

O Provão de Psicologia: Objetivos, Problemas, Conseqüências e Sugestões¹

Ricardo Primi

Universidade São Francisco.

J. Landeira-Fernandez

Pontifícia Universidade Católica do Rio de Janeiro e

Universidade Estácio de Sá

Cílio Ziviani²

Universidade Gama Filho

RESUMO - O Governo Federal, através do Ministério da Educação, criou o Exame Nacional de Cursos (Provão), com objetivo de avaliar os diferentes cursos de graduação e apontar eventuais falhas na formação acadêmica de seus alunos. O primeiro Provão de Psicologia foi realizado em 2000 e desde então vem sendo realizado anualmente por todos os alunos de Psicologia que estão se formando. O propósito do presente trabalho é o de discutir se os dois objetivos dessa forma de avaliação - discriminar os diferentes cursos de Psicologia e apontar falhas na formação de seus alunos - vêm sendo alcançados. Inicialmente apresentamos alguns problemas encontrados no primeiro Provão de Psicologia realizado em 2000, bem como de suas possíveis conseqüências. Finalmente, o trabalho propõe sugestões para a melhora deste sistema de avaliação.

Palavras-chave: validade de instrumento; avaliação educacional.

The Psychology National Examination: Aims, Problems, Consequences and Suggestions

ABSTRACT - The Brazilian Federal Government, through the Ministry of Education, devised the National Course Examination in order to evaluate the several Brazilian undergraduate courses and point out possible academic education deficiencies. The first Psychology National Course Examination was carried in 2000 and since then it has been annually carried out for all Psychology senior students. The purpose of the present work is to argue whether the two main purposes of this form of evaluation - to discriminate the different courses of Psychology and to point out deficiencies in Psychology education - are being reached. Initially we present some problems found in the first Psychology National Course Examination and possible consequences of these problems. Finally, we put forward some suggestions that might improve the purposes of this type of Psychology academic evaluation.

Key words: instrument validity; educational evaluation.

Desde a sua regulamentação, os cursos de Psicologia passam por intensos debates em relação ao processo de ensino e aprendizagem com objetivo de melhorar a sua qualidade e fortalecimento da sua profissionalização. Esses debates ganharam uma nova dimensão com a implementação da nova Lei de Diretrizes e Base (LDB), na qual o modelo antigo que adotava um currículo mínimo foi substituído por um modelo mais aberto, guiado pelas diretrizes curriculares. Calçado nessas diretrizes, o Governo Federal, através do Ministério da Educação (MEC), desenvolveu dois mecanismos de avaliação de todos os cursos de graduação em andamento. Esses dois sistemas são apresentados na figura 1. Um desses sistemas envolve a avaliação das condições de ensino por meio da visita de especialistas ao local onde o curso é ministrado. O projeto didático-pedagógico do curso bem como o suporte logístico (Instalações) e os recursos humanos (Professores) desse projeto são avaliados por essa comissão de especialistas.

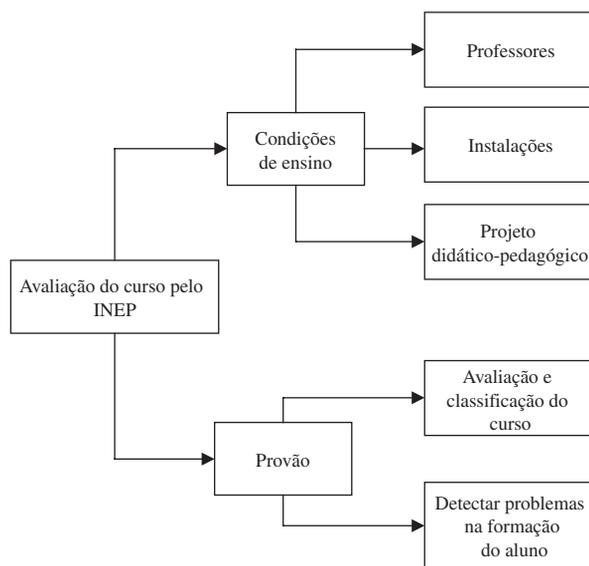


Figura 1. Mecanismos desenvolvidos pelo INEP para avaliação dos cursos de graduação em funcionamento.

Além desse mecanismo, o governo desenvolveu também um sistema com o propósito de avaliar as habilidades e competências específicas que o aluno adquiriu ao longo da sua

1 Desejamos prestar nossos agradecimentos ao INEP, em particular ao Prof. Orlando Pilati, ao Prof. Bernardo Jablonski, pela leitura crítica e sugestões bem como ao CNPq, FAPERJ, FAPESP.
2 Endereço: Laboratório de Cognição Social - Programa de Pós-Graduação em Psicologia - Universidade Gama Filho, Rua Manoel Vitorino, 553 - 20748-900 Rio de Janeiro. E-mail: cilio@cilio.com

formação. Esse sistema de avaliação, denominado Exame Nacional de Cursos (ENC) ou Provão como é popularmente conhecido, foi criado em 1995 por meio da Lei 9.131 e tem como objetivo classificar os diferentes cursos de graduação de acordo com o desempenho dos seus respectivos alunos, bem como detectar eventuais falhas na sua formação de acordo com os princípios expressos nas diretrizes curriculares. O desenvolvimento do Provão de Psicologia é produto de uma comissão de 7 membros designada por Portaria ministerial, cujo objetivo é o de determinar as competências, habilidades e conteúdos específicos que o Provão deve abordar. A partir dessas diretrizes, uma empresa especializada na confecção desse tipo de avaliação (Cesgranrio ou Carlos Chagas) dá início a elaboração de questões específicas que irão compor a avaliação.

