

A PROVA DO ENADE DE PSICOLOGIA 2006: CONCEPÇÃO, CONSTRUÇÃO E ANÁLISE PSICOMÉTRICA DA PROVA

Ricardo Primi¹ – Universidade São Francisco (LabAPE, USF)
Cláudio S. Hutz – Universidade Federal do Rio Grande do Sul (UFRGS)
Marjorie Cristina Rocha da Silva – Universidade São Francisco (LabAPE, USF)

RESUMO

O Exame Nacional de Desempenho dos Estudantes (ENADE) é elaborado com o objetivo de avaliar as habilidades acadêmicas e competências profissionais desenvolvidas pelos estudantes de Psicologia ao longo de sua trajetória no ensino superior, bem como obter informações sobre suas características socioeconômicas. As informações do ENADE são usadas, em última instância, como parte dos conceitos dos cursos e instituições. Este trabalho apresenta a análise da prova de 2006, respondida por 26613 estudantes ingressantes e concluintes de psicologia. Apresentam-se os procedimentos de construção da prova, uma análise psicométrica empregando a análise fatorial dos itens por informação completa e calibração dos parâmetros dos itens empregando o modelo Rasch e de créditos parciais (para as questões dissertativas). Apresenta-se uma análise dos mapas de itens para se estabelecer referências de interpretação das notas, que permitiu realizar uma caracterização das competências e habilidades dos estudantes pesquisados comparando-se o desempenho dos concluintes em relação aos ingressantes.

Palavras-chave: Avaliação do ensino superior; Modelo de Rasch; Competências; Habilidades.

THE ENADE PSYCHOLOGY EXAM: CONCEPTION, CONSTRUCTION AND PSYCHOMETRIC ANALYSIS

ABSTRACT

The National Exam of Student Performance (ENADE) is a test conducted with the objective of assessing the academic abilities and professional competences developed by psychology students throughout their years of study in higher education as well as information about their socioeconomic characteristics. The information from ENADE is used ultimately as part of a system that evaluates undergraduate programs and institutions. This paper presents an analysis of the 2006 exam, which was answered by 26,613 freshmen and graduating students in psychology. The test construction procedures are presented as well as a psychometric analysis of the exam based on full information item factor analysis and item calibration using the Rasch and Partial Credit (for essay questions). An analysis of item maps is presented to establish reference points for scale interpretation that allowed a characterization of the abilities and competences of students comparing the performance of freshmen with graduating students.

Keywords: Assessment of higher education; Rasch Model; Skills; Abilities.

A avaliação tem um papel central na formulação e implementação de políticas públicas e, assim sendo, os sistemas de avaliação em larga escala buscam levantar informações sobre a eficiência e qualidade das organizações que provêm bens públicos fundamentais à população, tais como saúde, educação e segurança. Essas informações são essenciais para a gestão dos recursos públicos uma

vez que clarificam possíveis modos de ações interventivas e regulatórias criadas para melhorar a qualidade do sistema.

É fundamental produzir indicadores para o controle de qualidade das instituições de ensino. A medida da qualidade das instituições pode ser definida por indicadores do quanto cada uma contribui para o desenvolvimento de habilidades acadêmicas, competências profissionais e incremento do conhecimento de seus alunos. Em sociedades democráticas, tais avaliações servem como instrumentos de prestação de contas (*accountability*), utilizados para examinar se as organizações, às quais são determinados alguns papéis, têm cumprido suas obrigações.

O Sistema Nacional de Avaliação do Ensino Superior (SINAES), atual sistema nacional de avaliação das Instituições de Educação Superior (IES), compreende três dimensões: avaliação da instituição, avaliação dos cursos de graduação e o

¹ Agência financiadora: Este trabalho foi produzido a partir do financiamento do Edital Observatório da Educação CAPES/INEP e do CNPq.

Correspondências: Ricardo Primi, Universidade São Francisco, Laboratório de Avaliação Psicológica e Educacional (LabAPE), Mestrado e Doutorado em Psicologia, Rua Alexandre Rodrigues Barbosa, 45, CEP 13251-900, Itatiba, São Paulo.

E-mail: rprimi@mac.com

Exame Nacional de Avaliação do Estudante (ENADE). Dessa forma, pretende-se olhar a Instituição a partir de uma perspectiva ampla (avaliação institucional), de um ponto de vista mais específico o curso (avaliação do curso) e, com um foco ainda mais reduzido, o estudante, por meio de seu desempenho no ENADE (Limana & Brito, 2005).

De acordo com a legislação em vigor (Lei nº 10.861 de 14 de Abril de 2004, art. 5º e seus respectivos parágrafos), o ENADE tem por objetivo aferir:

o desempenho dos estudantes em relação aos conteúdos programáticos previstos nas diretrizes curriculares dos respectivos cursos de graduação, suas habilidades para ajustamento às exigências decorrentes da evolução do conhecimento e suas competências para compreender temas exteriores ao âmbito específicos de sua profissão ligada à realidade brasileira e mundial e a outras áreas do conhecimento.

Portanto, o conjunto de habilidades acadêmicas, competências profissionais e conhecimentos que devem ser avaliados pelo ENADE são aqueles definidos nas diretrizes

curriculares de cada área. Consequentemente, na construção do ENADE tenta-se basicamente operacionalizar a medida dos construtos definidos nas diretrizes de cada curso. No processo de formulação do exame é definida uma comissão assessora da área que tem a atribuição de criar as recomendações necessárias para a banca de elaboração da prova em termos de diretrizes, objetivos e especificações do exame; outras especificações necessárias à Avaliação in loco dos Cursos de Graduação (ACG) em Psicologia; e elaborar os produtos resultantes dos processos de construção do ENADE e da ACG.

Inicialmente, a comissão assessora discute as diretrizes para elaboração da prova definindo uma matriz de especificação que descreve detalhadamente os objetivos, as questões em termos de conteúdos, habilidades acadêmicas e competências profissionais a serem abordadas pelos itens, a dificuldade esperada e o formato de cada item. Tendo definido esses pontos é publicada uma portaria com as especificações gerais da prova. Posteriormente, o perfil do coordenador e da banca de elaboração da prova é definido. Essas especificações são apresentadas na Tabela 1.

Tabela 1. Perfil do Coordenador e dos Membros da Comissão de Elaboração de Provas para a área de Psicologia ENADE-2006.

O coordenador da Comissão deverá apresentar os seguintes requisitos:

- a) ter o título de Doutor;
- b) conhecer o curso de Graduação em Psicologia e estar integrado ao funcionamento e estrutura do curso de Graduação em Psicologia;
- c) ter participado de comissões acadêmicas e científicas como, por exemplo, comissões de agências de fomento, de organização eventos científicos, órgãos colegiados, conselhos etc.;
- d) produção científica qualificada e atual na área;
- e) experiência em coordenação de grupos;
- f) preferencialmente com experiência em avaliação educacional.

Critérios para a composição dos membros da banca de elaboração da prova

- a) ter o título de Doutor;
 - b) ser Psicólogo, para os casos da elaboração de questões envolvendo práticas profissionais;
 - c) ser professor em cursos de graduação em Psicologia, com experiência mínima de 5 anos;
 - d) apresentar produção científica sistemática, recente e atualizada na área de Psicologia em que elaborará as questões;
- E, além disso, deve-se compor uma banca:
- e) oriunda de diferentes regiões geográficas do Brasil;
 - f) representando a diversidade metodológica, teórica e epistemológica da Psicologia;
 - g) não ser composta por dois membros ou mais de um mesmo curso.

Recomendações gerais na elaboração dos itens:

- a) Evitar apresentar conteúdos muito específicos a determinadas abordagens;
 - b) Possibilitar aos candidatos o estabelecimento de relações, garantindo as dimensões de ciência e profissão;
 - c) Vincularem-se ao núcleo comum de formação das Diretrizes Curriculares em Psicologia, garantindo o conjunto de competências e habilidades previstas na Portaria;
 - d) Contemplar as diversas abordagens em Psicologia, garantindo a pluralidade de referenciais teóricos;
 - e) Contemplar a tensão entre concepções clássicas e contemporâneas em Psicologia;
 - f) Apresentar exemplos de aplicação de conceitos psicológicos que se relacionem com aspectos da realidade brasileira.
-

Após a definição das diretrizes da prova, discute-se a matriz de especificações da prova com o coordenador da banca de elaboração contratado pela empresa licitada para sua elaboração. Nesses encontros são apresentados alguns itens como exemplo para serem discutidos com a comissão, mas que não irão posteriormente à prova. Terminadas as rodadas de discussão, a banca elabora a prova em caráter sigiloso, isto é, a comissão passa a não ter mais influência e conhecimento na elaboração dos itens da prova que, a partir desse ponto, fica totalmente a cargo da empresa licitada.

As provas do ENADE, em sua versão final, são compostas por 30 questões de componente específico e 10 questões de formação geral, idênticas para todas as áreas, totalizando 40 questões. Das questões específicas, 26 são em formato de múltipla escolha e quatro em formato de respostas dissertativas.

Pressupostos das dimensões avaliadas pela prova

O ENADE é formulado com a pretensão de avaliar um conjunto de competências acadêmicas e habilidades profissionais tomando como base os conhecimentos definidos nas Diretrizes Curriculares Nacionais para os cursos de graduação em Psicologia em uma amostra nacional de estudantes para que, a partir destas informações, seja possível conhecer o processo de formação do psicólogo nas IES Brasileiras, no que diz respeito ao desenvolvimento destas competências. Ele também pretende avaliar o desempenho em questões transdisciplinares, envolvendo um conhecimento mais geral.

Porém, um aspecto essencial a ser discutido neste artigo se refere à questão da validade da prova (Primi, 2006), isto é, em que medida o ENADE atingiu os objetivos definidos pelos planos iniciais da prova. A resposta a essa questão é possível por meio do emprego de ferramentas analíticas da Psicometria e Teoria de Resposta ao Item na análise das características da prova, como será apresentado a seguir. Portanto, em seguida relata-se a análise dos itens e da escala geral, formada a partir da agregação dos itens, procurando verificar as características positivas e negativas da prova e como elas afetam as interpretações pretendidas.

Ao considerarmos mais aprofundadamente os objetivos do ENADE, há dois elementos importantes que são tomados como objeto de estudo: (a) o construto e (b) a avaliação do processo

de formação. Em relação ao primeiro aspecto, embora não muitas vezes tão evidente nas discussões sobre o ENADE, é preciso deixar clara a distinção entre as manifestações observáveis, isto é, o acerto e erro de cada estudante nos itens do ENADE, indicando o domínio de determinados conteúdos específicos, e o construto latente inobservável que se pretende inferir a partir do desempenho, isto é, o nível de competência, habilidades e conteúdos que o estudante adquiriu. Portanto, inferimos a posição do aluno no construto a partir de suas respostas ao teste, isto é, agregamos os observáveis do estudante (itens corretamente respondidos) resumindo esta informação em um número que pretende estimar a posição do aluno no construto latente. A questão da validade que está implícita nesse caso se associa ao grau de credibilidade que temos na interpretação que fazemos da nota quando a usamos para estimativa do nível de competência do estudante. Como em toda estimativa, sempre existirá um erro associado, o que resta saber é se este erro é grande o suficiente a ponto de impedir que se faça qualquer tipo de inferência confiável sobre as diferenças no nível de competência entre alunos e cursos.

Em relação ao segundo aspecto, outra distinção não tão evidente precisa ser feita. Ela se refere ao fato de que o desempenho do estudante é observado como meio para se avaliar a qualidade do curso. Portanto, em última instância, as inferências são feitas sobre a qualidade do curso, isto é, tem-se claramente uma estrutura multinível na qual observamos o desempenho de um grupo de alunos (nível 1) dentro de um curso (nível 2) para agregar tais informações e construir uma nota para as instituições e poder compará-las entre si.

