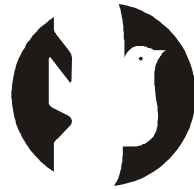


KARINE SILVA OLIVEIRA



UNIVERSIDADE
SÃO FRANCISCO

AVALIAÇÃO DO EXAME NACIONAL DE DESEMPENHO
DO ESTUDANTE PELA TEORIA DE RESPOSTA AO ITEM

ÍTATIBA
2006

KARINE SILVA OLIVEIRA

AVALIAÇÃO DO EXAME NACIONAL DE DESEMPENHO
DO ESTUDANTE PELA TEORIA DE RESPOSTA AO ITEM

Dissertação apresentada ao Programa de Pós-Graduação Stricto Sensu em Psicologia da Universidade São Francisco para obtenção do título de Mestre.

ORIENTADOR(A): CLAUDETTE MARIA MEDEIROS VENDRAMINI

ITATIBA
2006

UNIVERSIDADE SÃO FRANCISCO
PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU *EM PSICOLOGIA*
MESTRADO

AVALIAÇÃO DO EXAME NACIONAL DE DESEMPENHO
DO ESTUDANTE PELA TEORIA DE RESPOSTA AO ITEM

Autora: Karine Silva Oliveira

Orientadora: Claudette Maria Medeiros Vendramini

Este exemplar corresponde a redação final da dissertação de mestrado defendida por Karine Silva Oliveira para avaliação da comissão examinadora.

Itatiba, Setembro de 2006.

COMISSÃO EXAMINADORA

Claudette Maria Medeiros Vendramini

Amir Limana

Ricardo Primi

ITATIBA
2006

AGRADECIMENTOS

Agradeço primeiramente a Deus pelas conquistas até aqui alcançadas e pela força para superar todos os obstáculos.

Agradeço a minha família, meu pai e minha mãe, pelo apoio e incentivo que sempre me deram ao longo da vida. Pelos conselhos nos momentos difíceis e pela vibração nos momentos de alegria.

Ao Cleber que tem sido de fundamental importância nos últimos anos, por todo amor que me tem dedicado, por entender a minha ausência e por esperar com paciência.

Aos professores do programa de mestrado da Universidade São Francisco que tanto contribuíram com seus ensinamentos e em especial a Prof^ª. Claudette Vendramini pela orientação.

Aos novos amigos que conquistei ao longo do programa e aos velhos amigos pelo apoio mesmo à distância.

RESUMO

Oliveira, K. S. (2006). *Avaliação do Exame Nacional de Desempenho do Estudante pela Teoria de Resposta ao Item*. Dissertação de Mestrado, Programa de Pós Graduação Stricto Sensu em Psicologia, da Universidade São Francisco, Itatiba, SP.

O Exame Nacional de Desempenho do Estudante-ENADE foi criado com o intuito de obter informações sobre a evolução dos processos educativos em cada área do conhecimento e suas relações com a totalidade. Neste processo, foram incorporadas questões relacionadas às habilidades e às competências adquiridas pelos estudantes ao cursarem o Ensino Superior. O objetivo deste trabalho é analisar as propriedades psicométricas do ENADE por meio da Teoria de Resposta ao Item-TRI. Foi utilizado neste estudo, um banco de dados que contém os resultados do ENADE 2004 de estudantes do curso de Medicina de todo país ($n = 8.124$). As análises foram realizadas apenas com as questões objetivas tanto do componente de Formação Geral quanto do componente de Formação Específica, totalizando 28 itens. Ao ser analisada a dimensionalidade dos componentes do ENADE, verificou-se a pouca explicação acrescentada por cada uma das dimensões ao modelo, sendo considerado para este trabalho a predominância do primeiro fator. De um modo geral pode-se assumir que os dados se ajustam com maior confiabilidade ao modelo logístico de um parâmetro de Rasch nos dois componentes do ENADE, do que nos modelos de 2 e 3 parâmetros, os quais não atendem aos critérios mínimos estabelecidos.

Palavras-chave: Educação Superior, Habilidade, Avaliação Dinâmica, Competência, Validade de Construto, Fidedignidade.

ABSTRACT

Oliveira, K. S. (2006). *Assesment the National Test of Students' Performance through Item Response Theory*. Master degree dissertation. *Stricto Sensu* Psychology Post Graduation Program, Universidade São Francisco, Itatiba, SP.

The National Test of Students' Performance – ENADE was created in order to obtain information about the evolution of education processes in each area of knowledge and their relation to the total. Matters related to abilities and competences students acquired when they were in college were incorporated in this process. The goal of this study is to analyze the psychometric properties of ENADE through Item Response Theory - ITR. A data bank which contains the results of ENADE 2004 from medicine students from all over the country was used in this study. The analysis were made considering only the objective type questions from the component of General Formation and the Specific Formation, totalizing 28 items. When analyzing the dimensions of ENADE, little explanation from each of the dimensions to the model was found, considering the prevalence of the first factor in this study. In a general way, it can be assumed that the data are adjusted, with a higher reliability, to the logistic model of a Rasch parameter in both components of ENADE than the models of 2 and 3 parameters, which don't respond to the minimum criteria established.

Keywords: advanced education; ability; competence; dynamic evaluation; construct validity; reliability.