Os objetivos do Provão de Psicologia

Como apresentado na figura 2, o Provão de Psicologia, assim como o de qualquer outra área, tem dois grandes objetivos: avaliar e discriminar os diferentes cursos de graduação de Psicologia bem como apontar falhas na formação acadêmica de seus alunos.

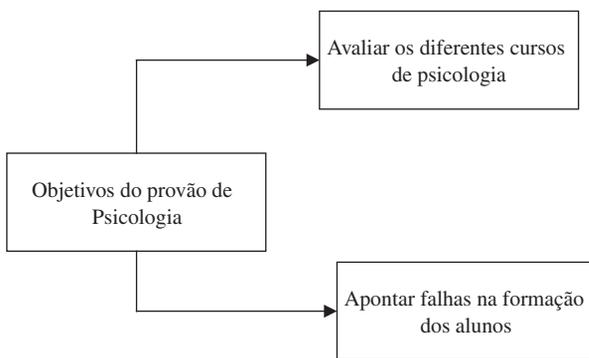


Figura 2. Objetivos do Provão de Psicologia.

O Provão, assim como o Exame Nacional do Ensino Médio (ENEM) e o Sistema de Avaliação do Ensino Básico (SAEB), consistem em ferramentas que permitem diagnosticar de forma relativamente objetiva a realidade acadêmica do país, tornando possível formular ações específicas para a melhora da qualidade do sistema de ensino brasileiro. Em particular, o Provão de Psicologia consiste em um instrumento de avaliação educacional que busca caracterizar o desempenho dos formandos em Psicologia em nível nacional. Reflete, portanto, as conseqüências do ensino superior na área da Psicologia na aprendizagem dos alunos. Permite assim formular ações específicas voltadas para a sua melhoria. Nesse sentido, é extremamente importante que o instrumento seja elaborado de tal forma que seus resultados realmente possam estabelecer avaliações justas e possibilitem a implementação de intervenções mais apropriadas.

O Provão tem sido alvo de críticas dos mais diferentes setores da sociedade. Grande parte dessas críticas ocorre na esfera ideológico-filosófica na ausência de qualquer argumentação de natureza empírica. É nosso entendimento que o conhecimento só avança quando idéias básicas são desafiadas, submetidas a testes empíricos e, dependendo dos

resultados, são fortalecidas ou alteradas. Nesse sentido, o Provão pode ser considerado uma medida positiva no processo de melhoria da qualidade dos cursos de graduação, desde que seus fundamentos possam ser testados empiricamente. A partir daí, problemas com o instrumento de avaliação podem ser detectados e medidas adequadas podem ser tomadas no sentido de melhorar o instrumento.

Dessa forma, o principal fundamento do Provão a ser avaliado é a sua validade. Na psicometria, a validade reporta-se ao conjunto de evidências científicas que atestam a legitimidade das interpretações baseadas nas informações obtidas pelos instrumentos de avaliação (*American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999*). Uma das suposições fundamentais do Provão é a de que o desempenho de um conjunto de alunos de um determinado curso tem a capacidade de refletir a qualidade ensino desse curso. Sob o enfoque da validade, pode-se questionar se o Provão tem sido um instrumento adequado para alcançar este objetivo; ou seja, se as interpretações calcadas no desempenho dos alunos de um curso frente ao Provão podem ser adotadas como parâmetros de qualidade desse curso. Outro fundamento do Provão, que caracteriza o seu segundo objetivo, é o de que o desempenho dos alunos como um todo pode também apontar eventuais falhas na formação acadêmica desses alunos. Neste artigo, apresentamos alguns problemas que o Provão de Psicologia têm encontrado em relação a esses dois objetivos, bem como de possíveis conseqüências que esses problemas podem gerar. Finalmente, é também nossa intenção propor ações que eventualmente possam reduzir esses problemas e assim aumentar a validade ou a interpretação dos resultados em relação ao Provão.

Problemas e Conseqüências Relacionadas com a Avaliação dos Cursos de Psicologia Através do Provão

O Provão Diferencia o Desempenho Entre Calouros?

Recentemente realizamos um trabalho em que o Provão de Psicologia 2000 foi aplicado em calouros provenientes de quatro cursos que tinham obtido os conceitos "A", "B", "C" e "E". O objetivo deste estudo foi o de verificar se o desempenho dos calouros nestes quatro cursos podia ser diferenciado da mesma forma que o desempenho dos formandos desses respectivos cursos. Mais ainda, com base nesses resultados, procuramos verificar também o incremento do desempenho dos formandos em relação aos ingressantes frente ao Provão de Psicologia em cada um desses cursos (Landeira-Fernandez & Primi, 2002).

Dentre as várias habilidades que o Provão pode avaliar destaca-se a sua capacidade de detectar a qualidade da formação acadêmica que um curso agrega a um grupo de alunos ao momento de sua formação acadêmica. Nesse sentido, a comparação do desempenho de calouros e formandos frente ao Provão deveria levar a dois resultados principais: em primeiro lugar, não deveriam existir diferenças entre as médias de desempenho dos calouros de diferentes cursos visto que os calouros ainda não receberam qualquer formação específica em Psicologia. Em segundo lugar, as diferenças só deveriam

aparecer ao final do curso, favorecendo os alunos de cursos de melhor qualidade, refletindo assim a possível influência do curso sobre o aluno.

