na prova de CE.

Tabela 1 Estatísticas descritivas dos estudantes na prova de CE por momento do curso

MOMENTO/ESTATÍSTICAS	
INGRESSANTES	
Número total	71.496
Média	-0,74
Desvio-padrão	0,54
Mínimo	-4,87
Máximo	3,02
CONCLUINTES	
Número total	58.066
Média	-0,46
Desvio-padrão	0,48
Mínimo	-4,84
Máximo	2,26

Fonte: autoria própria.

Gráfico 1 Distribuição de habilidades no componente específico (concluintes)



Fonte: autoria própria

As variáveis independentes foram centralizadas a fim de permitir a comparação entre elas. Primeiramente, foi analisado o modelo nulo, sem variáveis explicativas, para identificar qual parte da variância total poderia ser atribuída aos níveis 1 e 2. A notação formal do modelo é a seguinte:

$$\begin{aligned}y_{ij} &= \beta_{0j} + e_{ij} \\ \beta_{0j} &= \beta_0 + \mu_{0j} \\ \mu_{0j} &\sim N(0, \sigma_{\mu j}^2) \\ e_{ij} &\sim N(0, \sigma_e^2)\end{aligned}$$

A notação y_{ij} representa o desempenho na parte específica do Enade do estudante i no curso j . O desempenho é previsto por um parâmetro β_{0j} , composto pela média geral da amostra β_0 e pela média dos concluintes do curso j , do qual o estudante i é parte; μ_{0j} representa o desvio da IES em relação à média geral e mais um resíduo e_{ij} , relacionado ao estudante, que indica o quanto ele se afasta da previsão com base na média da IES.

Assim, esse modelo separa a variância total em dois componentes: *i*) a variância dos estudantes com relação à média de seu curso σ_e^2 , que pertence ao nível 1; e *ii*) a variância da média dos estudantes de uma IES j com relação à média geral $\sigma_{\mu j}^2$, que pertence ao nível 2.

Vale salientar que o modelo tenta explicar a variância total ($\sigma_{\mu j}^2 + \sigma_e^2$) empregando os parâmetros da IES no nível 2. Assim, serve para verificar a variância explicada $\sigma_{\mu j}^2$ versus a não explicada σ_e^2 , o que permite calcular o coeficiente de correlação intraclasse (r_{ICC}) ao dividir-se a variância do nível macro (IES) pela variância total. Portanto, ela pode variar de 0 a 1, com base na fórmula:

$$r_{ICC} = \frac{\sigma_{\mu j}^2}{\sigma_{\mu j}^2 + \sigma_e^2}$$

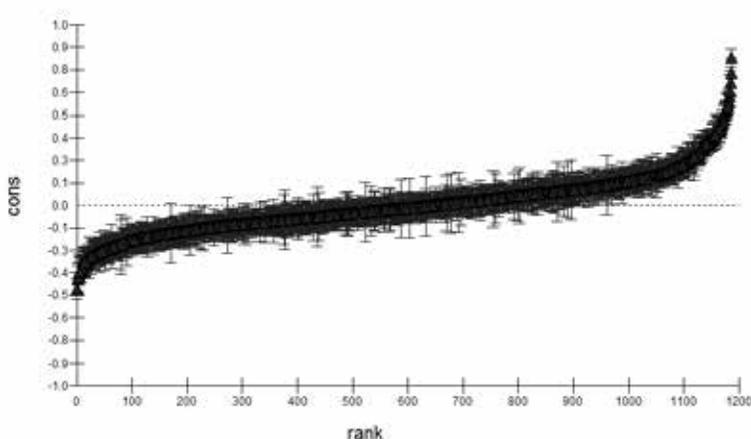
O valor nulo ou próximo de zero indica que as instituições são homogêneas entre si e que o desempenho escolar do aluno é independente do local que ele frequenta (FERRÃO; FERNANDES, 2003; JESUS; LAROS, 2004). Indica ainda que a análise multinível não precisa ser aplicada. No caso do presente estudo, tal resultado mostraria que a diferença nos desempenhos dos

estudantes não pode ser atribuída necessariamente à IES frequentada pelo estudante.