SUMÁRIO

LISTA DE FIGURAS	VII
LISTA DE TABELAS	VIII
LISTA DE ANEXOS	IX
INTRODUÇÃO	1
TEORIAS DA INTELIGÊNCIA.....	11
HABILIDADE E COMPETÊNCIA PARA O ENADE.....	14
TEORIAS PSICOMÉTRICAS	21
TEORIA CLÁSSICA DOS TESTES (TCT).....	21
TEORIA DE RESPOSTA AO ÍTEM (TRI).....	25
A TEORIA DE RESPOSTA AO ÍTEM NA AVALIAÇÃO EDUCACIONAL	34
MÉTODO.....	41
FONTE DE DADOS	41
MATERIAL	43
PROCEDIMENTO.....	44
PLANO DE ANÁLISE DE DADOS	45
RESULTADOS	47
CONSIDERAÇÕES FINAIS	63
REFERÊNCIAS	68
ANEXOS.....	75

LISTA DE FIGURAS

FIGURA 1. CURVAS CARACTERÍSTICAS DE ALGUNS ITENS DE UMA PROVA (VENDRAMINI, SILVA & CANALE, 2004).....	30
FIGURA 2. CURVAS CARACTERÍSTICAS DOS OITO ITENS DO COMPONENTE DE FORMAÇÃO GERAL, MODELO DE UM PARÂMETRO	57
FIGURA 3. CURVAS CARACTERÍSTICAS DOS OITO ITENS DO COMPONENTE DE FORMAÇÃO GERAL, MODELO DE TRÊS PARÂMETROS	58
FIGURA 4. CURVAS CARACTERÍSTICAS DOS 20 ITENS DO COMPONENTE DE FORMAÇÃO ESPECÍFICA, MODELO DE UM PARÂMETRO	60
FIGURA 5. CURVAS CARACTERÍSTICAS DOS 20 ITENS DO COMPONENTE DE FORMAÇÃO ESPECÍFICA, MODELO DE TRÊS PARÂMETROS	60

LISTA DE TABELAS

TABELA 1. DISTRIBUIÇÃO DOS ESTUDANTES DE MEDICINA POR REGIÃO	43
TABELA 2. DISTRIBUIÇÃO DOS ESTUDANTES DE ACORDO COM A SITUAÇÃO NO CURSO.....	43
TABELA 3. PARÂMETROS DESCRITIVOS DOS ITENS DO COMPONENTE DE FORMAÇÃO GERAL.	48
TABELA 4. PARÂMETROS DESCRITIVOS DOS ITENS DO COMPONENTE DE FORMAÇÃO ESPECÍFICA	50
TABELA 5. CARGAS NÃO ROTACIONADAS DOS FATORES PRINCIPAIS DO COMPONENTE DE FORMAÇÃO GERAL	52
TABELA 6. CARGAS NÃO ROTACIONADAS DOS FATORES PRINCIPAIS DO COMPONENTE DE FORMAÇÃO ESPECÍFICA.....	54
TABELA 7. PARÂMETROS DOS ITENS DO COMPONENTE DE FORMAÇÃO GERAL, MODELOS DE 1, 2 E 3 PARÂMETROS	56
TABELA 8. PARÂMETROS DOS ITENS DO COMPONENTE DE FORMAÇÃO ESPECÍFICA, MODELOS DE 1, 2 E 3 PARÂMETROS	59
TABELA 9. CARACTERÍSTICAS DO COMPONENTE DE FORMAÇÃO GERAL	61
TABELA 10. CARACTERÍSTICAS DO COMPONENTE – FORMAÇÃO ESPECÍFICA	62

LISTA DE ANEXOS

ANEXO 1. PROVA DO ENADE – COMPONENTE DE FORMAÇÃO GERAL	75
ANEXO 2. PROVA DO ENADE – COMPONENTE DE FORMAÇÃO ESPECÍFICA	80

INTRODUÇÃO

A avaliação educacional ganhou, nos últimos anos, uma atenção especial do governo brasileiro e já são vários os projetos de avaliação elaborados visando as Instituições de Ensino Superior. O Ministério da Educação (MEC) começou a acumular esforços no sentido de avaliar a educação em 1980. Na Educação Superior as primeiras experiências que surgiram tinham a preocupação com o controle da qualidade das Instituições de Ensino Superior, devido ao crescimento exacerbado de instituições e do número de estudantes matriculados. A avaliação atuava como uma forma de prestação de contas dos investimentos feitos pelo setor público (MEC, 2004d).

Em 1993 surgiu o Programa de Avaliação Institucional das Universidades Brasileiras (PAIUB), que se sustentava no princípio da adesão voluntária das universidades e concebia a auto-avaliação como etapa inicial de um processo que se estendia a toda instituição e se completava com a avaliação externa, tendo como preocupação a totalidade da instituição, com todas as suas dimensões e funções. O PAIUB estabeleceu um diálogo entre a comunidade acadêmica e a sociedade a respeito dos objetivos a serem atingidos com a avaliação da Educação Superior (MEC, 2004d). Segundo Bertolin (2004), o PAIUB sofreu muitas críticas e por não ter o objetivo de gerar concorrência ou orientar o financiamento federal, foi considerado não publicável. Por esses motivos, no ano de 2001, o PAIUB foi desconsiderado pelo MEC como programa de avaliação.