Em relação à primeira hipótese, observamos que o Provão de Psicologia 2000 (Landeira-Fernandez & Primi, 2002) bem como o de 2001 (Landeira-Fernandez, 2002) é capaz de diferenciar o desempenho dos calouros da mesma forma que o dos formandos. Essas diferenças significativas já entre os calouros indicam que o Provão é sensível às competências e habilidades adquiridas previamente ao ingresso no curso. De fato, o incremento médio nos diferentes cursos ou o “ Δ ” agregado ao aluno (diferença entre a média de desempenho do formando em relação ao calouro) é bem parecido. Por exemplo, no Provão de Psicologia de 2000, observou-se que o incremento médio do curso com conceito “A” foi igual a 18%, no curso “B” igual a 19% no curso “C” igual a 14% e no curso “E” igual a 12%. É interessante notar que neste estudo o curso com conceito “B” sofreu um incremento médio maior que o do curso “A”, o que poderia sugerir que o curso com conceito “B” agrega mais qualidade acadêmica em relação ao curso “A”. Entretanto há o problema do efeito “teto”, isto é, a limitação da variação possível no curso “A” já que o desempenho destes alunos tende a ser o mais alto e mais próximo do valor máximo.

Como os cursos públicos gratuitos têm uma procura muito grande, é possível que a seleção imposta pelo vestibular possa explicar as diferenças no desempenho dos formandos no Provão de Psicologia observada aproximadamente 4 ou 5 anos mais tarde. Embora as habilidades e competências avaliadas pelo vestibular sejam importantes para a formação profissional, elas não constituem, por definição, componentes específicos da formação em Psicologia. Esta formação específica, por sua vez, deve ser adquirida durante o curso e o Provão deveria estar mais associado a este componente específico do que aos aspectos gerais que diferenciam os alunos no momento da entrada no ensino superior. Dessa forma, o Provão deveria sofrer pouca influência dessas competências mais gerais pré-existentes refletindo muito mais os componentes específicos adquiridos no curso de Psicologia.

Um argumento frequentemente levantado em relação a este problema é o fato de que as universidades com baixa demanda deveriam adequar seu sistema de ensino ao perfil dos alunos que entram de tal modo a fazê-lo sair no mesmo nível atingido pelos alunos de cursos com boa qualidade (cursos com “B” e “A”). Conseqüentemente, o fato do desempenho dos formandos nessas universidades ser baixo refletiria uma fraqueza em seu sistema de formação. Embora seja verdade que as universidades precisem adaptar seus métodos e sistemas de ensino à sua clientela, é preciso levar em consideração o trajeto de desenvolvimento da inteligência (Ackermann & Heggstad, 1997; Horn & Noll 1997; Matsunaga & Horn, 2000). Este argumento presume ser possível criar um projeto pedagógico de forma a conseguir um incremento de desempenho médio de 35% que seria necessário para fazer um aluno que ingressa em um curso “E” sair com um desempenho médio de um aluno de curso “A”. Entretanto, ainda que grosseiro, um simples exercício com os dados daquele outro trabalho que investigou o desempenho dos calouros e formandos no Provão de Psicologia 2000 (Landeira-Fernandez & Primi, 2002) sugere que este incremento é muito difícil

de ser alcançado. Em nossa amostra, um aluno ingressante em um curso com conceito “E” teve um desempenho médio igual a 27,9% enquanto que um formando de um curso “A” igual a 62,1%. Se hipoteticamente calouros de um curso “E” frequentassem um curso com conceito “A” e sofressem o incremento médio encontrado neste curso (18%), então ao final do curso o desempenho desses alunos frente ao Provão seria igual a 45,9%, indicando que esse desempenho estaria um pouco abaixo do desempenho dos alunos formandos em um curso “C” (48,5%) e não muito diferente desempenho que eles já obtêm em seu curso original (40,4%).

Em conclusão, pode-se dizer que o resultado no Provão reflete tanto a influência do curso quanto as diferenças de perfis dos alunos que ingressam em determinada instituição. Muito provavelmente, há um viés presente nas notas do Provão beneficiando os alunos que ingressaram em cursos mais competitivos com vestibulares com alta taxa candidato-vaga. No entanto, a impressão que se tem é a de que o conceito do Provão indica exclusivamente a qualidade do curso, desconsiderando o segundo aspecto ligado às diferenças pré-existentes antes que o curso possa exercer qualquer influência.

A Média do Desempenho dos Alunos é um Bom Representante do Curso?