Ainda com relação à distinção entre construto latente e variável observada, é preciso salientar que a visão de construto latente, bastante comum na psicologia, não é muito comum em outras áreas do conhecimento, também envolvidas na mensuração da qualidade de sistemas educativos. Na perspectiva psicológica, os construtos têm existência própria, demonstrada pela literatura como dimensões causais do comportamento. Portanto, segundo essa perspectiva, as competências são entidades latentes que procuramos definir e medir para poder explicar o comportamento. A psicologia procura definir empiricamente quais dimensões latentes subjazem ao comportamento e explicam as diferenças individuais em termos de personalidade, inteligência e, incluindo nesta última, as dimensões das competências profissionais e habilidades

acadêmicas (Gottfredson & Saklofske, 2009). Nessa perspectiva da psicologia, as variáveis observadas são indicadores do construto, que existem por si só, são latentes e as causas dessas variáveis observadas. De certa forma, essa perspectiva implica que devemos chegar ao mesmo resultado independentemente dos indicadores que utilizamos, desde que eles sejam indicadores válidos das mesmas variáveis latentes.

Outras áreas concebem a mesma situação de maneira diferente. Por exemplo, na economia, em que se agrega um conjunto de variáveis para se obter um índice econômico com objetivos pragmáticos, não há a pressuposição de uma variável causal latente como causa dos indicadores. O índice é somente uma componente principal, isto é, uma agregação/resumo de um conjunto de variáveis observadas. A mudança dos indicadores muda o sentido do índice. Não há uma preocupação teórica e prática em se definir a variável latente, mas sim em obter um índice pragmático que possa ser usado como um indicador. Embora em termos práticos essa distinção não faça diferença, em teoria, elas são bastante importantes, pois revelam atitudes e esforços de pesquisa muito diferentes quando se considera uma ou outra perspectiva. Em nossa opinião, na educação, assim como na psicologia, estamos lidando com variáveis latentes, isto é, os construtos competências, habilidades e conhecimentos se referem, em última instância, a aspectos taxonômicos da inteligência humana. Assim todos os esforços de medida desses construtos devem também se preocupar com a definição e comprovações empíricas das suposições sobre essas variáveis latentes.

De maneira mais geral, pensando-se na prova, pode-se questionar em que medida ela consegue representar o construto planejado e assim permite diagnosticar estudantes e cursos em termos dos diferentes níveis no construto. Para responder a essa questão é preciso definir qual é o construto subjacente à prova. Essa não é uma questão fácil de responder. Em princípio, o construto é definido pelas diretrizes gerais da prova em termos de competências profissionais (por exemplo, realizar orientação, aconselhamento psicológico e psicoterapia) e habilidades acadêmicas (por exemplo, analisar, descrever e interpretar relações entre contextos e processos psicológicos e comportamentais) aplicadas aos eixos de conteúdo. Mas, ao examinar mais detalhadamente essas definições, fica evidente que elas são amplas e se referem não só a um construto, mas, eventualmente,

a um conjunto mais complexo de construtos. Por exemplo, uma análise de conteúdo das seis habilidades acadêmicas (ver Tabela 2), considerando os modelos recentes da Psicologia da Inteligência (Carroll, 1993; McGrew, 2009; Primi, 2003), indica que elas estão relacionadas a pelo menos quatro habilidades amplas: leitura e escrita, habilidades de raciocínio geral e quantitativo, habilidades de raciocínio sobre conteúdos comportamentais (percepção de emoções e conhecimento emocional) e de interação humana e habilidades de gerenciar relacionamentos sociais e emoções nos outros.

Ao examinarmos as competências profissionais do ponto de vista da sua estrutura dimensional, isto é, o número e organização das dimensões subjacentes, fica ainda mais evidente a complexidade do construto que se pretende avaliar com a prova. Como discutido em detalhe em outros trabalhos (Bastos, 2005; Perrenoud, 1997; Primi e cols., 2001), competência é um conceito hierarquicamente mais amplo e integrativo. As competências se referem à mobilização de determinados conhecimentos - tanto em termos de informação (factual) quanto em termos de “saber como fazer” (conhecimentos de procedimentos) - e de habilidades acadêmicas em torno de uma situação problema em um domínio relativamente específico do universo profissional, de tal forma que resulte em uma resposta, em um nível alto de maestria, satisfatória para resolução do problema segundo os critérios dessa área profissional. Portanto, a competência se refere à mobilização desses recursos, resultando na resposta adequada ao problema profissional encontrado.

A concepção geral do que é competência, isto é, a ideia de que ela significa uma mobilização efetiva de recursos (habilidades e conhecimentos), é menos controversa do que a definição de quais competências um profissional deve desenvolver no percurso da formação profissional. Portanto, a complexidade está na definição do conjunto de competências definidoras de um determinado campo profissional. Como competência é um conceito combinatório, ou seja, se refere à combinação de habilidades acadêmicas e conteúdos profissionais de maneira aplicada a uma situação problema, sua natureza é fundamentalmente multidimensional/relacional, tornando a operacionalização da avaliação bastante difícil. A avaliação se torna complexa, pois, tradicionalmente, os requisitos técnicos de construção de instrumentos de avaliação apontam que é mais desejável que se

tenha um conjunto simplificado de dimensões menores facilitando a criação de instrumentos que

meçam cada dimensão separadamente (unidimensionalidade) de maneira mais confiável.

Tabela 2. Descrição das habilidades acadêmicas amplas da prova do ENADE de Psicologia.

	Leitura e Escrita	Raciocínio (percepção de relações e correlatos) Geral e Quantitativo	Raciocínio sobre comportamentos e interação humana (percepção de emoções e conhecimento emocional)	Gerenciamento de relacionamentos social-emocional (administração de emoções em si e nos outros)
Habilidades ...saber				
a) Levantar informação bibliográfica em indexadores, periódicos, livros, manuais técnicos e outras fontes especializadas através de meios convencionais e eletrônicos;	■			
b) Ler e interpretar comunicações científicas e relatórios na área da Psicologia;	■			
c) Dominar e utilizar os fundamentos lógicos dos diferentes métodos de investigação científica;		■		
d) Planejar e realizar várias formas de entrevistas com diferentes finalidades e em diferentes contextos;				■
e) Analisar, descrever e interpretar relações entre contextos e processos psicológicos e comportamentais;			■	
f) Descrever, analisar e interpretar manifestações verbais e não verbais como fontes primárias de acesso a estados subjetivos;			■	
g) Utilizar os recursos da matemática, da estatística e da informática para a análise e apresentação de dados e para a preparação das atividades profissionais em Psicologia.		■		

Como parte dos esforços em se tentar operacionalizar as competências definidas nas diretrizes em itens do ENADE, a comissão tentou explicitar a natureza das competências que se desejava avaliar a partir de uma matriz relacionando competências e eixos de conteúdo (ver Tabela 3). Como se pode notar, por exemplo, a competência 1 (elaborar relatos científicos, pareceres técnicos, laudos e outras comunicações profissionais) envolve a mobilização de habilidades acadêmicas de leitura escrita e dos conhecimentos dos eixos 2, 3, 5 e 6, em situações específicas do exercício profissional. Todas as doze competências elencadas nas diretrizes estão associadas mais fortemente a eixos de conteúdos e habilidades, formando um conjunto combinatório complexo maior do que as doze unidades iniciais.

Diante dessas limitações, a comissão decidiu por elaborar a matriz de especificação da prova iniciando pelos seis eixos de conteúdo, definindo mais detalhadamente âmbitos do conhecimento que cada questão devia explorar. Evidentemente, buscou-se criar uma prova com maior abrangência possível da matriz de competências/habilidade/conteúdos. No entanto, o aspecto mais viável para se iniciar a operacionalização é a dimensão conteúdo. Por meio das matrizes de relacionamento expostas acima, a comissão pretendeu indicar as habilidades acadêmicas e as competências profissionais potenciais que deveriam estar na base da resolução dessas questões.

Tabela 3. Matriz buscando relacionar Competências com conteúdo para operacionalizar a criação de itens – ENADE/Psicologia.

Competências	Habilidades (gerais)	Eixos
Competências ... saber	Leitura e Escrita	<p>Raciocínio (percepção de relações e correlatos) Geral e Quantitativo</p> <p>Raciocínio sobre comportamentos e interação humana (percepção de emoções e conhecimento emocional)</p> <p>Gerenciamento de relacionamentos social-emocional (administração de emoções em si e nos outros)</p> <p>1) Fundamentos epistemológicos, teórico-metodológicos e históricos das principais vertentes do pensamento em psicologia.</p> <p>2) Fundamentos, métodos e técnicas de coleta e análise de informações para investigações científicas e avaliação de fenômenos psicológicos.</p> <p>3) Fenômenos, processos e construtos psicológicos, entre os quais, processos básicos (cognição, emoção, motivação e aprendizagem), processos do desenvolvimento, interações sociais, saúde psicológica e psicopatologia, personalidade e inteligência.</p> <p>4) Interfaces com campos afins do conhecimento (neurociências, sociologia, antropologia, filosofia).</p> <p>5) Práticas profissionais nos principais domínios de atuação do Psicólogo priorizando as intervenções nos processos educativos, de gestão, de promoção de saúde, clínicos e de avaliação.</p> <p>6) Princípios e normas éticas para a pesquisa e para a prática profissional.</p>
<p>a) Analisar o campo de atuação profissional e seus desafios contemporâneos;</p> <p>b) Analisar o contexto em que atua profissionalmente em suas dimensões institucional e organizacional, explicitando a dinâmica das interações entre os seus agentes sociais;</p> <p>c) Identificar e analisar necessidades de natureza psicológica, diagnosticar, elaborar projetos, planejar e agir de forma coerente com referenciais teóricos e características da população-alvo;</p> <p>d) Identificar, definir e formular questões de investigação científica no campo da Psicologia, vinculando-as a decisões metodológicas quanto à escolha, coleta, e análise de dados em projetos de pesquisa;</p> <p>e) Escolher e utilizar instrumentos e procedimentos de coleta de dados em Psicologia, tendo em vista a sua pertinência;</p> <p>f) Avaliar fenômenos humanos de ordem cognitiva, comportamental e afetiva, em diferentes contextos;</p> <p>g) Realizar diagnóstico e avaliação de processos psicológicos de indivíduos, de grupos e de organizações;</p> <p>h) Coordenar e manejar processos grupais, considerando as diferenças individuais e socioculturais dos seus membros;</p> <p>i) Atuar inter e multiprofissionalmente, sempre que a compreensão dos processos e fenômenos envolvidos assim o recomendar;</p> <p>j) Atuar profissionalmente, em diferentes níveis de ação, de caráter preventivo ou terapêutico, considerando as características das situações e dos problemas específicos com os quais se depara;</p> <p>k) Realizar orientação, aconselhamento psicológico e psicoterapia;</p> <p>l) Elaborar relatos científicos, pareceres técnicos, laudos e outras comunicações profissionais, inclusive materiais de divulgação.</p>		

Assim, pretendeu-se definir de maneira ampla as possíveis habilidades e competências avaliadas pelas questões mapeando-se uma área ampla de conteúdos básicos do núcleo comum da formação profissional. Porém, essas especificações foram feitas teoricamente. As especificações dos itens são originadas nos conteúdos dos eixos temáticos.

Análise fatorial dos Itens da Prova

A primeira análise empírica dos dados da prova, com o objetivo de verificar que dimensões do construto foram de fato operacionalizadas e avaliadas pela prova, é a análise fatorial dos itens. Essa análise examina a matriz de correlação entre todos os pares de itens e descobre grupos de itens altamente correlacionados entre si. Dois itens estão altamente correlacionados quando o estudante que acerta um deles tende a acertar o outro e quando erra um deles tende a errar o outro. Com base nesse fato, podemos inferir que os dois exigem uma mesma competência/habilidade/conteúdo. A análise fatorial considera simultaneamente todos os pares possíveis de correlação entre os itens da prova e indica quais itens formam grupos por estarem altamente correlacionados. E, por meio da análise de conteúdo desses grupos de itens, podemos inferir qual aspecto do construto os itens exigem. Dessa forma, pode-se entender melhor qual ou quais construtos foram avaliados pelo conjunto particular de itens da prova agrupados pelo fator.