A partir de 1995, no governo Fernando Henrique Cardoso, foram criados mecanismos de avaliação que funcionaram como instrumentos de classificação das instituições gerando a concorrência entre elas, por meio de vários instrumentos como o Exame Nacional de Cursos

(ENC), a visita das Comissões de Especialistas (ACE) e a Avaliação Institucional de Centros Universitários (Bertolin, 2004). Do ENC participaram estudantes concluintes do curso de graduação, focava o curso em sua dimensão de ensino, e tinha função classificatória, visando construir bases para uma possível fiscalização, regulação e controle por parte do Estado, baseada na lógica de que a qualidade de um curso é igual à qualidade de seus estudantes (MEC, 2004d). Assim, um curso que recebeu nota “A” não necessariamente significava um curso de boa qualidade, já que a prova não conseguia mostrar o quanto de conhecimento foi adquirido pelo estudante no curso. De acordo com Bertolin (2004) pode-se dizer que um curso que recebeu um estudante “E” e o formou “C” agregou muito mais conhecimento e prestou um serviço social de maior relevância do que um curso que recebeu um estudante “B” e o formou “A”.

O ENC teve como objetivo era orientar as ações do MEC para estimular a melhoria da qualidade de ensino e apoiar as decisões da Secretaria de Educação Superior (SESU) quanto ao reconhecimento e renovação de cursos de graduação. Segundo o MEC (2004d), o ENC possuía dois instrumentos, o primeiro era um exame composto por uma prova de conhecimento constituído de questões de múltipla escolha e/ou questões discursivas conforme adoção das comissões de curso. O segundo instrumento consistia em um conjunto de questionários: o questionário de pesquisa, que colhia informações socioculturais dos graduandos e caracterizava os cursos, a partir do ponto de vista dos graduandos, quanto a recursos e instalações disponíveis, estrutura curricular e desempenho docente; e o questionário sobre impressão das provas, apresentado aos estudantes durante a realização da prova com o objetivo de conhecer a opinião dos participantes a respeito do instrumento aplicado.

Utilizado para verificar a aquisição de conteúdos e medir competências dos estudantes, o ENC avaliava a qualidade dos cursos pelo desempenho dos estudantes no exame, não constituindo uma medida de avaliação da aprendizagem, ou seja, seus resultados refletiam

aproximações muito pobres do que efetivamente ocorre no processo ensino-aprendizagem, não sendo suficientes para identificar as qualidades acadêmicas dos cursos ou instituições. Além disso, o tipo de tratamento estatístico utilizado limitava as possibilidades de interpretação dos resultados e permitia que conclusões errôneas fossem tiradas sobre a qualidade dos estudantes e das instituições de Ensino Superior (MEC, 2004d).

O ENC como instrumento de avaliação utilizado pelo MEC sofreu diversas críticas como a falta de articulação com um conjunto integrado de avaliações com princípios, objetivos e ações claramente definidos; o fato de ter sua motivação mais fora do que dentro da escola, reproduzindo representações equivocadas do mundo acadêmico; seus resultados eram mais voltados à construção da reputação institucional do que à qualidade institucional; desconsiderar o perfil acadêmico do estudante tornando inviável a análise do valor agregado pela instituição aos conhecimentos e habilidades dos estudantes; a ausência de comparabilidade entre as provas ao longo do tempo comprometendo a capacidade de avaliar as perspectivas dos cursos; os boicotes por parte dos estudantes e a falta de critério para lidar com as provas entregues em branco; a divulgação dos resultados desvinculados de outros processos avaliativos, atribuindo ao ENC a centralidade do processo de avaliação e a adoção de políticas de premiação e punição de instituições com base em conceitos gerados por um instrumento e uma metodologia deficientes que não expressam com confiabilidade a qualidade dos cursos (MEC, 2004d).

Além dessas críticas, a administração do ENC tornou-se mais complexa e onerosa em função do aumento crescente do número de instituições, cursos e áreas, tornando os investimentos do Estado incompatíveis com a qualidade dos resultados produzidos. Como consta no documento do MEC (2004d) o ENC além de possuir deficiências técnicas e não orientar a população quanto à real qualidade dos cursos, aumentava a cada ano seu peso administrativo. E mesmo os que argumentavam que seu propósito não era o de avaliar os cursos, mas de verificar

até que ponto os estudantes concluintes haviam atingido normas ou padrões estabelecidos por especialistas, admitiam que esses padrões refletiam aproximações muito pobres do que realmente ocorria no processo de ensino-aprendizagem.

Por estas e outras deficiências, no ano de 2003, o governo de Luis Inácio Lula da Silva designa os membros da Comissão Especial de Avaliação da Educação Superior (CEA) com a finalidade de “analisar, oferecer subsídios, fazer recomendações, propor critérios e estratégias para a reformulação dos processos e políticas de avaliação da Educação Superior e elaborar a revisão crítica dos seus instrumentos, metodologias e critérios utilizados” (MEC, 2004c, p.83).

O Sistema Nacional de Avaliação da Educação Superior (SINAES) foi o produto do trabalho de quatro meses da CEA, integrada por acadêmicos, especialistas da temática da Avaliação e da Educação Superior, por membros do ministério da Educação e representantes da União Nacional dos Estudantes (Barreyro, 2004). As características fundamentais do SINAES são a avaliação institucional como centro do processo avaliativo e a integração de forma harmônica das várias dimensões componentes da avaliação. A proposta do SINAES baseia-se na concepção de avaliação como um processo que vincule a dimensão formativa a um projeto de sociedade comprometido com a igualdade e justiça social (MEC, 2004a).