Portanto, nesse modelo, no nível 1 ($N = 45.550$), estão os estudantes agrupados nos cursos das IES brasileiras que participaram do Enade 2006, concluintes do curso de Administração, que, por sua vez, constitui o nível 2 (1.121 cursos). De acordo com o modelo nulo, a estimativa da variância entre as instituições é de 0,031. A variância entre os estudantes dentro das instituições é igual a 0,204 e a variância total é de 0,235. Desse modo, foi possível calcular o coeficiente de correlação intraclasse, que mede o quanto a variação total decorre das diferenças entre as instituições. Segundo o resultado observado, somente 13,2% da variação total se deve à variação entre os cursos dessas instituições.

O gráfico 2 mostra os valores do parâmetro das 1.121 instituições, ordenados do menor para o maior, de modo que é possível visualizar a existência de pouca variabilidade entre os cursos. Uma parcela maior da variância ocorre dentro dos cursos. Pode-se inferir que há efeito pequeno dos *clusters* formados pelas IES no desempenho dos estudantes.

Gráfico 2 Ranking das instituições no componente específico, no modelo nulo



Fonte: autoria própria.

Na etapa seguinte, o modelo foi ampliado incluindo-se as variáveis do nível 1 (nível socioeconômico do aluno, por exemplo) e do nível 2 (média

no Enade dos estudantes ingressantes nos cursos), com o objetivo de remover a variância irrelevante para a avaliação da instituição. A notação geral do modelo é:

$$\begin{aligned}
 y_{ij} &= \beta_{0j} + (X_{1ij} + \dots + X_{vij}) + e_{ij} \\
 \beta_{0j} &= \beta_0 + (\beta_{1j} + \dots + \beta_{vj}) + \mu_{0j} \\
 \mu_{0j} &\sim N(0, \sigma_{\mu_j}^2) \\
 e_{ij} &\sim N(0, \sigma_e^2)
 \end{aligned}$$

Nesse modelo, os termos $X_{1ij} + \dots + X_{vij}$ indicam as variáveis X_{vij} preditoras consideradas no nível 1. Já $\beta_{1j} + \dots + \beta_{vj}$ indicam as variáveis preditoras β_{vj} do nível 2. É importante ressaltar que a variância total dos termos nesse modelo ($\sigma_{\mu_j}^2 + \sigma_e^2$) é menor que no modelo nulo, e a magnitude da redução indica o montante de variância explicado pelas variáveis que foram inseridas:

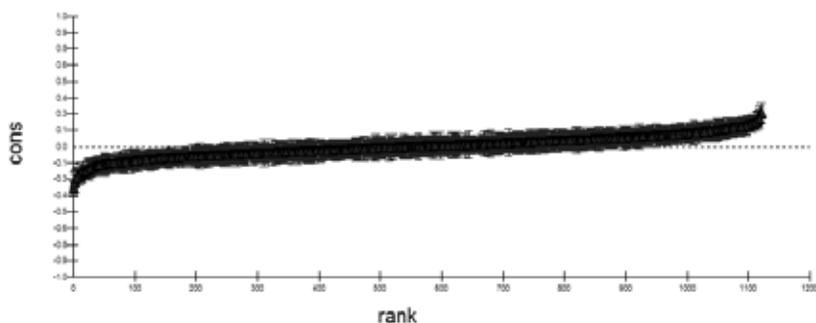
$$r = \frac{\sigma_{Tot_nulo}^2 - \sigma_{Tot_odt}^2}{\sigma_{Tot_nulo}^2}$$

Como já dito, esse modelo inclui, *a priori*, as principais variáveis do nível 2 relacionadas ao perfil cognitivo dos estudantes, a saber: o desempenho no CE e na FG como medidas do perfil típico médio dos ingressantes na IES. Assim, é feito o controle do desempenho inicial no nível 2.

Nesse modelo, a variância residual das médias das IES ($\sigma_{\mu_j}^2$) representa uma parcela de variância mais ‘pura’, pois remove parte da variância que não poderia ser atribuída à qualidade da IES, que compreende fatores externos relacionados ao perfil de sua clientela. Verificou-se que a variância residual das médias das IES reduziu-se de 13,2% (do modelo nulo) a 5% após inclusão dos desempenhos médios dos ingressantes no modelo. Em seguida, com a inclusão das variáveis preditoras do nível 1 (escolaridade dos pais, renda familiar, formas de sustento e trabalho), a variância foi reduzida a 4%.

Ao visualizar o gráfico 3, semelhante ao 2 (feito com base no último modelo), é possível verificar que a variabilidade entre os cursos é ainda menor no modelo final. Desse modo, é possível inferir que as diferenças se dão mais entre os alunos do que entre os cursos.