Uma outra questão importante refere-se às limitações da utilização da média como representante dos cursos. O procedimento adotado até o ano 2000 consistiu em calcular as médias dos alunos de um determinado curso, ordenar as médias dos cursos e atribuir o conceito “E” a 12% dos cursos com as médias mais baixas, “D” aos 18% cursos seguintes, “C” aos 40% seguintes, “B” aos próximos 18% e “A” a 12% dos cursos com as médias mais altas. O procedimento iniciado em 2001 padroniza estas médias e atribui conceito “E” aos cursos cujas médias estiverem abaixo de um desvio padrão da média geral, “D” aos cursos entre o intervalo -1,0 e -0,5 desvios padrão, “C” aos cursos que estiverem na faixa de -0,5 a 0,5, “B” aos cursos que estiverem na faixa 0,5 a 1,0 e “A” aos cursos cuja média estiver acima de um desvio padrão. Este novo procedimento eliminou a rigidez antes praticada, que definia *a priori* a proporção de cursos que iriam receber os cinco conceitos, permitindo com que essas proporções fossem definidas pela distribuição natural das médias. Entretanto, como geralmente distribuições de médias aproximam-se da distribuição normal, pode-se prever que se terá aproximadamente 16% de cursos com conceito “E”, 14% de cursos “D”, 39% de cursos “C”, 14% de cursos “B” e 16% de cursos “A” (a esse respeito veja o teorema da tendência central, Howell, 1997). Evidentemente, este procedimento é melhor que o anterior, pois se as distribuições forem assimétricas negativas poderão existir mais cursos com conceitos acima da média na faixa “B” e “A” do que cursos na faixa “E” e “D”, fato que nunca poderia ocorrer se fosse empregado o sistema antigo.

O problema principal deste procedimento consiste na tentativa de se caracterizar os alunos de um determinado curso em função de uma única nota: a média que estes alunos obtêm no Provão. A justificativa para sua utilização é estatística, já que em distribuições normais, a média é o valor mais próximo da tendência dominante em um grupo. No entanto a média

sozinha não é suficiente. A estatística ensina que para caracterizar uma variável, é necessário observar não só medidas de tendência central (média, moda e mediana), mas também medidas de dispersão (desvio-padrão e variância), e também aspectos da distribuição como a assimetria e curtose. Ainda que todas essas informações sejam utilizadas, é preciso lembrar que elas servem para o entendimento do comportamento das variáveis, mas não para caracterizar os indivíduos que compõem o grupo como é subentendido nas interpretações dos resultados do Provão. Nele, uma propriedade abstrata da distribuição é utilizada para caracterizar os indivíduos que compõem um grupo. Dessa forma, seria uma prática justa caracterizar todos os alunos de um determinado curso em relação ao conceito que esse curso obteve no Provão?

A resposta para esta pergunta parece ser que não. Essa prática pode ser altamente injusta, especialmente para os cursos que obtêm conceitos “C” no Provão. De modo geral, as pessoas tendem a generalizar o conceito de um curso correspondente à média do grupo para todos os alunos daquele curso, de tal forma que um determinado aluno formado em um curso “C” pode ser tomado como mediano e certamente como “mais fraco” academicamente em relação a um outro aluno formado em um curso com conceito “A”. Baseando-se nesta interpretação errônea, algumas empresas fazem uma triagem dos candidatos que concorrem a uma vaga no mercado de trabalho, aceitando somente aqueles que se formaram em cursos cujo conceito foi “A” ou “B”. Na realidade, o que realmente ocorre é que os cursos com conceitos “C” são muito heterogêneos. Há alunos em ambos os extremos, que fazem com que os cursos venham a obter uma nota média. No entanto, é possível que cursos com o mesmo desempenho possam apresentar substanciais diferenças entre os seus alunos.

Como ilustração desse argumento, apresentamos na tabela 1 uma análise comparativa de um curso que obteve conceito “C” no Provão de Psicologia 2000 (curso 1) comparando o desempenho dos alunos com os de outros três cursos que obtiveram respectivamente conceitos “C” (curso 2), “B” (cursos 3 e 4). Na primeira, quarta e sétima linhas da tabela 1 são apresentadas as porcentagens de alunos do curso 1 que obtiveram desempenho em cada uma das quatro faixas delimitadas pelos percentis 25, 50 e 75 (quartis). É possível observar que nessa instituição, 16,9% dos seus alunos estão

entre aqueles com desempenho mais baixo no Provão, 30,5% com desempenho médio baixo, 29,9 com desempenho médio alto e 22,7% entre os 25% com desempenho mais alto. Nas segunda, quinta e oitava linhas são apresentadas as mesmas proporções para os alunos do curso 2, também com conceito “C”, e para os alunos dos cursos 3 e 4 ambos com conceito “B”. Nas linhas seguintes apresentam-se as diferenças entre essas porcentagens.

Nas sétima e oitava colunas da tabela 1 são apresentados os Qui Quadrados associados a estas discrepâncias nas porcentagens, e se elas são estatisticamente significativas (coluna relacionada com valor de p). É interessante notar que os cursos 1 e 2, ambos com conceitos “C”, apresentaram diferenças significativas. Por outro lado, cursos 3 e 4, ambos com conceitos “B”, apresentaram as mesmas faixas de desempenho (ausência de diferenças significativas) em relação ao curso 1 que obteve conceito “C”. Isto indica que se considerarmos a distribuição de alunos nas quatro faixas, os alunos do curso 1 com conceito “C”, são mais parecidos com os alunos dos cursos 3 e 4 que obtiveram conceito “B”, do que com os alunos do curso 2 com conceito “C”. Embora os cursos 3 e 4 tenham apresentado uma proporção maior de alunos na faixa 1 e menor na faixa 4, essa diferença não pode ser generalizada na afirmação que os alunos dessas universidades vêm de populações diferentes, como se supõe a partir de seus conceitos. Se fossem divulgadas as médias dos cursos, essas diferenças ficariam mais evidentes. No entanto, como justifica o INEP (2000):

Desde o primeiro Exame, foi evitada a idéia de simplesmente apresentar as médias de cada um dos cursos, para que não fossem estimuladas as comparações individuais, mas sim, evidenciada a diferença entre grupos de instituições. Desta forma, a opção por atribuir conceitos pareceu a mais adequada, separando-se os cursos em grupos homogêneos segundo seus desempenhos. (p. 2).