Na presente análise, submetemos os dados a uma análise fatorial por informação completa (*Full Information Factor Analysis*) implementada pelo TESTFACT, que é um procedimento baseado na Teoria de Resposta ao Item (TRI) (Bock, Gibbons & Muraik, 1986; Primi & Almeida, 1998; Wilson, Wood & Gibbons, 1991). Analisamos os dados de 26.613 estudantes nos 30 itens do componente específico (26 questões de múltipla escolha e 4 questões dissertativas). A pontuação de 0-100 nas questões dissertativas foram dicotomizadas para possibilitar a análise pelo TESTFACT, uma vez que este modelo só trabalha com itens dicotômicos. O esquema de dicotomização foi o seguinte: escores 0-30 receberam “0” e escores 40-100 receberam “1”.

A média geral dos estudantes na prova, em termos de proporção de acerto, foi 0,452 com desvio padrão de 0,498. A consistência interna geral foi $KR20=0,786$. A matriz de correlações tetracóricas era composta por 435 pares de correlação válidos, cuja média foi $r=0,109$ e

$DP=0,11$. Assim, um primeiro ponto que chama atenção é a baixa magnitude das correlações entre os itens. Os *eigenvalues* dos cinco primeiros fatores foram (5,28; 1,65; 1,25; 1,16 e 1,09), sugerindo a existência de um fator predominante. O segundo e terceiro fatores também foram considerados para uma análise exploratória em razão de sua relativa saliência. Na Tabela 4 são apresentados os resultados da extração de três fatores rotados pelo procedimento Promax. Nela se podem ver as cargas fatoriais que são as correlações de cada item com os fatores extraídos (grupos de itens com alta intercorrelação entre si), a estimativa da dificuldade do item pela TRI (valores abaixo de zero indicando itens mais fáceis e acima de zero itens mais difíceis) e as comunalidades (a proporção de variância explicada pelos fatores) e a correlação entre os fatores.

O que se pode entender a partir desses resultados sobre o que foi operacionalizado nessa prova? Primeiramente, é preciso mencionar as baixas correlações entre os itens. Isso indica que a variabilidade do acerto e erro nas questões (especialmente nas de múltipla escolha) tende a ser única, isto é, não está fortemente compartilhada ou correlacionada com outros subconjuntos de itens da prova. Pode-se imaginar cada item como uma mini prova de conhecimentos específicos independente dos outros itens. Isso é consistente com a natureza complexa do construto avaliado, conforme descrito anteriormente, e também com a natureza especializada do conhecimento envolvido nessa fase educacional. Resultados como esse são frequentes em provas educacionais (*achievement*) e são bem diferentes dos padrões que os pesquisadores que lidam com testes e escalas psicológicas estão acostumados a ver. Nelas, um princípio frequentemente adotado é a maximização da consistência, isto é, a construção de várias questões medindo uma mesma dimensão subjacente (Cronbach, 1996). Já as provas educacionais são construídas com princípios fundamentalmente diferentes, pois elas têm como propósito maximizar a representação de um domínio multidimensional. Isso se reflete nos índices de consistência. O resultado disso, em termos de interpretação das notas, é que diferentes estudantes com a mesma nota 5 (em uma escala de zero a dez), por exemplo, não necessariamente podem ser caracterizados com tendo adquirido as mesmas competências, pois suas configurações de acerto aos itens, isto é, quais itens cada um acerta, podem ser diferentes.

Tabela 4 - Matriz de cargas fatoriais dos itens da prova do ENADE de Psicologia.

DISPLAY	20.	PROMAX	ROTATED	FACTOR	LOADINGS	Diff	Commun.
			1	2	3		
1	CE11		0.215	-0.073	0.106	0.421	0.058
2	CE12		0.573	-0.048	-0.011	-1.247	0.297
3	CE13		0.043	0.074	-0.173	0.463	0.024
4	CE14		0.054	0.016	0.071	0.557	0.013
5	CE15		-0.021	-0.016	0.089	0.539	0.006
6	CE16		0.057	-0.009	-0.095	0.452	0.008
7	CE17		0.181	-0.088	0.193	0.332	0.075
8	CE18		0.214	-0.019	0.033	-0.336	0.048
9	CE19		0.468	-0.076	0.027	-0.451	0.196
10	CE20		0.550	-0.056	0.042	-1.010	0.291
11	CE21		0.474	-0.041	-0.072	-0.630	0.186
12	CE22		0.337	0.044	0.282	-0.213	0.299
13	CE23		0.387	-0.046	0.278	-0.084	0.287
14	CE24		0.165	-0.089	0.191	0.471	0.067
15	CE25		0.562	-0.045	0.128	-0.465	0.361
16	CE26		0.402	-0.050	0.006	-0.115	0.145
17	CE27		-0.021	-0.081	0.261	0.941	0.055
18	CE28		0.188	-0.020	0.108	0.212	0.058
19	CE29		0.437	-0.103	0.224	0.149	0.264
20	CE30		0.083	-0.020	-0.029	0.734	0.005
21	CE31		0.476	-0.043	0.102	-0.143	0.253
22	CE32		0.117	-0.014	0.177	0.528	0.058
23	CE33		0.421	0.078	-0.142	-0.439	0.181
24	CE34		0.162	-0.017	0.068	0.479	0.036
25	CE35		0.153	0.096	-0.121	-0.121	0.038
26	CE36		0.488	0.061	-0.005	-0.471	0.271
27	CE37		0.133	0.522	0.082	0.309	0.416
28	CE38		0.251	0.365	0.242	1.039	0.476
29	CE39		0.114	0.630	0.109	0.508	0.565
30	CE40		-0.123	0.979	-0.038	-0.147	0.820
EMPIRICAL							
			1	2	3		
	RELIABILITY:		0.6305	0.5854	0.1768		
DISPLAY 21. PROMAX FACTOR CORRELATIONS							
			1	2	3		
		1	1.000				
		2	0.532	1.000			
		3	0.406	0.419	1.000		

Apesar dessa baixa correlação, há evidência de um fator geral que pode ser indicado tanto pelo primeiro *eigenvalue* quanto pelas correlações entre os três fatores extraídos. Isso quer dizer que apesar da relativa independência dos itens, a parcela que está correlacionada tende a formar uma dimensão predominante. Como, então, conciliar esses dados aparentemente contraditórios: as baixas correlações e, ao mesmo tempo, a presença de um único fator geral indicando a unidimensionalidade?

É mais fácil compreender a ideia da unidimensionalidade a partir da organização dos itens, ordenando-os em uma linha em função de sua dificuldade. Os itens têm diferentes níveis de dificuldade e isto ocorre porque há diferenças na complexidade das competências/habilidades/conteúdos avaliados. Quando ordenamos os itens do menos complexo (à esquerda) ao mais complexo (à direita) temos uma hierarquia de complexidade da aquisição dessas

competências. Assim, a existência de um fator geral significa que os estudantes que acertam os itens complexos tendem a acertar os itens anteriores, considerados como mais fáceis na hierarquia. Os escores mais baixos no teste indicam o acerto dos itens mais fáceis. Os aumentos progressivos no número de acertos indicam o acerto em itens progressivamente mais complexos da hierarquia. Portanto, a existência de um fator geral indica essa “sistematicidade” nos perfis de acerto às questões, indicando também uma organização progressiva da aquisição das competências/habilidades/conteúdos principais representadas pelas questões. Essa hierarquia será mais detalhada quando analisarmos o mapa de itens da prova mais adiante.

Embora haja evidência a favor dessa hierarquia, as baixas correlações entre os itens e, conseqüentemente, a baixa consistência, indica que essa sistematicidade nem sempre acontece, podendo ocorrer casos de alunos com configurações

diferentes das esperadas pela hierarquia como, por exemplo, com escores medianos que acertaram itens difíceis e erraram os itens mais fáceis, ao invés de terem acertado os itens fáceis e médios como seria esperado pelo modelo unidimensional. Os padrões inesperados são indicativos de independência das competências/habilidades/conteúdos, isto é, indicam que o estudante pode mobilizar seus recursos para responder a uma determinada questão, mas isto não garante que ele saberá responder a outra, ainda que esta seja mais fácil do que a anterior. Isso pode ocorrer quando as competências/habilidades/conhecimentos requeridos por dois itens não são exatamente os mesmos.

A presença de um fator geral subjacente indica que os padrões esperados são mais frequentes do que os inesperados. A magnitude da associação entre os itens, que se reflete na força do fator (*eigenvalue*), nos informa qual o resultado do jogo entre os padrões esperados e inesperados. Se for muito alta, os padrões inesperados são bastante raros. Se forem mais baixas, esses padrões são mais frequentes. Para a conciliação dos dados aparentemente contraditórios devemos considerar a existência desses dois padrões nos dados, geral e específico, ocorrendo simultaneamente, mas com uma predominância do primeiro em relação ao segundo em razão do fator geral.

Além desse fator geral, é possível notar ainda a existência de mais dois subfatores (ou facetas), agrupando conjuntos de questões que nos indicam quais competências/habilidades/conhecimentos específicos definem a variável subjacente. O primeiro fator é o geral da prova, já que possui o maior *eigenvalue* e tem a maior quantidade de itens com cargas superiores a 0,30 e se correlaciona com os outros dois. É interessante analisar o conteúdo dos itens com maior carga fatorial (CE12, CE20, CE21, CE31 e CE36), pois eles nos dão um indicativo de como interpretar o sentido da variável latente aos itens em termos de competências/habilidade/conteúdos avaliados. Esses itens são relativamente mais fáceis e, em termos de conteúdo, tratam da relevância dos fatores culturais e sociais como elementos importantes na psicologia contemporânea, que parece ser um teor geral da prova. Em termos de habilidade, esses itens se associam à habilidade básica de ler e interpretar textos já que as alternativas erradas são mais facilmente excluídas por serem contraditórias à alternativa certa.

O segundo fator agrupa as questões dissertativas. O formato dessas questões permite que se avaliem habilidades e competências mais distintas das questões de múltipla escolha. Elas permitem uma estimação mais confiável do nível de competência dos alunos em termos da sua capacidade de mobilizar os conhecimentos adquiridos para oferecer uma solução adequada a uma demanda comumente encontrada na prática profissional (Primi & Muniz, 2004). Também permitem a avaliação de habilidades acadêmicas de expressão escrita. Os itens com carga mais elevada foram CE40, CE39 e CE37, associados aos eixos: práticas profissionais nos principais domínios de atuação do Psicólogo e princípios e normas éticas para a prática profissional. A partir do tipo de problema proposto se pode entender que esses itens requerem a competência de mobilizar os conhecimentos sobre análise de necessidades psicológicas de indivíduos e/ou grupos, diagnóstico e avaliação de processos psicológicos, elaboração de projetos de caráter preventivo ou terapêutico em contextos multiprofissionais, levando-se em conta os elementos peculiares do contexto. Portanto, esse fator se refere a uma competência profissional da prática clínica e com grupos e organizações.

O terceiro fator tem importância limitada, do ponto de vista estatístico, em razão da baixa consistência (cargas fatoriais menores do que 0,30). Entretanto, ele aponta para um conjunto específico de competências/habilidades/conteúdos que trazem informações importantes para o entendimento do construto avaliado pela prova, permitindo uma reflexão sobre eventuais vieses dentro do processo de formação. A maioria dos itens que define esse fator (CE17, CE22, CE23, CE27, CE29 e CE38) se refere ao eixo dos processos básicos, especialmente envolvendo conhecimentos de psicopatologia, aprendizagem e fundamentos e métodos de análise de dados em investigações científicas. A questão dissertativa CE38 também apresenta uma associação nesse fator, reforçando a ideia de que a variável latente que esse fator representa se refere a competências de mobilização de conhecimentos básicos dos métodos envolvidos em investigações científicas e habilidades de interpretar informações estatísticas. Os itens desse fator têm dificuldade mediana a difícil, sugerindo que essas competências são adquiridas mais tardiamente.