Gráfico 3 Ranking das instituições no componente específico após controle



Fonte: autoria própria.

Assim é possível analisar os efeitos das variáveis explicativas de ambos os níveis. Os resultados encontrados podem ser vistos na tabela 2, que apresenta as estatísticas referentes ao modelo nulo e final encontrado neste estudo.

Tabela 2 Modelo sem variáveis explicativas criado após a inserção das variáveis de controle

VARIÁVEIS EXPLICATIVAS	MODELO NULO (SEM VARIÁVEIS EXPLICATIVAS)			MODELO DE REFERÊNCIA (APÓS INSERÇÃO DE VARIÁVEIS DE CONTROLE)		
	EFEITO	EP	RAZÃO TOTAL	EFEITO	EP	RAZÃO TOTAL
Efeitos Fixos						
Intercepto (β_0)	-0,467			-0,453		
ce.measure_mean.0 (β_{1j})				0,504	0,035	14,4
nt_fg_mean.0 (β_{2j})				0,009	0,001	9
Renda (X_{1ij})				0,029	0,002	14,5
Sus_trab (X_{2ij})				0,055	0,002	27,5

VARIÁVEIS EXPLICATIVAS	MODELO NULO (SEM VARIÁVEIS EXPLICATIVAS)			MODELO DE REFERÊNCIA (APÓS INSERÇÃO DE VARIÁVEIS DE CONTROLE)		
	EFEITO	EP	RAZÃO TOTAL	EFEITO	EP	RAZÃO TOTAL
Nível 2 (variância de μ_{0j} $\sigma_{\mu_j}^2$)	0,031			0,011		
Nível 1 (variância de e_{ij} σ_e^2)	0,204			0,201		
Correlação intraclasses		0,13			0,04	
Deviance		72969,06			57251,29	
Número de parâmetros estimados		3			7	
Diferença nos deviances					3929,444	
Diferença do número de parâmetros					4	

Fonte: autoria própria.

Uma última etapa consistiu em analisar as diferenças de desempenho dos estudantes segundo os conceitos do Enade, antes e após o controle das variáveis pelo modelo final da análise multinível (tabela 3). Antes do controle, há, como esperado, uma diferença marcante entre os cursos das IES associados ao seu conceito, uma vez que cursos com conceito mais alto apresentaram médias mais altas. No entanto, depois do controle, as diferenças entre os cursos com conceitos distintos diminuem radicalmente e se tornam muito pequenas. Assim, grande parte das diferenças entre os cursos pode ser atribuída ao perfil socioeconômico e cognitivo dos alunos que ingressam em cada IES.

Tabela 3 Diferenças de desempenho dos estudantes por conceito antes e após controle de variáveis externas

CONCEITO	ANTES DO CONTROLE		APÓS O CONTROLE	
	MÉDIA	ERRO-PADRÃO DA MÉDIA	MÉDIA	ERRO-PADRÃO DA MÉDIA
1	-0,84	0,017	-0,05	0,019
2	-0,63	0,004	-0,02	0,004
3	-0,45	0,003	0,00	0,002
4	-0,24	0,006	0,02	0,006
5	0,02	0,009	0,04	0,010

Fonte: autoria própria.

Discussão

Este estudo decompôs a variância do desempenho no Enade do estudante concluinte separando a parcela mais relevante para a construção de indicadores de qualidade dos cursos. O resultado mais importante indica que a maior parte da variabilidade das notas do Enade está dentro dos cursos, e não entre os cursos. Verificou-se que a correlação intraclasse diminuiu de 0,13 para 0,04 depois da inserção das variáveis de controle. Esse valor é compatível ou até menor que o valor da correlação intraclasse de outros estudos similares (FERRÃO; FERNANDES, 2003; JESUS; LAROS, 2004; SOARES, 2004; SOARES et al., 2001, 2004), apesar da ressalva de que foram feitos no âmbito do ensino fundamental e não do ensino superior.