De acordo com Limana e Brito (2005), o SINAES é parte importante de uma política nacional sobre a educação superior, enfatiza os dinamismos dos processos e relações entre as dimensões externas e internas, particular e global, quantitativo e qualitativo e os diversos objetos e objetivos da avaliação. Tendo em vista essa nova concepção de avaliação como um sistema dinâmico buscou-se um novo modelo de exame considerando o modelo de avaliação dinâmica do potencial de aprendizagem que é mais adequado a esta nova filosofia.

Esse processo de avaliação dinâmica foi incorporado pelo SINAES visando atender às novas exigências do trabalho que requer não só uma flexibilidade técnico-instrumental, como

também a flexibilidade intelectual, tendo em vista as necessidades contínuas dos processos de produção de bens e serviços. Assim, as instituições de Ensino Superior devem preparar seus estudantes para serem profissionais que possuam mais do que saberes escolares ou técnico-profissionais. É necessário que os estudantes desenvolvam e tenham a capacidade de mobilizar esse conhecimento adquirido durante sua formação para a solução de problemas e para o enfrentamento de imprevistos nas situações de trabalho, principalmente porque as novas atividades profissionais são cada vez mais marcadas pela imprevisibilidade das situações, demandando, o tempo todo, escolhas e opções, ampliando as operações mentais e cognitivas envolvidas nas atividades (Souza, 2005).

Um dos instrumentos utilizado pelo SINAES como parte do processo de avaliação da educação superior é o Exame Nacional de Desempenho dos Estudantes (ENADE) que veio substituir o ENC e conforme a Lei nº. 10.861 no art. 5º do MEC (2004b, p. 137-138):

O ENADE tem como 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 específico de sua profissão ligada à realidade brasileira e mundial e a outras áreas do conhecimento. Para isso ele será aplicado aos estudantes de todos os cursos de graduação, ao final do primeiro e do último ano de curso.

O ENADE, de acordo com relatório do MEC (2004b), foi formulado com o objetivo de verificar as competências e habilidades básicas das áreas, os conhecimentos sobre conteúdos básicos e profissionalizantes e verificar o desempenho em questões transdisciplinares, envolvendo o conhecimento mais geral. É examinado o desempenho de estudantes ingressantes e

concluintes porque o que se pretende analisar são as mudanças, os ganhos do estudante ao longo de sua trajetória na Instituição.

O ENADE foi criado com o intuito de obter informações e permitir análises sobre a evolução dos processos educativos em cada área do conhecimento e suas relações com a totalidade, incorporando questões relacionadas ao domínio da aprendizagem e da habilidade dos estudantes. Tem como meta principal acompanhar o processo de aprendizagem e o próprio desempenho do estudante frente aos conteúdos que compõe o currículo dos cursos, bem como suas habilidades e competências, esperando que no final o nível de habilidade dos estudantes atinja níveis mais elevados que no início, sinalizando ganhos para o estudante, o curso e a instituição. Pressupõe-se que o bom desempenho no ENADE indique o que o estudante é capaz de fazer com o conhecimento adquirido na universidade e não apenas “o quê” e “o quanto” ele aprendeu (Vendramini, 2005).

Os estudantes que participaram do exame foram, como descreve Vendramini (2005), selecionados aleatoriamente a partir de uma relação completa de ingressantes e concluintes enviadas pelas instituições de ensino. Segundo a autora, a amostragem foi aleatória, nela todos os elementos da população têm probabilidade conhecida e não nula de pertencerem à amostra e a representatividade da amostra permite que sejam feitas inferências confiáveis de características educacionais da população, possibilitando orientar a tomada de decisões relacionadas aos cursos avaliados.

O aperfeiçoamento dessa nova proposta de exame pode permitir inferências confiáveis sobre o valor que as instituições de ensino agregam aos estudantes e que não eram possíveis serem feitas através do exame anterior. Esse aperfeiçoamento inclui avaliações sucessivas do processo de aprendizagem levando em consideração as habilidades iniciais dos estudantes como também as habilidades efetivas desenvolvidas por eles durante o curso (Vendramini, 2005). Uma

das vantagens de se considerar a mudança educacional em um processo de avaliação é que a real evolução do perfil acadêmico do estudante é considerada na análise (Limana & Brito, 2005).

Para se avaliar a mudança educacional é preciso avaliações sucessivas do mesmo estudante ao longo do tempo. Para Primi (2005), o ENADE 2004 avaliou estudantes ingressantes e concluintes do mesmo ano o que impossibilita interpretações sobre mudanças ou sobre o valor agregado ao estudante no decorrer do curso, mas é possível verificar as diferenças entre concluintes e ingressantes desde que se assuma como verdadeira a suposição de que o nível de desempenho dos estudantes concluintes é o mesmo dos estudantes ingressantes de hoje. Assim, as diferenças entre os concluintes e ingressantes do mesmo ano poderiam ser consideradas indicadoras de mudança. Também neste estudo as interpretações das diferenças entre ingressantes e concluintes serão consideradas indícios de mudanças, não sendo possível ainda fazer afirmações conclusivas sobre essas interpretações como indicadores de mudança.