Um fato notável é a existência de alguns itens com carga negativa nesse fator (CE13, CE33 e CE35). Esses itens se referem à crítica ao positivismo na perspectiva histórico-crítica, à

interferência de aspectos políticos na prática do psicólogo organizacional e ao uso de técnicas projetivas. Essas cargas fatoriais indicam que um conjunto de estudantes com notas altas no fator tendem a errar as questões CE13, CE33 e CE35 e, o inverso, os estudantes que acertam estas questões tendem a ter nota baixa nesse fator. Esse resultado talvez indique a dicotomização da estrutura de conhecimento de pensamentos psicológicos antitéticos, um mais científico por um lado e outro mais histórico-crítico do outro, sendo complexo pensar no referencial do outro. Esses dados indicam que um subgrupo de estudantes com um pensamento mais predominante do primeiro tipo erra questões com o segundo tipo de pensamento e vice-versa. Tal verificação estimula a reflexão sobre o processo de formação que, em nossa opinião, deveria criar um domínio mais diversificado de conhecimentos e não dicotomizado. A Psicologia é uma ciência nova, em desenvolvimento, e como tal abarca teorias explicativas antitéticas de um mesmo fenômeno. Uma formação básica adequada requer a explicitação das tensões e o conhecimento dos argumentos antitéticos e a apropriação do debate lógico entre eles, mas não a tomada de posição em relação a um dos argumentos somente. Os dados observados no terceiro fator levantam a questão se uma formação mais ampla, voltada para a síntese e não para as teses e antíteses, está sendo atingida.

Análise dos itens e da escala pelo modelo de Rasch

A segunda análise submeteu os dados à estimação dos parâmetros da Teoria de Resposta ao Item, seguindo o modelo de Rasch, buscando definir a métrica da escala geral subjacente aos itens (Embretson 2006; Primi, Fernandez & Ziviani, 2003; Ziviani & Primi, 2002; Wright & Stone, 1979). Georg Rasch (1980), matemático psicometrista dinamarquês, desenvolveu um modelo matemático para análise de provas educacionais de leitura de exames em larga escala em seu país que, subsequentemente, atingiu um largo espectro de aplicabilidade nas ciências humanas e da saúde. Esse modelo é usado para agregar dados de respostas aos itens calibrando-se medidas intervalares para os itens (nível de dificuldade) e para os estudantes (nível de competência). A grande vantagem do modelo está nos recursos interpretativos que ele traz como discutidos por Embretson (2006), Woodcock (1999) e Primi (2004). Algumas dessas possibilidades serão exploradas mais adiante.

Não é propósito deste artigo explicar o modelo e por isto será feita uma apresentação

apenas em linhas gerais, evitando detalhes técnicos. A ideia fundamental do modelo está na expressão da relação entre os elementos da situação de testagem na qual uma pessoa responde a um problema por meio da Curva Característica do Item (CCI). A CCI indica a probabilidade de acerto em função da competência das pessoas (na variável latente) que respondem ao teste considerando-se simultaneamente a dificuldade do item. O modelo formal e gráfico é apresentado nas duas primeiras linhas da Figura 1. À esquerda está sua formulação matemática e à direita a CCI dos itens CE17 e CE16. Esses gráficos contêm no eixo x o traço latente Theta que representa o nível geral de competência dos estudantes. No eixo y está indicada a probabilidade de acerto. A curva cinza é a CCI modelada para esses itens. As cinco linhas (em cada gráfico) mostram o padrão encontrado para as cinco respostas (as respostas D no CE17 e B no CE16 correspondem à resposta certa que coincide com a curva modelada já que o modelo tenta reproduzir a probabilidade de acerto). Essas curvas encontradas são calculadas a partir dos dados criando-se subgrupos com níveis de competência progressivamente crescentes e calculando-se a proporção de estudantes em cada grupo, que acertam a questão e colocando-se esses pontos no gráfico, desenhando as curvas. A curva modelada CCI é a melhor “cópia” da curva observada a partir da equação matemática que contém o parâmetro de dificuldade do item.

Como pode ser visto, para o item CE17, a CCI modela a alternativa D, que é a escolha correta nesta questão. Conforme representado na CCI, quanto maior a competência do estudante maior será a probabilidade de que ele responda corretamente. Portanto, na medida em que a competência aumenta (Person Logit) aumenta também a probabilidade de se acertar o item. Para os valores -1, 0 e 1, as probabilidades são aproximadamente 0,21, 0,39 e 0,60 no item CE17. O parâmetro calibrado da dificuldade b do item corresponde ao valor de theta associado à probabilidade $p=0,50$ de acerto (para este item $b=0,55$). As quatro curvas das alternativas representam a probabilidade de escolha das quatro alternativas restantes, B, E, C e A, em razão dos níveis de competência. Sua tendência é diminuir à medida que aumenta a capacidade já que a alternativa correta passa a ser mais escolhida entre os alunos com alta competência. Nota-se também que a curva cinza teórica é praticamente coincidente com a curva empírica de probabilidade X theta da

alternativa D, indicando um bom ajuste do modelo aos dados no caso desse item.

$$P_{ij}(\theta_j) = \frac{e^{D(\theta_j - b_i)}}{1 + e^{D(\theta_j - b_i)}}$$

Onde,

$P_{ij}(\theta_j)$ = probabilidade de que a pessoa j, com habilidade θ_j acerte o item i;

θ_j = a habilidade da pessoa j,

b_i = dificuldade do item i,

D = constante de ajuste igual a 1,7,
e = constante matemática igual a 2,72.

$$P_{nix}(\theta) = \frac{e^{\sum_{j=0}^x (\theta_n - \delta_{ij})}}{\sum_{r=0}^{m_i} \left[e^{\sum_{j=0}^r (\theta_n - \delta_{ij})} \right]}$$

Onde $\sum_{j=0}^0 (\theta_n - \delta_{ij}) \equiv 0$,

$P_{nix}(\theta)$ indica a probabilidade de o sujeito n ter o escore x no item i. Os escores no item têm a notação $x = 0, \dots, m_i$ para um item com $K_i = m_i + 1$ categorias de resposta. Os valores

$\delta_{ij} (j = 1, \dots, m_i)$ indicam os valores dos limiares de transição entre a categoria j-1 e a categoria j. Esses valores indicam o ponto de intersecção entre as curvas da categoria j-1 e j. Em situações ideais, esse ponto indicará a o momento a partir do qual a j passa a ser a mais provável, portanto o passo de transição entre a categoria menor j-1 e a categoria em consideração a qual este parâmetro se refere.

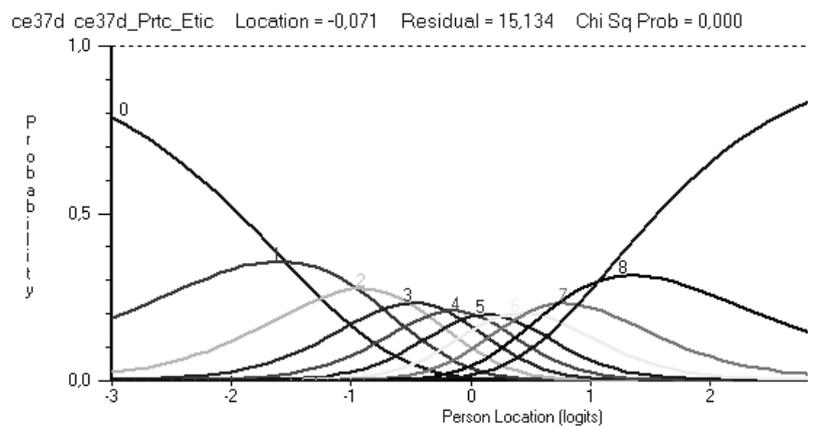
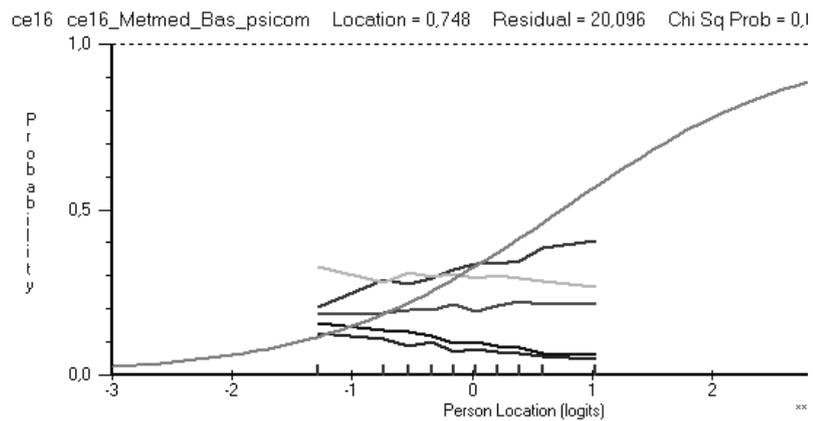
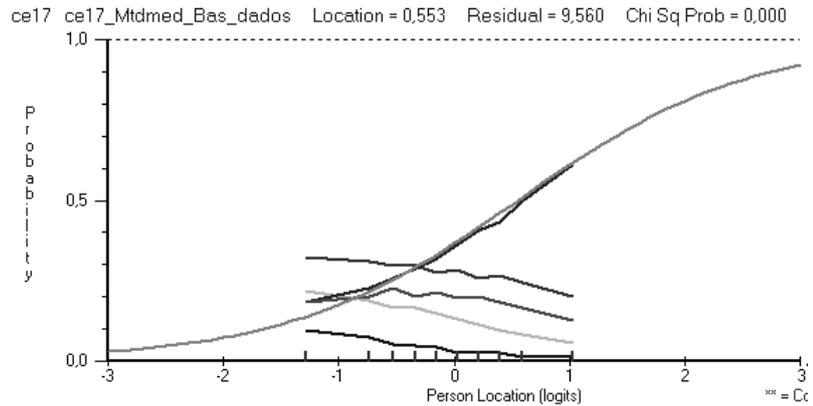


Figura 1 - Curvas características dos itens CE17, CE16 e Ce37.

Com a calibração do modelo, cada estudante recebe uma medida indicando sua posição na variável latente e cada item, um parâmetro de dificuldade. A aplicação desses valores na fórmula permite prever a probabilidade de acerto do estudante ao item. Dito de outro modo, conhecendo-se o nível de competência do estudante e a dificuldade dos itens, pode-se calcular uma expectativa teórica do padrão de acertos aos itens. Se houver uma boa correspondência entre os dados e o modelo, essas previsões serão bem acuradas. Assim, por meio da Curva Característica do Item se estabelece uma relação que tem, de um lado, a probabilidade de acerto ($P_{ij}(\theta_j)$), e do outro, uma comparação entre competência do estudante e dificuldade do item ($\theta_j - b_j$). Nesse processo de comparação, se o estudante acertou o item é porque sua competência excedeu a dificuldade do item ($\theta_j > b_j$). Reversamente, se errou, é porque sua competência foi inferior à dificuldade do item ($\theta_j < b_j$). Portanto, uma vez conhecida a capacidade do estudante, pode-se estabelecer as expectativas de acerto nos itens que avaliam aquela competência. Por exemplo, quando uma pessoa tem um nível de competência igual ao índice de dificuldade do item, as chances são de 50% de que ela o acerte¹. Na medida em que o nível de competência aumente em relação à dificuldade do item, suas chances de acertá-lo serão maiores do que 50%. Na medida em que sua competência seja menor do que a dificuldade do item, suas chances de acertá-lo serão menores do que 50%. Portanto, levando-se em conta o nível de competência do estudante e a organização hierárquica dos itens pode-se prever quais ele acertou e fazer inferências sobre as competências adquiridas.