Portanto, os resultados sugerem que, antes do controle, grande parte da variabilidade das médias entre as instituições que irão compor as atuais notas dos cursos (de 1 a 5) é explicada pelos perfis socioeconômicos e o desempenho prévio dos estudantes. Assim, o indicador de qualidade com base na média dos concluintes é em grande parte confundido com as condições prévias dos estudantes, o que dificulta identificar a parcela que pode ser atribuída à influência do curso (SOARES, RIBEIRO; CASTRO, 2001; PRIMI, 2006a, 2006b). Tais dados apontam a relevância de se controlar estatisticamente o efeito dessas variáveis, caso contrário se atribuiria aos cursos e, assim, também às instituições um efeito que não cabe a estes (FERRÃO, 2003; FERRÃO; FERNANDES, 2003, PRIMI, 2006a, 2006b; BROOKE; SOARES, 2008).

A variabilidade das médias dos cursos, decorrentes dos fatores socioeconômicos e do desempenho prévio, que foi o foco de análise deste artigo, é um componente equivalente ao IDD. As variáveis de controle usadas e o método (multinível *versus* regressão com dados agregados) não são idênticos, mas são índices bastante similares. Uma questão importante neste estudo, e que também se refere ao IDD, é se 4% da variabilidade total do desempenho do Enade é um montante suficientemente útil para se criar um índice que diferencia a qualidade dos cursos. Um ponto importante nessa comparação é que, no cálculo da correlação intraclasse que acarreta essa proporção, o denominador refere-se à variância no Enade entre indivíduos, e o numerador à variância entre médias de cursos. Este apresenta resultado naturalmente menor do que aquele. Ao considerar-se que as amostras

de números de alunos de cada curso vêm de uma mesma população, a variabilidade das médias seria dada por $\sigma\sqrt{N}$. Essa diferença de escalas entre a média dos cursos (nível 2) e a média dos estudantes (nível 1) geralmente não é considerada na discussão das correlações intraclasse, mas é certamente importante na análise psicométrica desse índice, ao ser julgado se esse montante é útil ou não.

Desse modo, a nota final do Enade, com base no desempenho dos concluintes sem controle, contém uma parcela considerável de variância irrelevante que pode comprometer a validade do índice. Este estudo ressalta a importância de índices, como o IDD e, alternativamente, destaca a abordagem multinível como forma de construir índices potencialmente mais úteis para a avaliação da qualidade dos cursos. Seria importante para estudos futuros e para a avaliação da validade do IDD a análise da correlação desse índice com outros indicadores mais diretos da qualidade dos cursos, bem como análises econométricas de identificação do impacto dos cursos e do quanto esse se relaciona ao IDD.

Referências

AERA; APA; NCME. *Standards of Educational and Psychological testing*. Washington, DC: American Educational Research Association, 1999.

ANTONAKIS, J. et al. On making causal claims: a review and recommendations. *The Leadership Quarterly*, v. 21, n. 6, p. 1086-1120, 2010.

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

CARROLL, J. B. A model of school learning. *Teachers College Record*, v. 64, p. 723-733, 1963.

FERRAO, M. E. *Introdução aos modelos de regressão multinível em educação*. Campinas, São Paulo: Komedi, 2003.

FERRAO, M. E.; FERNANDES, C. O efeito-escola e a mudança – dá para mudar? Evidências da investigação brasileira. *Revista Iberoamericana sobre Calidad, Eficacia y*

Cambio en Educaci3n (Reice), v. 1, n. 1, 2003. Dispon3vel em: <<http://www.ice.deusto.es/rinace/reice/vol1n1/FerraoyFernandes.pdf>>. Acesso em: 8 jun. 2009.

GOLDSTEIN, H. The difficulty of ranking schools: the limits to 'value-added'. *New Economy*, v. 27, n. 4, p. 391-405, 2001.

GOLDSTEIN, H. et al. A multilevel analysis of school examination results. *Oxford Review of Educacion*, v. 19, n. 4, p. 425-433, 1993.

GOLDSTEIN, H.; WOODHOUSE G. Pesquisa sobre efic3cia escolar e politicas educacionais. In: BROOKE, N.; SOARES, J. F. (Org.). *Pesquisa em efic3cia escolar: origem e trajet3rias*. Belo Horizonte: Editora UFMG, 2008. p. 411-424.

JESUS, G. R.; LAROS, J. A. Efic3cia escolar: regress3o multin3vel com dados de avalia3o em larga escala. *Avalia3o Psicol3gica*, v. 3, n. 2, p. 93-106, 2004.

LANDEIRA-FERNANDEZ, J.; PRIMI, R. Compara3o do desempenho entre calouros e formandos no Prov3o de Psicologia 2000. *Psicologia: Reflex3o e Cr3tica*, v. 15, n. 1, p. 219-234, 2002.