O ENADE busca aferir o desempenho do estudante, seus resultados podem mostrar se o estudante teve ou não desenvolvimento nas suas competências, se é capaz de trabalhar os conteúdos aprendidos e sua capacidade para usar esse conhecimento no seu desenvolvimento (MEC, 2004d). Seus resultados ainda não permitem aferir o valor agregado, nem apontar os indicativos de ganho, mas com provas equivalentes do ENADE realizadas no início e final do curso, indícios de melhora do desempenho poderão ser analisados, confirmando a idéia de que as questões do exame devem buscar o que o sujeito é capaz de fazer com o conhecimento aprendido na escola. Neste processo, não apenas aspectos relativos aos construtos psicológicos estão envolvidos como também aspectos relacionados à multiplicidade de fatores que interatuam e afetam a ação dos indivíduos (Limana & Brito, 2005).

O ENADE foi aplicado pela primeira vez no dia 07 de novembro de 2004 em treze cursos: Agronomia, Educação Física (licenciatura e bacharelado), Enfermagem, Farmácia, Medicina,

Medicina Veterinária, Nutrição, Odontologia, Fisioterapia, Fonoaudiologia, Serviço Social, Terapia Ocupacional e Zootecnia. Participaram do exame estudantes de todo o país, o curso com maior número de participantes foi educação física e com menor número foi o curso de terapia ocupacional. O Estado de São Paulo foi o que mais inscreveu estudantes no ENADE/2004.

Segundo Andrade, Tavares e Valle (2000) a avaliação possui um importante papel nas políticas de Educação Superior, uma vez que os custos do Ensino Superior se tornam cada vez mais elevados, indispensáveis à informação e à prestação de contas da qualidade e da amplitude dos serviços que essas instituições prestam à sociedade, em ensino, pesquisa e extensão. Uma das preocupações existentes refere-se à metodologia utilizada para analisar os dados obtidos por estas avaliações, e a busca por uma metodologia precisa que permita além da obtenção de resultados por escores brutos ou padronizados, a construção de escalas de habilidades, levando a um acompanhamento do progresso do conhecimento adquirido pelo estudante ao longo do tempo.

A avaliação da qualidade das provas do ENADE passa pela verificação de sua adequação em detectar as diferenças de desempenho entre ingressantes e concluintes, e se a prova detecta variações na magnitude dessas diferenças que poderiam ser atribuídas aos cursos. Existem várias possibilidades de análises e todas deverão ser utilizadas para se escolher aquela que melhor se ajusta ao modelo de avaliação proposto pelo SINAES.

Não é fácil medir variáveis educacionais, e para que esse processo seja confiável e permita generalizações confiáveis de resultados amostrais é fundamental que sejam utilizados instrumentos válidos e fidedignos. Em psicologia, os testes são medidas que devem reunir certas qualidades, como: padronização, precisão e validade. Embora as provas ou testes educacionais não tenham o mesmo significado que testes psicológicos, que se constituem em medidas objetivas e padronizadas para medir algum aspecto de conduta humana, com o progresso alcançado na psicometria, várias técnicas de construção de testes são utilizadas na construção de

exames escolares. A utilização dessas técnicas torna o processo de avaliação educacional mais objetivo.

Na área da Avaliação Educacional, sabe-se que as formas e os procedimentos de avaliação podem adotar diferentes abordagens. Uma das preocupações existentes refere-se à metodologia utilizada para analisar os dados obtidos por estas avaliações, e se considerarmos as medidas objetivas (testes) utilizadas, pode-se fundamentar esta avaliação em dois modelos matemáticos mais conhecidos: a Teoria Clássica dos testes (TCT) e a Teoria de Resposta ao Item (TRI) (Andriola, 1998; Andrade, Tavares & Valle, 2000).

A Teoria de Resposta ao Item (TRI) vem sendo progressivamente introduzida na avaliação educacional por possibilitar análises qualitativas a partir de resultados brutos de uma prova, fornecendo informações mais precisas do desempenho do estudante e da qualidade das questões utilizadas (Vendramini, 2005). A TRI como modelo matemático considera, segundo Fletcher (1994), a interação entre sujeito e item reproduzindo a probabilidade do indivíduo acertar a questão considerando o processo de solução, a dificuldade e o fato de que um item não depende do outro, sendo a pontuação dada em função das respostas e dos parâmetros de cada item.

O trabalho apresentado busca verificar em que medida o ENADE, provas de formação geral e componente específico, enquanto instrumento de avaliação afere sobre o desempenho acadêmico do estudante universitário do curso de Medicina, uma vez que o curso de psicologia não foi avaliado pelo ENADE 2004 e considerando que em algumas universidades do país o curso de psicologia está inserido na área de ciências da saúde. E é pensando no compromisso com a melhoria permanente da qualidade da Educação Superior que se espera contribuir para o aperfeiçoamento do ENADE, permitindo a realização de inferências sobre as mudanças no processo educacional ocorridas nas instituições de Ensino Superior. Assim, este projeto tem

como objetivo analisar as propriedades psicométricas do ENADE, enquanto instrumento de avaliação do curso de graduação, por meio da teoria de resposta ao item, centrando a análise nas questões da prova e não na prova como um todo, com o intuito de fornecer bases para possíveis interpretações dos diferentes níveis de desempenho de universitários ingressantes e concluintes.

Nessa perspectiva compreender como se relacionam as habilidades, as competências dos estudantes podem auxiliar na construção de instrumentos de avaliação mais consistentes e fidedignos. Desta maneira, conceituar os termos habilidade e competência permitirá uma análise mais fundamentada dos resultados da prova.