A análise apresentada na tabela 1 contradiz diretamente a idéia de grupos homogêneos. Essa análise ilustra uma situação de grande heterogeneidade entre os alunos de dois cursos que receberam o mesmo conceito. Além disso, a análise mostra que um grupo de alunos de um curso com conceito “C” foi mais parecido com grupos de alunos de cursos com conceito

Tabela 1. Uma comparação da distribuição de alunos de um curso C com outros três cursos um “C” e dois “B” em função da distribuição de alunos em quatro faixas de desempenho.

Curso	Conceito	Faixas de Desempenho - Quartis				Qui Quadrado	p
		1 (1-25)	2 (26-50)	3 (51-75)	4 (76-100)		
Curso 1	C	16.9	30.5	29.9	22.7	8.72	0.0332
Curso 2	C	30.9	31.9	26.6	10.6		
Diferença		-14	-1.4	3.3	12.1		
Curso 1	C	16.9	30.5	29.9	22.7	4.75	0.1908
Curso 3	B	9.8	25.4	30.3	34.4		
Diferença		-7.1	5.1	-0.4	-11.7		
Curso 1	C	16.9	30.5	29.9	22.7	2.64	0.4505
Curso 4	B	13.6	24.3	30.1	32		
Diferença		3.3	6.2	-0.2	-9.3		

“B”. Como conseqüência, generalizações simplistas, como, por exemplo, afirmar que cursos “C” formaram alunos “mais fracos”, cometem injustiça em relação àqueles alunos que se formaram nessas instituições com desempenho entre os 25% melhores de todos os que se formaram nesse período.

As Diferenças Entre os Cursos são Muito Grandes?

Uma das principais características da medida é a sua capacidade de discriminação. Nesse sentido, o desempenho dos alunos de psicologia nas questões discursivas dos Provões de 2000 e 2001 tem sido extremamente fraco, girando em torno de 25% de acerto, caracterizando assim a falta de expressividade psicométrica dessa parte do Provão no processo de diferenciação dos diferentes cursos. Por outro lado, o desempenho dos alunos nas questões de múltipla escolha tem sido em torno de 50% de acerto, caracterizando assim a máxima variabilidade que o instrumento pode oferecer. Nesse sentido, é importante destacar que a diferença entre os 5 conceitos (“A”, “B”, “C”, “D” e “E”) é muito pequena. Considerando-se o desempenho dos formandos que prestaram o Provão no ano de 2000 observou-se, por exemplo, que uma instituição com conceito “A” obteve média 22,9 (DP=4,5), outra com conceito “B” 21,2 (DP=3,9) e uma terceira com conceito “C” obteve 20,0 (DP=4,5) nas questões de múltipla escolha. Esses dados mostram claramente que pequenas diferenças têm a capacidade de atribuir conceitos distintos entre os cursos. Uma conseqüência dessa estreita margem no desempenho entre cursos com diferentes conceitos é o fato de que pode haver um favorecimento no conceito dos cursos que concentrem professores responsáveis pela formulação dessas questões. Uma sugestão que poderia ser levantada para evitar esse possível viés seria a de não incluir na prova mais de uma questão oriunda de professores de um mesmo curso.

Outra conseqüência que a pequena margem entre cursos com diferentes conceitos pode acarretar é o favorecimento de cursos com ênfases curriculares em áreas que foram mais privilegiadas na prova. Por exemplo, no Provão de 2000 houve uma concentração de questões na área de Psicologia comunitária. Esse fato pode aumentar artificialmente o conceito de cursos que privilegiem essa área.

Problemas Relacionados com o Diagnóstico de Deficiências Curriculares: O Provão de Psicologia é Capaz de Apontar Falhas na Formação Acadêmica do Aluno?

Além de avaliar e atribuir conceitos aos diferentes cursos de graduação, o Provão também tem o como objetivo detectar eventuais falhas na formação acadêmica dos alunos que estão terminando o curso. Em um estudo recente comparando o desempenho dos calouros e dos formandos no primeiro Provão de Psicologia (Landeira-Fernandez & Primi, 2002), observou-se um baixo incremento no desempenho dos formandos em relação aos calouros nas questões que abordaram conteúdos de natureza básica. Por outro lado, questões de natureza aplicada apresentaram maiores incrementos. Isso também foi observado por Pasquali (2002) na análise dos dados dos Provões de 2000 e 2001. Esse melhor desempenho dos formandos nas questões abordando conteúdos aplicados

está de acordo com relatos da literatura que apontam para o fato de que os cursos de Psicologia têm uma grande orientação para a área aplicada (Biaggio, 2000; Nunes, 1993).

Dentre as várias habilidades e competências adquiridas ao longo da formação e profissionalização do psicólogo, Hutz e Adair (1996) observaram, através do padrão de citação em periódicos brasileiros, uma ênfase no pensamento artístico concomitante à desvalorização do pensamento científico. Dessa forma, é possível que o mau desempenho dos formandos em questões de conteúdo básico no Provão de Psicologia 2000 esteja relacionado à falha dos cursos no processo de ensino-aprendizagem desse tipo de conhecimento ou a eventuais deficiências na associação desses conteúdos a disciplinas de caráter mais prático ou profissionalizante.