No caso de questões dissertativas, o modelo é um pouco mais complexo. Na terceira linha da Figura 1, apresentam-se as curvas do item CE37 e, à esquerda a formulação matemática. Esses itens são pontuados por examinadores em uma escala de 0 a 100, que foram recodificados para uma escala de dez pontos (0, 1, 2, 3, 4, 5, 6, 7, 8, 9-10). Nesse modelo, cada aluno recebe também uma medida de sua competência e os itens são caracterizados por parâmetros de transição. A CCI nesse caso mostra a relação entre as probabilidades de pontuação 0,1,2, entre outras e o valor de Theta. Em cada região da escala há uma nota observada mais provável. Por exemplo, prevê-se que estudantes com nível de competência -2 recebem o escore 0. Na região -1,2

recebem pontuação 1 e assim por diante. Portanto à medida que se progride em competência, aumenta-se a probabilidade de pontuações superiores. A dificuldade do item é caracterizada pelos pontos de transição das curvas. Esses pontos indicam o momento que uma determinada pontuação (2 por exemplo) passa a ser mais provável que a anterior (1). Esse modelo é chamado modelo de créditos parciais (Wright & Masters, 1981).

Evidentemente, os dados podem ser incompatíveis com o modelo em casos nos quais, por exemplo, os padrões de resposta ao item não acontecem como o modelo prevê. Isso pode ocorrer por vários motivos que precisam ser investigados durante a fase analítica do instrumento na qual se executa a calibração dos parâmetros. Essa falta de correspondência modelo-observações ocorreu, por exemplo, no item CE16, exemplificado na Figura 1. Nota-se que a curva modelada, embora próxima da curva da alternativa B, considerada inicialmente correta, não adere bem aos valores empíricos encontrados. Isso ocorreu provavelmente por um problema na formulação do item. Os estudantes mais competentes não acertam mais do que os estudantes menos competentes, talvez por se confundirem com alternativas que também parecem possíveis (C e E). Isso parece ser a explicação mais plausível nesse caso já que, antes mesmo de analisar os resultados da prova, a comissão já tinha anulado esse item com base na análise de conteúdo que detectou essa ambiguidade. Portanto, na análise para calibração dos parâmetros, é preciso verificar a adequação entre as expectativas teóricas e os dados observados para verificar a viabilidade de aplicação do modelo.

Como no ENADE misturam-se questões de múltipla escolha com questões dissertativas, foi necessário analisar os dados com o modelo de créditos parciais. Tal modelo, mais recente na família de modelos de Rasch, permite a calibração simultânea de itens com diferentes estruturas dicotômicas e politômicas. Nesse caso, os 34 itens de múltipla escolha foram modelados com a fórmula tradicional conforme apresentado. Aos itens dissertativos corrigidos com escores politômicos (0 a 10) ajustou-se o modelo de créditos parciais. A calibração conjunta permite que sejam aplicados modelos distintos a conjuntos de itens e se obtenham as medidas dos sujeitos na mesma escala.

Dos 30 itens da prova do componente específico, dois foram eliminados pela comissão

por possuírem problemas na formulação. Os 28 itens restantes foram submetidos à análise de Rasch para calibração dos parâmetros pelo programa WINSTEPS (Linacre & Wright, 1991). O modelo de Rasch se baseia na suposição da unidimensionalidade e, portanto, antecipa os padrões de resposta discutidos anteriormente. Como na análise fatorial não fica tão clara a predominância do primeiro fator, é preciso analisar os resultados da calibração verificando-se em que medida dados correspondem com a expectativa do modelo. Considerando os critérios técnicos discutidos na literatura (índices de ajuste e análise fatorial dos resíduos, Smith & Smith, 2004) concluiu-se que os dados se ajustaram adequadamente ao modelo.

Na Tabela 5, são apresentados resultados gerais das medidas dos estudantes (retângulo superior) e dos itens (retângulo inferior). Dessas informações, as mais importantes são as seguintes: (a) as medidas dos estudantes (*measure*) variaram de um mínimo de -3,78 a um máximo de 1,89 com média -0,09 e desvio padrão 0,64; (b); a precisão

geral dessas medidas foi de 0,76, atingindo um nível aceitável. Caso esse valor tivesse sido muito baixo, haveria um problema com a confiabilidade das medidas de competência dos estudantes, já que se fosse criada outra prova, pretendendo medir os mesmos construtos, mas com itens diferentes, não teríamos garantia de observar, nos mesmos alunos, desempenhos próximos aos encontrados nessa edição do ENADE; (c) o índice de separação para a medida dos estudantes foi de 1,79, indicando que se pode ter de dois a três estratos de escores significativamente diferentes entre os estudantes. O índice de separação é calculado pela razão da variância verdadeira (corrigida pela estimativa de variância de erro) com o erro de medida, indicando, portanto, a dispersão das medidas dos estudantes em unidades de erro padrão, ou seja, quantas vezes a dispersão verdadeira entre as medidas supera o erro padrão da medida; (d) as medidas de desajuste dos itens (INFIT e OUTFIT) estiveram em níveis aceitáveis, atingindo no máximo 1,24. Em resumo, esses dados indicam um bom ajuste dos valores calibrados pelo modelo com os dados observados.

Tabela 5 - Estatísticas gerais dos estudantes e itens na prova após a calibração, segundo o modelo de Rasch.

Resultados para os estudantes (medidas Theta)

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	20.3	27.8	-.09	.28	.94	-.1	1.01	.0
S.D.	8.4	1.3	.64	.07	.50	.9	.31	.9
MAX.	51.0	28.0	1.89	1.17	6.30	4.9	9.90	9.9
MIN.	1.0	4.0	-3.78	.22	.19	-2.4	.14	-2.9
REAL RMSE	.31	ADJ.SD	.56	SEPARATION	1.79	P	RELIABILITY	.76
MODEL RMSE	.29	ADJ.SD	.57	SEPARATION	1.95	P	RELIABILITY	.79
S.E. OF P MEAN	= .00							

VALID RESPONSES: 99.4%

P RAW SCORE-TO-MEASURE CORRELATION = .97

CRONBACH ALPHA (KR-20) P RAW SCORE RELIABILITY = .76

Resultados para os itens (medidas b)

SUMMARY OF 28 MEASURED (NON-EXTREME) IS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	16337.1	22366.1	.00	.01	1.00	.6	1.01	.7
S.D.	14495.2	81.0	.97	.00	.07	8.2	.13	8.7
MAX.	71773.0	22507.0	1.58	.02	1.12	9.9	1.24	9.9
MIN.	3904.0	22244.0	-2.40	.00	.80	-9.9	.71	-9.9
REAL RMSE	.01	ADJ.SD	.97	SEPARATION	66.06	I	RELIABILITY	1.00
MODEL RMSE	.01	ADJ.SD	.97	SEPARATION	67.14	I	RELIABILITY	1.00
S.E. OF I MEAN	= .19							

UMEAN=.000 USCALE=1.000

Na Tabela 6, são apresentadas as estatísticas dos itens. Entre elas, as mais importantes são o valor estimado da dificuldade b de cada item (*Measure*), os valores de ajuste por item (INFIT MNSQ e OUTFIT MNSQ) e as correlações item

total (PTMEA CORR.). Esses dados mostram que todos os itens apresentam ajustes aceitáveis, a despeito de alguns possuírem baixa correlação item-total (Itens CE13, CE15, CE27 CE13 CE30). Esses

itens também foram os que apresentaram os índices OUTFIT MNSQ mais elevados.

Como foi dito anteriormente, uma vantagem do modelo de Rasch é que as estimativas de dificuldade dos itens e competências dos estudantes têm uma métrica comum. Disso decorrem implicações práticas importantes. Por exemplo,

podemos interpretar que um aluno com competência média $-0,09$ teria aproximadamente 35% de chance de acerto ao item CE11 ($b=0,637$), mas mais de 90% de chance de acerto do item CE12 ($b=-2,401$). Portanto, com o valor da medida de competência do estudante é possível prever quais itens ele terá chance de acertar e errar.

Tabela 6 - Estatísticas dos itens (dificuldade e índices de ajuste) após calibração pelo modelo de Rasch.

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEA CORR.	EXACT OBS%	MATCH EXP%	I
9	7608	22411	.637	.015	1.06	9.9	1.09	9.9	.17	65.9	67.3	ce11_FndBas_his
10	20077	22429	-2.401	.022	.97	-2.0	.89	-5.5	.28	89.7	89.6	ce12_FndBas_prof
11	7182	22327	.725	.015	1.12	9.9	1.17	9.9	.07	65.3	68.6	ce13_FndBas_sist
12	6495	22292	.880	.015	1.07	9.9	1.14	9.9	.13	70.7	71.2	ce14_FndBas_sig
13	6628	22409	.855	.015	1.10	9.9	1.18	9.9	.07	69.3	70.8	ce15_FndBas_sensocom
15	8358	22409	.481	.014	1.06	9.9	1.08	9.9	.18	63.0	65.1	ce17_Mtdmed_Bas_dados
16	14211	22403	-.686	.014	1.06	9.9	1.07	9.9	.19	64.3	66.7	ce18_Mtdmed_Bas_corr
17	15187	22417	-.894	.015	1.00	-.2	1.00	-.6	.29	70.1	69.8	ce19_Procbas_Bas_mem
18	18978	22428	-1.933	.019	.96	-3.6	.90	-6.2	.32	85.1	84.8	ce20_Procbas_Bas_desado
19	16533	22391	-1.215	.016	1.00	.6	1.01	.7	.27	75.0	74.8	ce21_Procbas_Bas_desinf
20	13281	22306	-.503	.014	.94	-9.9	.92	-9.9	.39	66.9	64.5	ce22_Procbas_Bas_psicopat
21	12151	22397	-.270	.014	.95	-9.9	.94	-9.9	.36	64.4	62.5	ce23_Procbas_Bas_saudoenc
23	15322	22355	-.932	.015	.93	-9.9	.90	-9.9	.39	72.3	70.4	ce25_Procbas_Bas_represoc
24	12295	22353	-.301	.014	1.02	4.2	1.02	4.2	.26	61.0	62.6	ce26_Procbas_Bas_psican
25	3904	22363	1.584	.018	1.05	4.8	1.24	9.9	.08	82.5	82.5	ce27_Procbas_Bas_aprendiz
26	9375	22297	.267	.014	1.05	9.9	1.06	9.9	.20	59.5	62.9	ce28_Intrfc_Bas_gestao
27	9998	22328	.147	.014	.97	-8.7	.96	-7.2	.33	64.8	62.1	ce29_Intrfc_Bas_locultura
28	5146	22244	1.214	.016	1.08	9.7	1.21	9.9	.07	76.6	76.9	ce30_Intrfc_Bas_neuroc
29	12580	22321	-.361	.014	.96	-8.9	.95	-9.6	.35	64.3	63.1	ce31_Intrfc_Bas_intelgen
30	6767	22316	.820	.015	1.04	6.7	1.08	9.3	.18	69.2	70.2	ce32_PraticEsc
31	15033	22251	-.880	.015	.99	-2.0	.99	-1.3	.30	70.7	69.6	ce33_PraticTrab
32	7106	22278	.742	.015	1.05	9.3	1.11	9.9	.16	68.5	68.9	ce34_PraticSaud
33	12283	22252	-.309	.014	1.08	9.9	1.09	9.9	.17	57.6	62.7	ce35_PraticDiag
34	15325	22245	-.947	.015	.94	-9.0	.92	-9.9	.37	72.2	70.7	ce36_PraticGrp
37	51033	22507	.746	.005	.92	-8.2	.90	-9.9	.67	27.4	27.0	ce37d_Pratic_Etic
38	24785	22507	.911	.004	.86	-9.9	.71	-9.9	.53	56.5	52.5	ce38d_Mtdmed_Bas
39	38025	22507	1.158	.005	.80	-9.9	.75	-9.9	.67	33.4	29.6	ce39d_Pratic_Esc_Tarb
40	71773	22507	.464	.004	.94	-6.4	.94	-5.0	.72	24.5	24.1	ce40d_Pratic_Clinc
MEAN	16337.1	22366.1	.000	.014	1.00	.6	1.01	.7		64.7	64.7	
S.D.	14495.2	81.0	.970	.004	.07	8.2	.13	8.7		14.7	15.0	

A aplicação prática mais interessante dessa métrica comum é chamada mapa do construto ou mapas de respostas esperadas (Smith, 2004). O mapa do construto dos itens da prova de Psicologia do ENADE-2006 é apresentado na Figura 2. Nessa figura, a escala das medidas das competências dos sujeitos é desenhada na linha horizontal (na parte inferior da figura), variando de -5 a $+5$. Os valores T, S, M indicam o ponto em que encontramos a média (M), um desvio padrão (S) e dois desvios padrão (T) das medidas de competência dos 26.613 estudantes. Os números acima dessas letras indicam o número de estudantes em cada nível da escala, por exemplo, a partir deles podemos dizer que 2.924 estudantes tiveram thetas iguais à média ($-0,09$).