TEORIAS DA INTELIGÊNCIA

Existe um grande volume de teorias sobre a inteligência, sendo que sua discussão mostra controvérsias e grande diversidade de opiniões entre os teóricos. O conceito de inteligência permanece genérico, ambíguo e de difícil definição. Mas mesmo com essas diferenças teóricas, algumas concordâncias quanto ao significado do termo inteligência podem ser encontradas entre os estudiosos.

Segundo Anastasi e Urbina (2000) a inteligência há muito tempo tem sido foco de interesse de pesquisadores, como também de diversas áreas, em especial na psicologia e na educação. Segundo as autoras, nos manuscritos chineses a inteligência já era associada ao desempenho acadêmico e, no processo de seleção para o serviço civil. A inteligência para solucionar problemas e o conhecimento acadêmico já fazia parte de provas de desempenho utilizadas.

Sternberg (2000), após um levantamento realizado em um Simpósio sobre inteligência em 1921, definiu a inteligência como a “capacidade para aprender a partir da experiência, usando processos metacognitivos para melhorar a aprendizagem, e a capacidade para adaptar-se ao ambiente circundante, que pode exigir diferentes adaptações dentro de diferentes contextos sociais e culturais” (p. 400).

Vários autores propuseram formas diferentes de organizar as teorias existentes. Almeida (1994), por exemplo, agrupou-as de acordo com o enfoque do objeto e com o método de trabalho. Dessa forma obteve três grupos de abordagens: desenvolvimentistas, que tem como seu principal representante Piaget; cognitivistas, com as contribuições de Gardner sobre as Inteligências Múltiplas e de Sternberg com a Teoria Triádica da Inteligência; e fatoriais ou psicométricas, com

o Modelo de Cattell-Horn-Carroll (CHC) a qual mais se aproxima das relações que se estabelecem com este estudo.

Forma parecida a essa, foi proposta por Taylor (1993, citado por McGrew & Flanagan, 1998) resultando, também, três grupos de teorias: psicométricas, processamento de informação e modificabilidade cognitiva. Sternberg (1990), propõe outra forma de arranjo das diversas teorias da inteligência analisando-as através das metáforas que norteiam a sua construção. Sua proposta distribui as teorias em sete grupos: geográfico (psicométricas), computacional, biológico, epistemológico, antropológico, sociológico e sistêmico.

Segundo Munhoz (2004), os autores acima citados, Almeida, Taylor e Sternberg, concordam que existe uma nítida separação das teorias psicométricas das outras por apresentarem características peculiares e um vasto volume de estudos. Ainda de acordo com a autora a abordagem psicométrica também se caracteriza pela busca ou observação das diferenças individuais considerando que a inteligência, varia em intensidade e estas variações podem ser mensuradas por meio de testes psicológicos. As teorias psicométricas ou fatoriais da inteligência são as mais antigas e caracterizam-se, principalmente, pelo uso da estatística e são, por isso, bastante influenciadas pelo desenvolvimento nesta área (McGrew & Flanagan, 1998).

Alguns autores (Sternberg, 2000; Anastasi & Urbina, 2000; McGrew & Flanagan, 1998; Carroll, 1993) consideram que o grande marco da abordagem psicométrica foi a apresentação da Teoria dos Dois Fatores de Inteligência de Charles Spearman (1863-1945). Esse autor é considerado o criador do método da análise fatorial, tendo buscado estabelecer uma matriz de correlações entre o desempenho acadêmico e os resultados dos indivíduos em diferentes testes, encontrando a presença de dois fatores: um amplo, que permeia o desempenho do indivíduo em todas as atividades apresentadas, denominado de Fator Geral e um outro conjunto de fatores de

grupo, específicos às tarefas ou ao grupo de tarefas propostas ao indivíduo, cujo papel, no desempenho intelectual, foi julgado irrelevante, de pouca aplicabilidade.

Para Sternberg (1990), a teoria de Spearman refere-se não a dois fatores de inteligência, mas a dois tipos de fatores que podem estar presentes ao mesmo tempo, enquanto o indivíduo realiza uma determinada tarefa. Assim tem-se apenas um Fator Geral que, de acordo com Spearman, seria a energia mental acionada pelos indivíduos, conforme sua vontade, para a realização de determinadas tarefas. No desenvolvimento de seus estudos, Spearman reconheceu que o Fator Geral poderia não ser suficiente para explicar toda a variância nos testes de habilidade mental. Isso promoveu o início de pesquisas realizadas por seus estudantes buscando identificar dimensões da inteligência, envolvendo fatores de grupo.

Raymond B. Cattell (1905-2003), por volta de 1940, lançou as bases da *Teoria da Inteligência Fluída e Cristalizada*, que englobava uma visão dicotômica da inteligência, diferente da proposta por Spearman, compreendendo dois grandes fatores sendo que, a inteligência fluída estava relacionada ao raciocínio que o sujeito apresenta quando frente a situações novas, enquanto que, a inteligência cristalizada era o resultado do conhecimento e a aculturação acumulada que refletiam as diferenças individuais (McGrew & Flanagan, 1998).

Em 1965, John L. Horn elaborou uma revisão da *Teoria de Inteligência Fluída e Cristalizada* de Cattell, estabelecendo uma estrutura multidimensional hierárquica em dois níveis com fatores gerais e específicos, agregando, também, o papel das habilidades na aprendizagem. A Teoria da Inteligência Fluída e Cristalizada de Cattell e Horn reflete uma visão multidimensional da inteligência em que as capacidades humanas variam de acordo com a tarefa apresentada, isto é, com a quantidade de habilidades e conhecimentos aprendidos ou não no passado (Munhoz, 2004).