No entanto, deve-se observar que o baixo incremento dos formandos em relação aos calouros está relacionado a um baixo desempenho dos calouros em questões que abordaram este tipo de conteúdo. Dessa forma, é possível que essas questões tenham abordado conteúdos extremamente específicos, tendo-se assim uma falsa impressão dessas deficiências.

Outro problema que reduz a possibilidade do diagnóstico de deficiências curriculares refere-se às questões mal formuladas. Dentre os três Provões de Psicologia já realizados, 3 questões, uma em cada Provão, foram anuladas: questão 13 do Provão 2000, questão 25 do Provão 2001 e a questão 14 do Provão de 2002. A questão 25 do Provão de 2001, por exemplo, aplicou equivocadamente os conceitos de precisão e validade enquanto que a questão 14 do Provão 2002 apresentou uma interpretação equivocada do teste WISC. Estas duas últimas questões foram amplamente debatidas em um grupo de discussão (*avalpsi@yahoogrupos.com.br*) composto, atualmente, por mais de 236 pessoas (parte delas associadas ao Instituto Brasileiro de Avaliação Psicológica – IBAP), incluindo professores, pesquisadores, profissionais, alunos de pós-graduação e alunos de graduação interessados em avaliação psicológica. O resultado destas discussões originou documentos que foram levados ao INEP por um dos autores do presente artigo que participa da Comissão de Psicologia, propondo a anulação dessas questões (cópias destes documentos estão disponíveis no site do IBAP: <http://www.ibapnet.org.br>).

Além destas, existem outras questões problemáticas nos Provões de Psicologia realizados em 2000 e 2001. Por exemplo, Ziviani e Primi (2002) aplicaram o modelo de Rasch (modelo de um parâmetro da Teoria de Resposta ao Item) para analisar os itens das provas de 2000 e 2001 com base no desempenho de todos os formandos que realizaram as provas. Esta análise identificou seis itens com problemas de clareza e ambigüidade na resposta certa. Estes problemas têm como conseqüência uma redução na validade de conteúdo em decorrência do decréscimo do número de itens que efetivamente podem cumprir sua função na avaliação. Além disso, a diminuição do número de itens válidos leva também a uma redução da fidedignidade ou precisão (consistência interna) da prova como um todo.

Algumas Sugestões para Melhoria do Provão de Psicologia

Os problemas levantados neste trabalho suscitam reflexões no sentido de buscar ações que possam levar a um

aumento da validade na construção e interpretação dos resultados de futuros Provões. Nesta seção, apresentamos algumas propostas específicas que buscam aliviar algumas das questões levantadas anteriormente.

Com relação ao fato de as questões discriminarem o desempenho dos calouros, um procedimento que poderia reduzir este viés seria o de formular questões que fossem mais ricas em conteúdos técnicos específicos ao curso de Psicologia. Dessa forma, é possível impedir que estas questões sofram influência de outras fontes de diferenças individuais não diretamente relacionadas ao domínio do conteúdo específico de Psicologia. Em um trabalho anterior (Landeira-Fernandez & Primi, 2002), observamos que 20 questões do Provão de Psicologia 2000 não discriminaram o desempenho dos calouros, sendo, portanto, questões que abordaram conteúdos específicos adquiridos ao longo do curso de Psicologia.

Entretanto, para que isto ocorra, é necessária uma ampla e exaustiva discussão sobre o conteúdo básico comum a todos os cursos de Psicologia. Um trabalho como este deverá enfrentar o desafio de equilibrar a generalidade com a especificidade na composição dos itens que, por um lado, deveriam ser suficientemente específicos e não discriminarem calouros, mas por outro lado, abordando conteúdos gerais ensinados em todos os cursos de Psicologia.

Esta sugestão conduz a uma outra discussão um pouco mais ampla em relação a diferentes tipos de sistema de avaliação educacional: sistema de referência à norma e sistema de referência ao domínio (Anastasi & Urbina, 2000). O Provão utiliza o sistema de referência à norma, onde o conceito de um determinado curso é obtido por meio da comparação com os escores obtidos por todos os outros cursos. Esse procedimento busca a maior variabilidade possível entre os diferentes cursos para que se possa, então, ordená-los hierarquicamente em postos (Bond, 1996; Stiggins, 1994). Esse sistema é capaz de atribuir conceitos através da posição relativa de um determinado curso em relação aos demais sem levar em consideração se determinados conteúdos específicos à formação do psicólogo foram ou não adquiridos pelos formandos que prestaram o Provão. Por outro lado, um sistema referenciado ao domínio, ou ao critério como às vezes é chamado, compararia o desempenho dos alunos de um curso em relação a um conjunto de conteúdos básicos, indicando o domínio dos alunos em relação a esses conteúdos, sem levar em consideração a posição relativa de um curso face aos demais. Basicamente o procedimento de referência à norma estabelece como parâmetro de comparação o desempenho do próprio do grupo, seja ele adequado ou não. Já no procedimento de referência ao critério, o parâmetro de comparação é o próprio conteúdo que se espera que os alunos tenham adquirido ao longo do curso. Num sistema referenciado ao domínio, um curso poderia ser avaliado através da proporção dos formandos que prestaram o Provão que foram capazes de atingir um determinado critério de competência, como por exemplo, 75% de domínio nos conteúdos avaliados.