MESSICK, S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, v. 50, n. 9, p. 741-749, 1995.

PRIMI, R. Evid3ncias de validade das provas do Enade – 2004. In: RISTOFF, D.; LIMANA, A.; BRITO, M. R. (Org.). *Enade: perspectiva de avalia3o din3mica e analise de mudan3as*. Bras3lia: Instituto Nacional de Estudos e Pesquisas Educacionais An3sio Teixeira (Inep), p. 59-76, 2006a.

PRIMI, R. A validade do *Enade para avalia3o da qualidade dos cursos de institui3es de ensino superior*. Projeto de Pesquisa. Itatiba: Universidade S3o Francisco, LabAPE, 2006b.

PRIMI, R.; FERNANDEZ, J.L.; ZIVIANI, C. O prov3o de psicologia: objetivos, problemas, consequ3ncias e sugest3es. *Psicologia: Teoria e Pesquisa*, v. 19, n. 2, p. 109-116, 2003.

PRIMI, R.; MUNIZ, M.; NUNES, C. H. S. S. Defini3es contempor3neas de validade de testes psicol3gicos. In: Claudio Simon Hutz. (Org.). *Avan3os e pol3micas em avalia3o psicol3gica*. S3o Paulo: Casa do Psic3logo, p. 243-265, 2009.

RAUDENBUSH, S. W. *Schooling, statistics, and poverty: can we measure school improvement?* New Jersey: Educational Testing Service, 2004a.

_____. What are value-added models of estimating and what does this imply for statistical practice. *Journal of Educational and Behavioral Statistics*, v. 29, n. 1, p. 121-129, 2004b.

RAUDENBUSH, S. W.; BRYK, A. S. *Hierarchical linear models: applications and data analysis methods (advanced quantitative techniques in the social sciences)*. London: Sage Publications, 2002.

RUBIN, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, v. 66, n. 5, p. 688-701, 1974.

SILVA, M. C. R. et al. Teoria de Resposta ao Item: equalização da prova de Administração do Enade. In: CAPOVILLA, F. C. (Org.). *Transtornos de aprendizagem: progressos em avaliação e intervenção preventiva e remediativa*, v. 1. São Paulo: Memnon, 2010. p. 239-246.

SNIJEDERS, T. A. B.; BOSKER, R. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. London: Sage Publications, 2003.

SOARES, J. F. O efeito da escola no desempenho cognitivo de seus estudantes. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educacion (Reice)*, v. 2, n. 2, p. 83-104. 2004. Disponível em: <<http://www.ice.deusto.es/rinace/reice/vol2n2/Soares.pdf>>. Acesso em: 20 mai. 2008.

_____. O efeito da escola no desempenho cognitivo de seus alunos. In: SOUZA, A. M. (Org.). *Dimensões da avaliação educacional*. Petrópolis: Vozes, 2005, p. 174-204.

SOARES, J. F., CASTRO, C. M.; RIBEIRO, L. M. *O Provão: os cursos A são os que mais oferecem aos estudantes?* Belo Horizonte: UFMG. Texto não publicado, 2000.

SOARES, J. F., RIBEIRO, L.; CASTRO, C. M. Valor agregado de instituições de ensino superior em Minas Gerais para os cursos de Direito, Administração e Engenharia Civil. *Dados*, v. 44, n. 2, p. 363-396, 2004.

URBINA, S. *Essentials of psychological testing*. New Jersey: John Wiley & Sons Inc., 2004.

Ricardo Primi

Doutor em Psicologia Escolar e do Desenvolvimento Humano pela Universidade de São Paulo.

Professor da Universidade São Francisco, Itatiba (SP), Brasil;

ricardo.primi@saofrancisco.edu.br

Marjorie Cristina Rocha da Silva

Doutora em Psicologia na Universidade de São Francisco.

Professora do Centro Universitário FIEO, UNIFIEO, Osasco (SP), Brasil;

silvamarjorie@yahoo.com.br

Daniel Bartholomeu

Mestre e Doutor na área de Avaliação Psicológica na Universidade de São Francisco.

Professor do Centro Universitário Salesiano de São Paulo (Unisal), Americana (SP), Brasil;

d_bartholomeu@yahoo.com.br

Artigo submetido em 07/06/2017

Aprovado em 26/04/2018