John B. Carroll (1916-2003), realizou uma meta-análise de dados de pesquisas sobre a inteligência e seus resultados apontaram para um modelo hierárquico da inteligência envolvendo três níveis, camadas ou estratos de processos mentais, sendo esta teoria conhecida como *Teoria dos Três Estratos*. A estrutura hierárquica encontrada mostrou que há um número finito de fatores em diferentes camadas que contribuem para uma grande parte da covariância encontrada na maioria das medidas de habilidades cognitivas (Carroll, 1993).

De acordo com Munhoz (2004), com o estudo de Carroll encerrou-se um período de buscas por componentes da inteligência e levou a um novo patamar as pesquisas da abordagem psicométrica. A tendência é ampliar o enfoque dos estudos da inteligência englobando variáveis cognitivas e não cognitivas num esquema ou modelo de processamento de informação. Um desses novos modelos é o resultado da união das teorias que apresentam o melhor embasamento empírico para este estudo, a Teoria de Inteligência Fluída e Cristalizada de Cattell e Horn e a Teoria dos Três Estratos de Carroll, dando origem ao que se denominou de *Teoria Cattell-Horn-Carroll das Habilidades Cognitivas*, ou Teoria CHC (Flanagan, McGrew & Ortiz, 2000).

Como já citado anteriormente, os estudos a respeito das habilidades humanas iniciaram-se com os avanços das pesquisas relativas à natureza da inteligência, sendo que as estruturas dessas habilidades estão diretamente relacionadas aos testes de inteligência. Será mostrado a seguir um breve histórico acerca dos estudos das habilidades e competências e os conceitos que melhor estruturam este estudo.

HABILIDADE E COMPETÊNCIA PARA O ENADE

O termo habilidade é encontrado na literatura psicológica desde o final do século XIX. Seus estudos iniciais surgiram a partir de considerações relativas à natureza da inteligência e à estrutura dessas habilidades relacionadas aos testes de inteligência, possuindo um caráter inato segundo muitos autores (Munhoz, 2004).

O trabalho de Spearman deu início ao desenvolvimento e à aplicação da análise fatorial para habilidades humanas. Como citado por Vendramini (2000), uma revisão da literatura sobre habilidades humanas e a aplicação da análise fatorial exploratória aos estudos dessas habilidades foi feito por John B. Carroll (1916-2003), que a partir dos dados obtidos em seu trabalho formulou a teoria dos três estratos, na qual o domínio total das habilidades intelectuais é mostrado em três níveis ou estratos. O terceiro estrato refere-se ao fator geral, o segundo às habilidades gerais de raciocínio, linguagem, memória e aprendizagem, percepção visual, percepção auditiva, habilidade de recuperação, conhecimento e proficiência e velocidade de processamento cognitivo. As habilidades específicas compõem o primeiro estrato, perfazendo um total de 69, e são diretamente observáveis pelo desempenho do sujeito (Carroll, 1993).

Na Teoria dos Três Estratos de Carroll as habilidades apresentam-se em três níveis: o fator geral; as habilidades gerais de raciocínio como memória, linguagem, percepção auditiva e visual, habilidade de recuperação, conhecimento, proficiência e velocidade de processamento cognitivo; e as habilidades específicas, observáveis por meio do desempenho do sujeito (Munhoz, 2004). Flanagan, McGrew e Ortiz (2000), separam as habilidades em habilidades acadêmicas e habilidades cognitivas, sendo as primeiras aquelas consideradas pelos aspectos relacionados ao conhecimento e desempenho acadêmico e as habilidades cognitivas aquelas consideradas características da inteligência.

Algumas questões conceituais têm surgido a respeito das habilidades humanas. Uma dificuldade para conceituá-la deve-se ao fato de existir mais de uma maneira de empregar a

palavra habilidade, podendo ser usada com referência àquilo que uma pessoa pode fazer ou como referência à causa que leva uma pessoa a fazer algo. O importante é que mesmo escolhendo concepções alternativas de habilidades é necessário o uso de palavras específicas e conceitos claramente definidos para a definição deste construto (Vendramini, 2000).

O conceito de habilidade que melhor estrutura este estudo foi apresentado por Brito, Munhoz, Primi, Gonçalves, Rezi, Neves, Sanches, e Marinheiro (2000), na qual a habilidade é uma característica psicológica individual que favorece um domínio rápido e fácil de uma determinada atividade, as habilidades são totalidades e seus componentes são identificados em separado apenas com o objetivo de pesquisa, seu desenvolvimento persiste ao longo do tempo, e em universitários seu desenvolvimento está centrado em tarefas mais específicas. E ainda conforme Primi, Santos, Vendramini, Taxa, Franz, Lukjanenko, e Sampaio (2001), a habilidade é o potencial do indivíduo para realizar determinada tarefa com relativa facilidade, sendo que esse potencial pode ser observado em várias situações, como quando o indivíduo apresenta respostas corretas e conhecimento de determinados conteúdos.

Segundo Carroll (1993), habilidade é a variação individual na força total disponível para a realização de uma classe definida de tarefas. A habilidade indica facilidade em lidar com um modelo de informação, não necessariamente implicando em competência.