Este segundo sistema faz muito mais sentido em avaliações educacionais. É muito mais relevante saber qual a proporção de alunos que um curso consegue formar com nível razoável de competência em relação a um conjunto pré-definido de conhecimentos adquiridos ao longo da formação,

do que saber se os formandos de um determinado curso têm em média desempenho melhor ou pior do que os formandos de outros cursos. Utilizar exclusivamente a referência à norma pode trazer resultados enganosos, já que mesmo os cursos com conceito “A” podem não ter demonstrado domínio de certos conteúdos que supostamente deveriam ter sido adquiridos.

Por outro lado, o grande problema com um sistema referenciado ao domínio está na complexidade de sua implementação, já que depende de uma definição muito precisa dos conteúdos que serão avaliados bem como de uma operacionalização arrojada desses conteúdos. Nesse sentido, Anastasi e Urbina (2000) observaram que este sistema de avaliação é mais facilmente implementado na avaliação de níveis mais básicos de aprendizagem, nos quais os conteúdos são mais facilmente definidos. Em níveis mais complexos há uma maior flexibilidade de conteúdos, o que dificultaria a definição de um conjunto básico de conceitos.

Embora um sistema de avaliação referenciado ao domínio seja muito mais difícil de ser desenvolvido, seria o mais adequado ao cumprimento de um dos objetivos do Provão, qual seja o de apontar falhas na formação dos alunos – ao nosso ver, uma das suas principais funções. Além disso, fomentaria uma maior discussão sobre a formação do Psicólogo, uma vez que seria necessário definir mais precisamente os conteúdos básicos propostos nas diretrizes curriculares e sua subsequente operacionalização em um instrumento de avaliação.

Uma avaliação deste tipo seria a que mais se aproximaria do ideal da avaliação, uma vez que ela poderia demonstrar os pontos fortes e fracos dos cursos na formação de seus alunos. Ainda, a referência ao domínio se afasta do modelo competitivo de diferenças individuais implícito na referência à norma, e volta-se para um modelo mais próximo à idéia de desenvolvimento de competências.

Uma sugestão para o desenvolvimento de um sistema referenciado ao domínio seria a análise do conteúdo dos itens dos Provões de Psicologia, bem como análises fatoriais desses itens, buscando entender quais agrupamentos de conteúdos são formados pelas provas já aplicadas, como foi feito por Pasquali (2002). A Teoria de Resposta ao Item pode ser empregada associando níveis de desempenho geral à performance nos itens, indicando para cada nível de desempenho quais conteúdos foram dominados (Primi & Almeida, 2001; Primi, 2002; Santos, Primi, Taxa & Vendramini, 2002). Uma vez definidos níveis de desempenho associados ao domínio de conteúdos, os cursos poderiam ser caracterizados em razão da proporção de formandos em cada nível de desempenho.

Como discutido anteriormente, a média utilizada como índice representante do desempenho dos cursos leva a situações problemáticas. Uma solução imediata seria a de deixar de atribuir um conceito global ao curso. Ao invés disso, o mesmo procedimento de atribuição de conceito aos cursos poderia ser aplicado aos alunos, independentemente do curso ao qual pertencam. A distribuição dos alunos em cada um dos cinco conceitos poderia, então, ser utilizada para a avaliação dos cursos. Neste caso, a identidade dos alunos continuaria preservada e teríamos uma visão mais realista dos cursos por meio da proporção de alunos com desempenhos “A”, “B”, “C”, “D” e “E” formados em cada curso. Neste esquema, os cursos poderiam ser ordenados através da proporção do

número de alunos que atingiram cada um desses cinco conceitos. Além disso, com o emprego da Teoria de Resposta ao Item, particularmente do modelo de Rasch, poderiam ser criados mapas de questões que indicam, para cada um dos cinco conceitos, quais conteúdos que foram dominados. Esses mapas fornecem informação muito útil aos cursos (Primi & Almeida, 2001; Ziviani & Primi, 2002).

Com relação ao problema das pequenas diferenças entre os cursos, uma forma de evitá-lo seria o da distribuição do número de questões da prova de forma proporcional aos diferentes conteúdos a serem avaliados. Dada a riqueza e diversidade da área da Psicologia, seria desaconselhável uma prova conter duas questões que abordem o mesmo conteúdo, assunto ou autor.

Quanto ao problema das questões mal formuladas, talvez uma solução será tornar seu processo de criação mais transparente. Nesse sentido, a preservação do sigilo absoluto acerca do autor ou da autora de questões que venham a compor a forma final do Provão deve ser mantida, não apenas durante a elaboração da prova, mas também no período que antecede a publicação final dos resultados. Entretanto, após essa publicação, a divulgação do autor ou da autora de cada questão estabelecerá a relação entre o juízo crítico da comunidade acadêmica acerca da questão ao seu autor ou a sua autora. Tal procedimento oferece as condições essenciais para a autoregulação de Provões subseqüentes, uma vez que estarão presentes os elementos característicos de um sistema orientado pelas contingências públicas do seu próprio comportamento. A abertura de sigilo da autoria das questões pode também desvelar a eventual concentração de itens oriundos de professores de uma mesma região ou instituição de ensino, que como discutido anteriormente, pode ser fonte de viés no processo de avaliação.

A tabela 2 apresenta um resumo dos dois grandes objetivos do Provão de Psicologia, possíveis problemas associados

a esses objetivos, conseqüências geradas por esses problemas e possíveis sugestões que podem aliviar esses problemas e suas conseqüências. Esperamos que estas reflexões possam contribuir para a melhoria deste importante sistema de avaliação do ensino superior.