Os números 0 : 1 na escala indicam os escores esperados nos itens (cujos nomes e áreas de conteúdo que avaliam estão escritos ao lado), associados aos níveis de competência na escala de -5 a $+5$. O ponto “:” indica o ponto de transição

entre 0 e 1, isto é, o ponto em que, quando se move da esquerda para a direita, o acerto ao item (1) passa a ser mais provável que o erro (0). Em itens dicotômicos, indicam-se os escores 0 e 1 e nas questões dissertativas 1 a 9. Nessas últimas, os pontos de transição “:” indicam onde 1 passa a ser mais provável que 0, 2 mais provável que 1 e assim sucessivamente. Os itens estão ordenados de baixo para cima, em ordem crescente de dificuldade. Pode-se observar, por exemplo, que no item mais fácil, o acerto passa a ser mais provável a partir do nível $-2,4$. Se considerarmos o nível $-0,9$ e colocarmos uma linha perpendicular nesse ponto até a linha superior, podemos observar o cruzamento dessa linha com os escores esperados (buscando em quais itens a linha passa ao lado direito do ponto “:”) e prever que os estudantes nesse nível de competência acertariam os cinco itens mais fáceis (CE12, CE20, CE21, CE36, CE25) e teriam escore 1 nas três dissertativas.

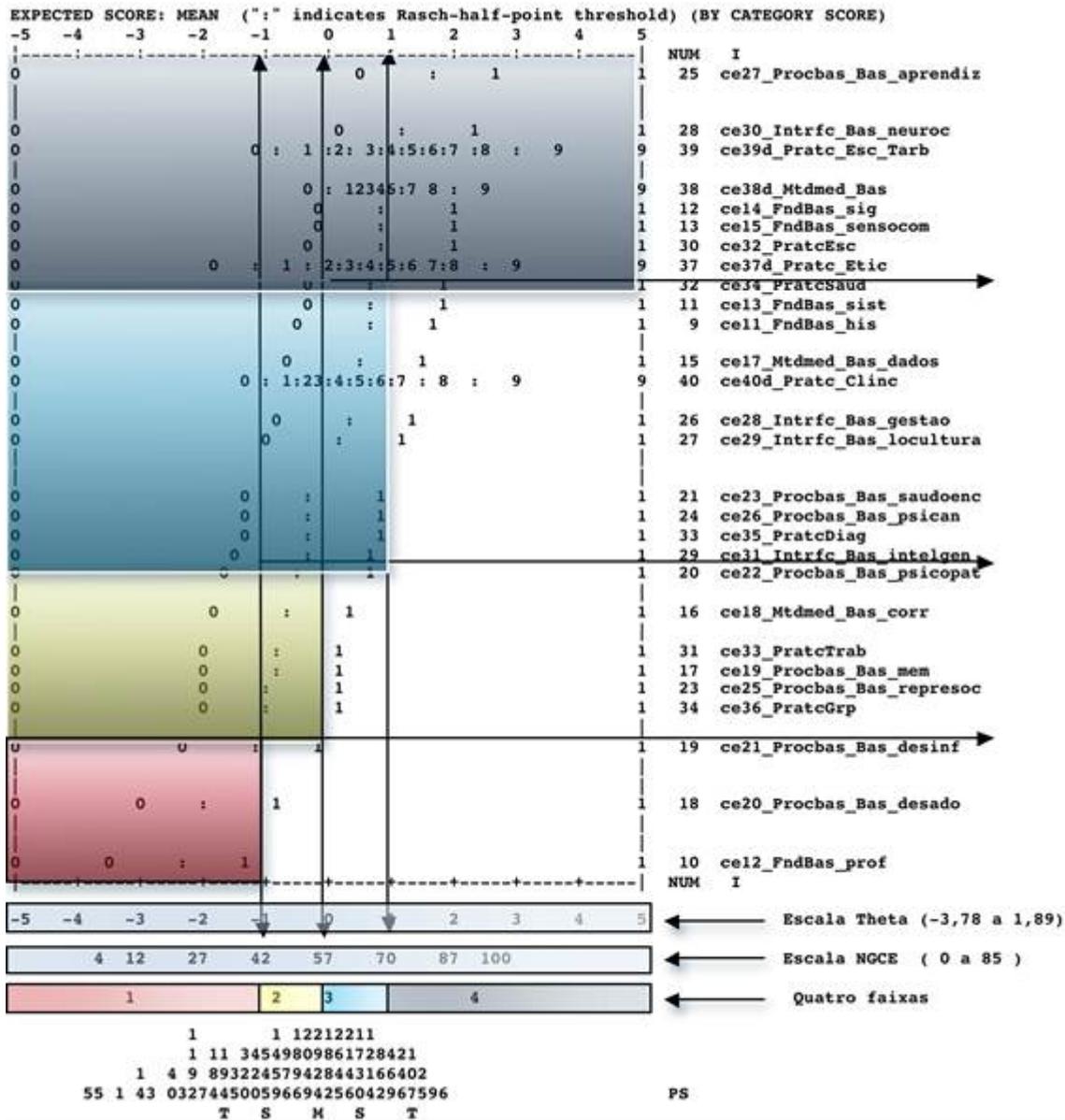


Figura 2 - Mapa dos itens da prova de Psicologia do ENADE-2006.

A partir desse mapa, é possível analisar a relação entre o nível de competência e em quais questões se espera que ocorra o acerto. A análise de conteúdo dessas questões permite compreender melhor o que cada nível significa em termos de competências/habilidades e conteúdos. Esse mapa possibilita a criação de faixas sobre a escala com sentido mais rico e qualificado do ponto de vista psicológico e educacional. A definição desses pontos é chamada *standard setting* na literatura especializada (Cizek, 1996; MacCann & Stanley, 2006). Essa análise permite, portanto, em adição à interpretação referenciada à norma, atualmente

usada no ENADE, a implementação de interpretações referenciadas ao critério (Smith, 2004) que, conforme sugerido em outro trabalho (Primi, Fernandez & Ziviani, 2003) parece ser bem mais adequada ao propósito do ENADE. Também ajuda a suplantar os problemas da métrica arbitrária existente nas ciências humanas (Embretson, 2006).

Seguindo esse procedimento, foi feita uma análise exploratória do mapa de itens procurando identificar pontos significativos associados a blocos de questões ao longo da escala e que pudessem definir níveis diferenciados de competência. Nessa análise, procurou-se definir pontos que

demonstrassem mudanças significativas em termos das competências implicadas nas questões. Nas regulamentações associadas ao SINAES, há a proposição de uma escala de cinco níveis. Seria, portanto interessante se fosse possível atender simultaneamente o critério de sentido substancial e interpretabilidade (Smith, 2004) e o atendimento a essa proposição. Entretanto, conforme relatado anteriormente, o índice de separação (a precisão das medidas das competências dos estudantes) indicou que a prova tem precisão para distinguir de dois a três estratos significativamente diferentes. Portanto, a precisão atingida pela prova torna difícil justificar cinco faixas de desempenho.

A partir da análise qualitativa e levando-se em conta os critérios anteriores, foram definidos três pontos, -1,0, -0,20 e 0,9 (na escala theta), dividindo a escala em quatro faixas que são indicadas por retângulos limitados pelos pontos de corte. Com isso é possível verificar quais questões

são mais prováveis de serem respondidas corretamente à medida que se progride no nível de competência implicado pelas quatro faixas. Os escores em Theta também foram transformados para uma escala mais simples de 0 a 100 (escala NGCE, Nota Geral no Componente Específico). Se um estudante acertasse todas as questões da prova, poderia obter 60 pontos (contando 9 pontos por questão dissertativa). A máxima pontuação encontrada foi 51, isto é, pontuação correspondente a 85% da pontuação máxima. Com base nisso, a escala theta foi convertida para a escala NGCE, para variar de 0 a 85. Os valores nessa escala, correspondentes aos pontos de corte, foram 42, 54 e 70. Também foram incluídas no gráfico as quatro faixas de desempenho criadas a partir da divisão da escala em razão desses quatro pontos. Com base nessas informações, o que podemos dizer dos estudantes localizados nas quatro faixas de desempenho?

Faixa I

Nesse nível, o estudante tende a acertar os itens CE12, CE20 e CE21 (este último está no limite). A análise dessas questões mostra que elas requerem habilidade elementar de compreensão de textos já que a resposta correta pode ser facilmente descoberta pela leitura do enunciado. Além disso, as alternativas erradas são claramente contraditórias à alternativa certa. Em termos de conteúdo, essas questões fazem exigência a um conhecimento genérico sobre a importância da expansão das práticas psicológicas para além da clínica individual e importância dos fatores socioculturais no desenvolvimento humano. Nas questões dissertativas, o estudante tem desempenho esperado 0 ou 1, ou seja, insuficiente. Portanto esse nível sugere um nível de competência insuficiente esperado para o concluinte.

Faixa 2

Nesse nível de competência, o estudante tende a acertar, além dos itens da faixa anterior, os itens CE36, CE25, CE19, CE33, CE18 e CE22 e a ter notas de 0-1 nas questões CE37, CE38 e CE39 e 0-3 na questão CE40. Nesse nível de competência, as questões requerem conhecimentos mais específicos sobre: processos básicos (memória e emoção), sistema de classificação diagnóstica, representação social, elementos da prática profissional envolvendo grupos e no contexto organizacional e do trabalho. Portanto, nesses itens, o estudante precisaria mobilizar os conhecimentos básicos adquiridos nessas áreas para saber interpretar o que era perguntado e distinguir qual alternativa de resposta era correta. Diferentemente do nível anterior, nesses itens era relativamente mais difícil chegar à resposta certa somente pela leitura do enunciado. Aqui já começa a ser necessária a mobilização de conhecimentos adquiridos no curso tais como, manejo de grupos, análise do campo e contexto de atuação, análise de processos psíquicos e sua relação com fenômenos sociais. No entanto, esses conhecimentos são requeridos em um nível básico sem muito aprofundamento. Porém, nas questões dissertativas, esse nível de competência corresponde a um desempenho muito baixo. Por exemplo, na questão CE40, que apresenta um caso clínico e exige que o estudante: analise a necessidade de atendimento psicológico, elabore uma hipótese diagnóstica e planeje uma proposta de avaliação e ou intervenção, o desempenho no nível 0-4 indica: “referência somente a um ou dois pontos abordados superficialmente ou apenas citados; problemas de linguagem; repetição de informações do enunciado, ou menção a dados que não permitem traçar hipóteses; julgamento moral ou apontamento de soluções incompatíveis”. Portanto, nesse nível de competência, observa-se o domínio de conhecimentos básicos, mas ainda não se evidencia a capacidade de mobilizar tais conhecimentos para oferecer soluções a problemas básicos da prática profissional, como a atuação em um caso clínico. Assim sendo, indica um de nível competência mínimo, porém ainda insuficiente.