Referências

- Ackermann, P.L., & Heggstad, E.D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, 121(2), 219-245.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association
- Anastasi, A. & Urbina, S. (2000). *Testagem Psicológica*. Porto Alegre: Artes Médicas.
- Biaggio, A.M.B. (2000). Brazil. Em *Encyclopedia of Psychology* (Vol 1, pp. 464-466). New York: Oxford University Press.
- Bond, L.A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research & Evaluation*, 5(2). Disponível online: <http://ericae.net/pare/getvn.asp?v=5&n=2>. Acesso: 2/07/2002.
- Horn, J.L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53-91). New York: Guilford.
- Howell, D.C. (1997) *Statistical methods for psychology*. Boston: Duxbury Press
- Hutz, C. & Adair, J. (1996). The use of references in Brazilian psychology journals reveals trends in thought and research. *International Journal of Psychology*, 31, 145-149.
- Instituto Nacional de Estudos e Pesquisas Educacionais (2001). *Os Novos Critérios de Avaliação do Exame Nacional de Cursos*.

Tabela 2. Objetivos do Provão, problemas, conseqüências e sugestões para melhoria.

Objetivo do provão	Problemas	Conseqüências	Sugestões
Avaliar os diferentes cursos de Psicologia	Questões discriminam desempenho entre calouros	Viés favorecendo cursos com altataxa candidato/vaga	Formular questões contendo conteúdo específico à área de Psicologia Mudar o sistema da avaliação com referência à norma para o sistema de referência ao domínio
	Média como representante dos cursos	Viés desfavorecendo subgrupos de alunos em cursos heterogêneos com conceito "C"	Não atribuir um conceito global ao curso mas sim a proporção do número de alunos de um determinado curso que atingiu determinados critérios.
	Diferença entre os 5 conceitos (A,B,C,D e E) muito pequena. Média de duas questões pode alterar conceito do curso.	Viés favorecendo cursos com ênfases curriculares nas questões que foram mais abordadas no provão Viés favorecendo cursos em que mais de um professor formulou questões para o provão	Distribuir o número de questões proporcionalmente com os diferentes conteúdos a serem avaliados. Evitar duas questões sobre o mesmo assunto Distribuir de forma aleatória os professores que formulam as questões. Desaconselhar mais de 1 professor por curso
	Questões mal formuladas	Índice de acerto da questão muito pequeno ameaçando a validade e fidedignidade do provão como um todo	Divulgação da autoria das questões após a divulgação dos resultados
Apontar falhas na formação dos alunos	Questões com conteúdo psicológico muito geral	Redução do número de questões que permitam avaliar conteúdos específicos de Psicologia	Ajustar a profundidade da questão, englobando conteúdos comuns ministrados entre os diferentes cursos de Psicologia

- [On Line]. Disponível: http://www.inep.gov.br/download/noticias/2001/criterios_prova02001.doc
- Ladeira-Fernandez, J. (2002). Comparação do desempenho entre calouros e formandos no Provão: evidências de falhas na formação. *Trabalho apresentado no I Congresso Brasileiro de Psicologia: Ciência e Profissão*. São Paulo, SP.
- Ladeira-Fernandez, J. & Primi, R. (2002). Comparação do desempenho entre calouros e formandos no Provão de Psicologia 2000. *Psicologia: Reflexão e Crítica*, 15(1), 219-234.
- Matsunaga, H. & Horn, J. (2000). Characterizing mature human intelligence expertise development. *Learning and individual differences*, 12, 5-53.
- Nunes, T. (1993). Psychology in Latin America: The case of Brazil. *Psychology and Developing Societies*, 5, 123-133.
- Pasquali, L. (2002). Provão (ENC) de Psicologia 2000 e 2001: Análise dos parâmetros psicométricos. Em: R. Primi (org.). *Temas em Avaliação Psicológica* (pp. 152-178). Campinas: IDB Digital / Instituto Brasileiro de Avaliação Psicológica (IBAP).
- Primi, R. (2002). Contribution to the Development of a Fluid Intelligence Scale. Manuscrito submetido para publicação.
- Primi, R. & Almeida, L.S. (2001). Teoria de Resposta ao Item. Em: E.M. Fernandes, & L.S. Almeida (Orgs.) (pp. 205-232). *Métodos e técnicas de avaliação: contributo para a prática e investigação psicológicas*. Braga, Portugal: Centro de Estudos em Educação e Psicologia, Universidade do Minho.
- Santos, A.A.A.; Primi, R., Taxa, F.O.S. & Vendramini, C.M.M. (2002) O Teste de Cloze na avaliação da compreensão em leitura. *Psicologia: Reflexão e Crítica*. 15(3).
- Stiggins, R.J. (1994). *Student-Centered Classroom Assessment*. New York: Merrill
- Ziviani, C. & Primi, R. (2002). Teoria de Resposta ao Item e o modelo de Rasch de mensuração: uma análise do Provão de Psicologia. Em: R. Primi (org.). *Temas em Avaliação Psicológica* (pp. 131-151). Campinas: IDB Digital / Instituto Brasileiro de Avaliação Psicológica (IBAP).

Recebido em 06.12.2002

Primeira decisão editorial em 11.05.2003

Versão final em 10.06.2003

Aceito em 01.07.2003 ■