Nesse nível de competência, o estudante tende a acertar, além dos itens das faixas anteriores, os itens CE31, CE35, CE26, CE23, CE29, CE28, CE17, CE11, CE13 e CE34 e a ter notas 4 a 6 na questão CE40, 2 a 5 na questão CE37 e CE38 e 2 a 4 na questão CE39. Assim como no nível anterior, a resolução das questões indicadas nesse nível requer a mobilização de conhecimentos específicos. Entretanto nesse nível é necessário um conhecimento mais aprofundado em termos de abrangência, contexto de aplicação e, especialmente, inter-relação de conceitos provenientes de diferentes domínios/abordagens. Uma característica das questões desse nível é a junção de diferentes subáreas da psicologia na análise de um tema em diferentes contextos, como, por exemplo: bases genéticas e evolutivas do comportamento e inteligência; psicodiagnóstico em um contexto jurídico; tensão normal–patológico no contexto escolar tratando-se do diagnóstico psiquiátrico (TDAH), psicanálise e neurociência; indivíduo, cultura e sociedade nos âmbitos organizacionais. Também há questões envolvendo interpretação de dados estatísticos. Nesse nível de competência observa-se um domínio mais abrangente dos conhecimentos básicos da psicologia. Esse avançado implica no entendimento de múltiplas visões do mesmo fenômeno a partir de enfoques diferenciados. Essa é uma marca do conhecimento em desenvolvimento da psicologia e é característico do nível mais alto de complexidade dessas questões. Assim o estudante que atinge esse nível é mais hábil em mobilizar conhecimentos adquiridos ao longo de sua formação para identificar os referenciais teóricos subjacentes e interpretar as informações em seu contexto escolhendo a alternativa de resposta mais coerente com proposições teóricas daquele referencial. Nas questões dissertativas esse nível de competência indica uma nota relativamente mais alta (4 a 6) na questão CE40 mencionada anteriormente como um problema clínico. Esse nível de desempenho indica uma resposta que atende parcialmente ao que foi solicitado com nível de elaboração simples. Nas outras questões esse nível indica um desempenho insuficiente. Isso sugere, portanto, que nesse nível de desempenho, embora haja um nível de conhecimento superior, ainda não fica claramente evidenciada a capacidade de mobilização desses conhecimentos de modo a pensar e redigir uma solução satisfatória para problemas da prática profissional que, na prova, associavam-se a condução casos na clínica, a solução de dilemas éticos relativos à prática profissional, propostas de projeto de atuação multiprofissional em comunidades e elaboração de projetos de pesquisa.

Em geral, pode-se qualificar essa faixa como um nível mínimo esperado de competência para um estudante concluinte. Se por um lado esse nível ainda não reflete um nível de maestria desejável, por outro, associa-se ao domínio mais elevado de conhecimento especializado. Ao lado disso, essa expectativa é condizente com o fato reconhecido que, ao término da graduação, o profissional recém formado médio não está suficientemente preparado para atuação prática requerendo, após a formação na graduação, continuidade em atividades supervisionadas, cursos de extensão e pós graduação para atingir um nível suficiente de competência profissional. Em suma, sugere-se que se qualifique esse nível de competência como um ponto mínimo esperado para estudantes concluintes ainda que não reflita um ponto ideal.

Faixa 4

Nesse nível de competência, além dos itens das faixas anteriores, o estudante tende a acertar os itens mais difíceis CE32, CE15, CE14, CE30 e CE27 e a ter notas mais próximas de um desempenho satisfatório nas questões dissertativas (acima de 6 nas questões CE38 e CE40, acima de 5 na questão CE37 e acima de 4 na questão CE39). Esse é o nível mais alto de competência em que é esperado que os estudantes acertem as questões mais complexas. A análise das questões mostra que, como no nível 3, elas também exigem conhecimentos específicos diversos, mas com uma complexidade ainda maior que a anterior já que a maioria das questões contrasta dados empíricos com explicações teóricas diversas para esses dados. Assim, nesse nível, o aluno tem que mobilizar conhecimentos específicos para interpretar dados e raciocinar sobre a coerência entre os dados e diferentes hipóteses explicativas desses dados. Isso ocorre tanto em um âmbito teórico, envolvendo hipóteses explicativas alternativas para o mesmo fenômeno, quanto na interpretação de dados de pesquisa como em delineamentos fatoriais nos quais se combina mais de uma variável independente. Esse nível também sugere a capacidade para interpretação crítica da realidade, desviando-se de explicações mais simples de senso comum. Em razão dessa necessidade de identificação de ideias (argumentos teóricos e dados empíricos) ou um conjunto de itens de informação e do estabelecimento de relações comparativas entre elas, o nível habilidade de leitura e raciocínio dessas questões é bem mais elevado. Esse nível reflete textos típicos na literatura psicológica em que há a discussão argumentativa de abordagens explicativas alternativas para o mesmo fenômeno.

Nas questões dissertativas, o desempenho esperado nesse nível é adequado na maioria das questões, indicando que o aluno consegue mobilizar seus conhecimentos, pensar e redigir propostas de atuação prática satisfatórias. Portanto, sugere-se que esse nível de competência seja qualificado como nível de competência superior ao valor mínimo esperado.

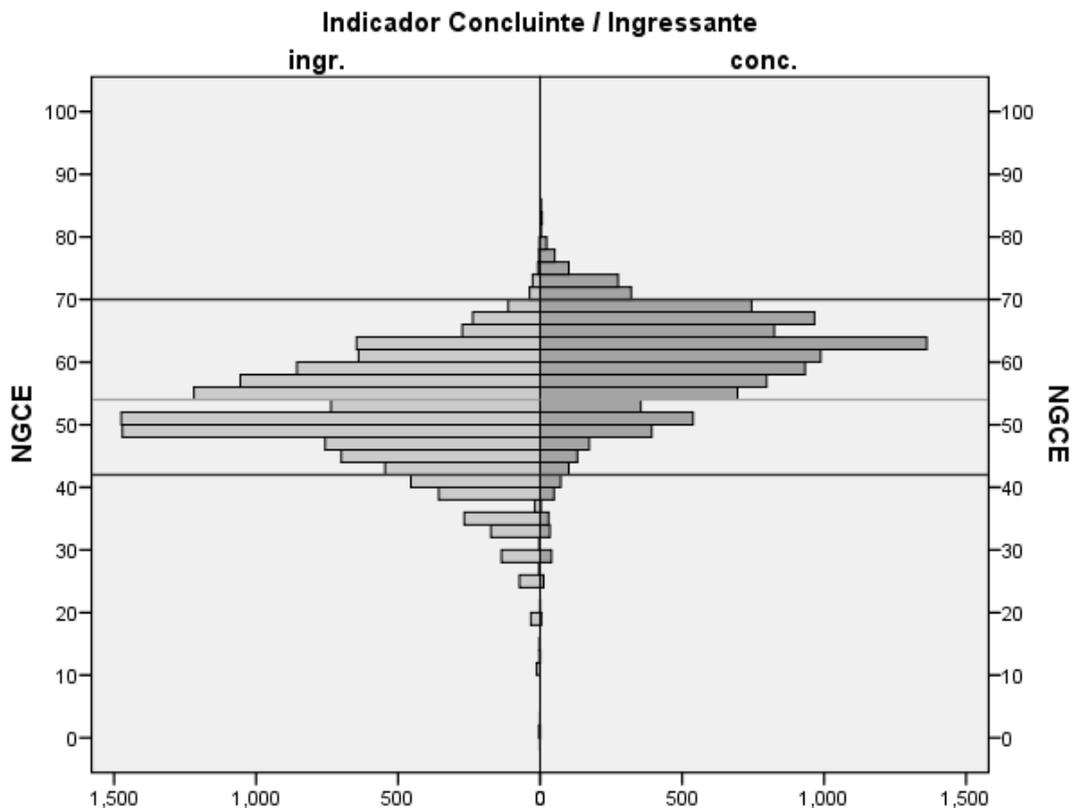


Figura 3 - Distribuição das notas padronizadas dos estudantes ingressantes e concluintes nas quatro faixas de desempenho.

Uma vez definidos esses níveis mais significativos em termos de competência/habilidades/conteúdos, surge a próxima questão: O que essa escala nos diz sobre o nível de competência dos recém formados em Psicologia? Na Figura 3, está apresentada a distribuição dos escores dos estudantes ingressantes e concluintes na escala de 0 a 85.

A diferença é altamente significativa ($t = -75,84$, $p < 0,0001$, $\text{Eta}^2 = 0,204$, $D = 0,909$). Na Tabela 6, está apresentada a proporção de estudantes ingressantes e concluintes em cada uma das quatro faixas. Conforme se constata, a maior diferença é observada nos níveis 2 e 3, sendo que entre os concluintes diminui-se a proporção de estudantes no nível 2 e aumenta-se o nível 3.

Tabela 7 - Distribuição dos estudantes ingressantes nos quatro níveis de competência definidos pela análise do mapa de itens.

Estudantes	Níveis de Competência – NGCE				Total	
		Insuficiente	Fraco	Comp. mínima		Comp. superior
Ingressantes	f	1534	5674	5056	77	12341
	%	12.4	46.0	41.0	0.6	
Concluintes	f	248	1685	7324	784	10041
	%	2.5	16.8	72.9	7.8	
Total	f	1782	7359	12380	861	22382
	%	8.0	32.9	55.3	3.8	

Na Tabela 8, apresenta-se a distribuição de estudantes ingressantes e concluintes nas quatro faixas de competência de cursos com diferentes conceitos tradicionais segundo o sistema adotado pelo INEP. Como pode ser observado, há uma associação entre os dois sistemas, especialmente em grupos de estudantes de cursos 4 e 5 em que a faixa superior aumenta para 11,3% e 28,9%.

Na Figura 4, são apresentadas as médias na escala NGCE dos estudantes ingressantes e concluintes em razão do conceito do curso (metade superior) e do IDD (metade inferior). Na Tabela 8, são apresentadas as ANOVAS 2 (momentos) X 5 (conceitos), testando a significância estatística das diferenças entre as médias na nota NGCE, dividindo-se o grupo dois momentos (ingressante e concluinte) e em grupos de estudantes de cursos com conceitos diferenciados (Conceito e IDD), bem como a interação entre estas duas variáveis,

demonstrando que todos os efeitos foram significativos.

Como pode ser notado, as médias aumentam em razão do conceito do curso de origem do estudante, mas dividindo-os pelo conceito tradicional, as médias dos ingressantes aumentam na mesma proporção, conforme já havia sido notado em outro estudo com os dados do Exame Nacional de Cursos (Landeira & Primi, 2002). Assim, as diferenças concluintes-ingressantes não são maiores nos cursos com conceito mais alto. De fato, nos cursos com conceito 5, a diferença entre concluinte ingressante é menor nos cursos 4 e 3. Já o IDD mostra outra configuração dos dados. Como esse índice foi construído para indicar diferenças entre ingressantes e concluintes, a magnitude dessas diferenças é diretamente refletida nos conceitos que passa a ser progressivamente maior de modo que se aumenta de 1 para 5.

Tabela 8 - Distribuição dos estudantes ingressantes nos quatro níveis de competência - análise do mapa de itens e conceito do curso.

Conceito/Estudante	Níveis de Competência - NGCE				Total		
	Insuf.	Fraco	Comp. mín.	Comp. sup.	Insuf.		
1	Ingressante	f	5	44	52	3	104
		%	4.8	42.3	50.0	2.9	
	Concluinte	f	12	47	76	17	152
		%	7.9	30.9	50.0	11.2	
2	Ingressante	f	172	536	315	1	1024
		%	16.8	52.3	30.8	0.1	
	Concluinte	f	70	315	700	86	1171
		%	6.0	26.9	59.8	7.3	
3	Ingressante	f	737	2502	1743	23	5005
		%	14.7	50.0	34.8	.5	
	Concluinte	f	132	1002	4019	248	5401
		%	2.4	18.6	74.4	4.6	
4	Ingressante	f	253	1118	1280	16	2667
		%	9.5	41.9	48.0	.6	
	Concluinte	f	30	275	2267	328	2900
		%	1.0	9.5	78.2	11.3	
5	Ingressante	f	4	44	206	15	269
		%	1.5	16.4	76.6	5.6	
	Concluinte	f	1	27	208	96	332
		%	.3	8.1	62.7	28.9	
Total	Ingressante	f	1171	4244	3596	58	9069
		%	12.9	46.8	39.7	.6	
	Concluinte	f	245	1666	7270	775	9956
		%	2.5	16.7	73.0	7.8	

Tabela 9 - Estatísticas da ANOVA testando as diferenças - média de desempenho no ENADE - conceito e momento.

Fonte	Soma Quadrática	gl	Média Quadrática	F	Sig.	Eta Parcial ao quadrado
Conceito						
Modelo	897198,4	9	99688,7	1051,3	0,001	0,31
in_grad	75791,1	1	75791,1	799,3	0,001	0,03
Conceito	458693,0	4	114673,2	1209,3	0,001	0,19
in_grad X conceito	23491,8	4	5872,9	61,9	0,001	0,01
Erro	1916101,3	20208	94,8			
Total	2813299,7	20217				
IDD						
Modelo	449393,5	9	49932,6	504,4	0,001	0,19
in_grad	173778,8	1	173778,8	1755,4	0,001	0,08
idd_conc	60833,2	4	15208,3	153,6	0,001	0,03
in_grad X idd_conc	32752,0	4	8188,0	82,7	0,001	0,01
Error	1898014,1	19173	98,9			
Total	2347407,6	19182				

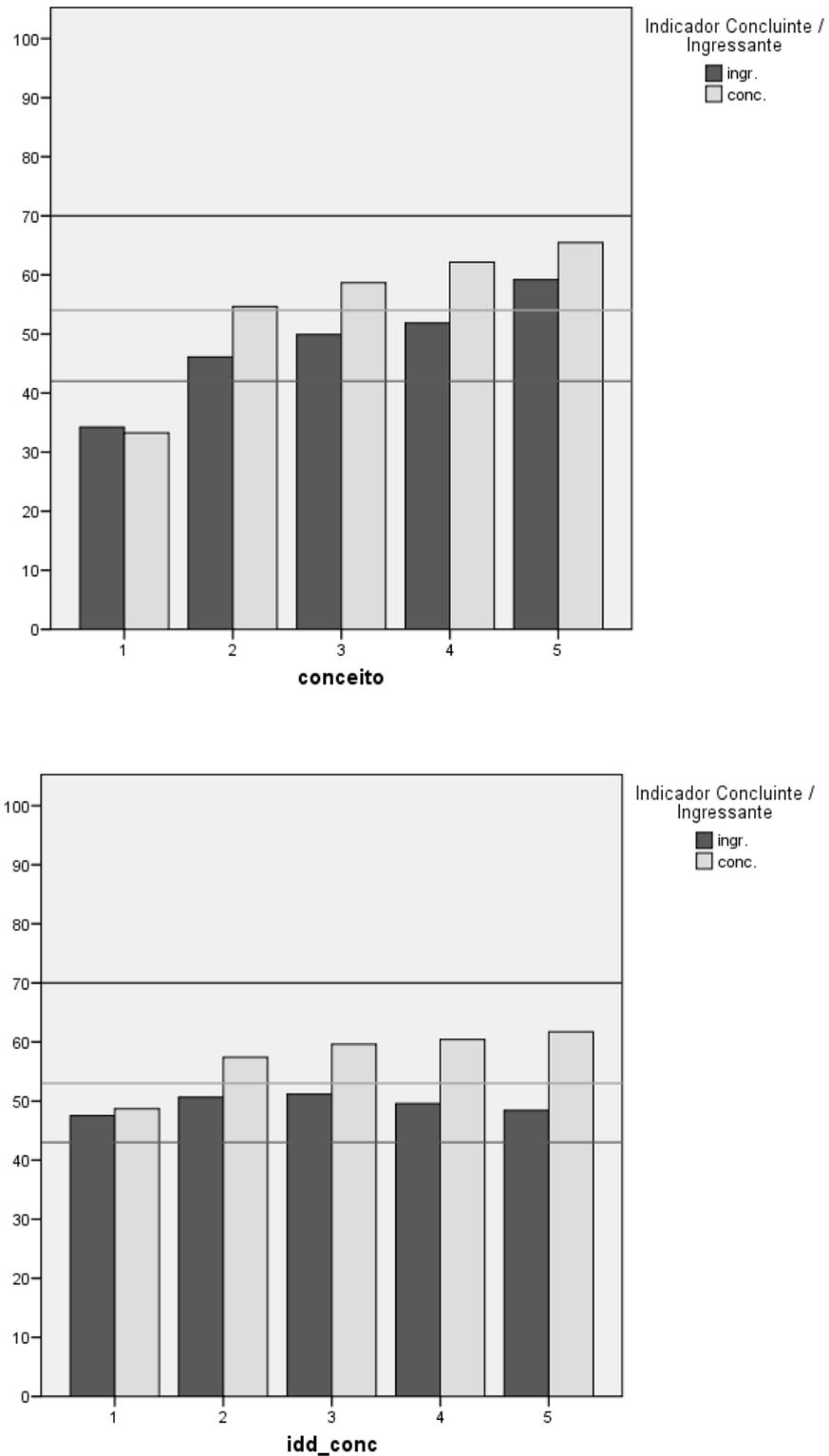


Figura 4 - Médias na escala NGCE em razão do conceito do curso e do IDD.

Em geral, este estudo ilustra que métodos derivados da Psicometria podem trazer avanços significativos não só ao ENADE de Psicologia, mas também para as provas educacionais em geral. A

primeira análise empírica dos dados da prova teve por objetivo verificar quais dimensões do construto foram avaliadas pela prova por meio da análise fatorial dos itens. A segunda análise de estimação

dos parâmetros da TRI buscou definir a métrica da escala geral subjacente aos itens. De maneira geral, é possível observar que a análise dos mapas de itens com base no modelo de Rasch permite o desenvolvimento de parâmetros interpretativos mais informativos a respeito das competências, habilidades e conteúdos que os escores na prova refletem. Especialmente os dados da Tabela 8 em conjunto com as interpretações propostas para as faixas são bastante úteis para descrever o nível de competência dos estudantes. Isso poderia ser apresentado para cada curso, fomentando o entendimento de seus estudantes. Por último se apresentam informações sobre a validade da prova em diferenciar estudantes ingressantes e concluintes. Tais dados corroboram padrões anteriormente encontrados de que provas desse tipo tem boa capacidade de indicação das diferenças entre estudantes concluintes e ingressantes e também mostram a grande influência que os resultados do curso sofrem em razão do perfil dos estudantes ingressantes. Em suma, este trabalho procurou descrever o processo de construção das provas, e aplicar procedimentos de análise para testar sua qualidade e tentar extrair o máximo de informação possível, dadas as limitações que sempre ocorrem ao se tentar criar um instrumento de medida.

Nessa perspectiva, ressalta-se a relevância de estudos como esse que se propôs a aprofundar a análise da representatividade do construto, além da interpretação e implicação dos resultados. Entende-se que a aplicação dos métodos aqui tratados possibilitam construir interpretações mais ricas e sobre o desempenho dos estudantes. Ressalta-se também a necessidade de controle de outros fatores e variáveis que podem interferir diretamente no desempenho dos estudantes (como variáveis pessoais e de contexto). Portanto, esse e outros trabalhos trazem informações mais fundamentadas para criticar e reformular sistemas avaliativos, tanto com base nos itens propostos, quanto em referência aos níveis variados de competências.

REFERÊNCIAS

Bastos, V. A. (2005). Diretrizes curriculares para a reforma curricular do curso de Psicologia. As diretrizes curriculares e as competências profissionais: os desafios da implementação.

Análise Psicométrica do ENADE de Psicologia 2006 293
Em *Associação Brasileira de Ensino em Psicologia (ABEP)*. I Seminário Mineiro de Formação em Psicologia, Belo Horizonte. ABEP.

Bock, D., Gibbons, R. & Muraki, E. (1986). Full-information item factor analysis. *Applied Psychological Measurement*, 12(3), 261-280.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Cronbach, L. J. (1996). *Fundamentos da testagem psicológica*. Porto Alegre: Artes Médicas.

Cizek, G. J. (1996). Standard-Setting Guidelines. *Educational Measurement Issues and Practices*, 15(1), 13-21.

Embretson, S. E. (2006). The Continued Search for Nonarbitrary Metrics in Psychology. *American Psychologist*, 61, 50-55.

Gottfredson, L. & Saklofske, D. H. (2009). Intelligence: foundations and issues in assessment. *Canadian Psychology*, 50(3), 183-195.

Landeira-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.

Limana, A. & Brito, M. R. F. (2005). O modelo de avaliação dinâmica e o desenvolvimento de competências: algumas considerações a respeito do ENADE. *Revista Avaliação*, 10(2), 9-32.

Linacre, J. M. & Wright, B. D. (1991). *Winsteps - Rasch-Model Computer Programs*. Chicago: MESA Press.

MacCann, R. G. & Stanley, G. (2006). The use of Rasch Modeling to improve Standard Setting. *Practical Assessment Research and Evaluation*, 11(2). Retirado em 12 de maio de 2008 de: <<http://pareonline.net/pdf/v11n2.pdf>>.

McGrew, K. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1-10.

Perrenoud, P. (1997). *Construir as competências desde a escola*. Porto Alegre: Artes Médicas.

- Primi, R. (2003). Inteligência: Avanços nos Modelos Teóricos e nos Instrumentos de Medida. *Avaliação Psicológica*, 1(2), 67-77.
- Primi, R. (2004) Avanços na Interpretação de Escalas com a Aplicação da Teoria de Resposta ao Item. *Avaliação Psicológica*, 3(1), 53-58.
- Primi, R. (2006). Evidências de Validade das Provas do ENADE-2004. Em D. Ristoff; A. Limana & M. R. Brito. (Org.). *ENADE: perspectiva de avaliação dinâmica e análise de mudanças*. (pp. 59-76), Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP.
- Primi, R. & Almeida, L. S. (1998) Considerações sobre a análise factorial de itens com resposta dicotômica. *Psicologia: Teoria, Investigação e Prática*, 3, 225-234.
- Primi, R., Fernandez, J. L. & Ziviani, C. (2003). O provão de psicologia: objetivos, problemas, consequências e sugestões. *Psicologia. Teoria e Pesquisa*, 19(2), 109-116.
- Primi, R. & Muniz, M. (2004). *PROJETO 991223. Relatório da análise psicométrica das provas da Avaliação de Aprendizagem 2004 – 1a e 2a unidades e sua correspondência com os testes-piloto*. Salvador: Agência de Avaliação, UFBA/ISP/FAPEX.
- Primi, R., Santos, A. A. A., Vendramini, C. M. M., Taxa, F., Franz, A. M., Lukjanenko, M. F. & Sampaio, I. S. (2001). Competências e Habilidades Cognitivas: diferentes definições dos mesmos construtos. *Psicologia: Teoria e Pesquisa*, 17(02), 01-09.
- Smith Jr., E. V. (2004). Metric Development and Score Reporting in Rasch Measurement. Em: Smith Jr., E. V. & Smith, R. M. (Eds.). *Introduction to Rasch Measurement: Theory, Models, and Applications*. (pp343-365) Minnesota: JAM Press, Maple Glover.
- Wilson, D. T., Wood, R. & Gibbons, R. (1991). *TESTFACT: Test scoring, item statistics, and item factor analysis*. Chicago, IL: Scientific Software International.
- Woodcok, R. W. (1999). What can Rasch-based scores convey about person's test performance? Em S. E. Embretson & S. L. Hershberger (Orgs.). *The new rules of measurement*. (pp. 105-127). Mahwah, NJ: Lawrence Erlbaum.
- Wright, B. D. & Masters, G. N (1981). A model for Partial Credit Scores. Research Memorandum, 31, Mesa. Psychometric Laboratory, University of Chicago.
- Wright, B. D. & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.
- 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).

ⁱ É contra intuitivo o significado 50% de acerto em um item que só pode ser certo ou errado. O fato é que o modelo e as probabilidades correspondem a uma idealização teórica. Um exemplo usado para facilitar o entendimento concreto dessa ideia é que se tivéssemos 10 itens com a mesma dificuldade um sujeito com habilidade igual a dificuldade desses itens tenderia a acertar 5 e errar 5, o que levaria a um escore total de 5 (50%).