Na literatura educacional e psicológica, o termo competência é recente. Começou a ser utilizado a partir de 1970 e está ligado à idéia de desempenho eficiente, ao cumprimento efetivo da ação e à capacidade de realização com êxito de uma determinada tarefa. As competências são desenvolvidas a partir de determinadas habilidades e estão ligadas à idéia de desempenho eficiente podendo ser tratadas como um processo de agir de maneira reflexiva e eficaz em um determinado tipo de situação, apoiado em um conjunto articulado e dinâmico de conhecimentos, habilidades e atitudes (Brito & cols., 2000).

A competência refere-se a um nível esperado de realização para uma determinada atividade. Três aspectos são fundamentais para o desenvolvimento da competência: primeiro, o indivíduo deve ter a habilidade (potencial) para a atividade; segundo, deve passar por experiências adequadas de aprendizagem; e terceiro, deve dedicar-se ou investir nessas experiências. Para que uma habilidade se transforme em competência será necessário investir em experiências de aprendizagem. Sem esse investimento não há competência mesmo que a pessoa tenha habilidade (Primi, 1998; Primi & cols., 2001).

A constatação da competência na academia pressupõe-se um conjunto de critérios estabelecidos referenciados no perfil do estudante que se quer formar, descrito no projeto pedagógico do estabelecimento de ensino. Esses critérios formam a base para o julgamento das competências dos estudantes analisados a partir de seu desempenho acadêmico (Munhoz, 2004).

Segundo Limana e Brito (2005, p.16):

Pode-se afirmar que o ENADE não mede habilidades enquanto um ‘traço’, mas apenas permite verificar um conjunto de componentes (fundado em uma ou mais habilidades) que se manifestam quando o estudante é colocado frente a uma tarefa que demanda o uso da(s) habilidade(s).

Formar competências e desenvolver habilidades pressupõe um trabalho diferenciado das instituições de ensino, principalmente porque não há correspondência entre a aquisição de saberes e a capacidade de mobilizá-los ou aplicá-los no cotidiano (Souza, 2005). Portanto, é importante medir quanto cada instituição de ensino contribui para a formação de competências, assim como, para o desenvolvimento de habilidades.

Na educação, essa medida, é entendida como a contribuição da escola ao estudante, ou seja, a diferença entre o conhecimento e competências do estudante ao iniciar a graduação e o conhecimento e competências ao finalizar a graduação. Segundo Limana e Brito (2005) há

poucos textos de pesquisa sobre esse assunto, sendo que a maioria descreve situações que envolvem o desempenho de professores.

No ENADE é utilizado o conceito de valor agregado, para identificar qual foi o ganho do estudante no processo de aprendizagem. É uma medida do progresso médio do estudante durante o tempo em que ele está exposto a um determinado ambiente, comparativamente com o que ele obterá em outros ambientes educacionais, não no sentido de classificação do melhor ao pior ambiente, mas com o objetivo de identificar quais fatores explicam a variabilidade deste progresso (Ferrão, 2003).

As diretrizes Curriculares para os cursos de graduação no país, elegeram o desenvolvimento de competências vinculadas ao domínio de conhecimentos, saberes e práticas próprias da área profissional como meta orientadora na formação acadêmica (Araújo, 2004). A avaliação do desenvolvimento de conhecimentos e competências realizadas pelo ENADE deve investigar no curso do processo pedagógico, as habilidades, os conhecimentos e os saberes fundamentais que compõem o perfil profissional (MEC, 2004a).

Conforme Souza (2005) a avaliação das habilidades e competências tem-se tornado uma necessidade em um contexto comprometido com a qualificação profissional, principalmente por não haver correspondência entre aquisição de saberes e a capacidade de mobilizá-los ou aplicá-los no cotidiano de trabalho. Assim, a avaliação das habilidades e competências constitui um processo que tem como objetivo verificar a capacidade do estudante no enfrentamento de situações concretas, focalizando não só a tarefa, mas a mobilização e articulação dos recursos disponíveis. Daí a importância de verificar se o ENADE enquanto instrumento de avaliação consegue aferir a forma como os estudantes estão conseguindo estabelecer a relação entre teoria, prática, realidade e princípios éticos, na busca das competências.

Na prática, as medidas de avaliação das habilidades são medidas de desempenho e vice-versa. O desempenho do indivíduo é função da capacidade e também do incentivo, em outras palavras, é o potencial existente no indivíduo somado à motivação que lhe é proporcionado pelo meio no qual está inserido. Desempenho escolar é o termo utilizado para designar métodos que constatarem o conteúdo aprendido nas disciplinas escolares e também a maneira mais adequada de lidar com estes conhecimentos (Dorsch, Hacker & Stapf, 2001). O desempenho acadêmico envolve atividades escolares ou acadêmicas, mas ambos envolvem a aplicação das habilidades intelectuais do indivíduo na conquista do sucesso. (Munhoz, 2004).

O termo desempenho acadêmico, em português, é em geral associado a rendimento acadêmico. A descrição do termo desempenho envolve a dimensão da ação e, o rendimento é o resultado da sua avaliação, expresso na forma de notas ou conceitos obtidos pelo sujeito em determinada atividade. Considera-se o termo desempenho acadêmico no sentido de atuação observada de um indivíduo ou grupo na execução de tarefas acadêmicas avaliadas em termos de eficiência e rendimento, que refletem ou indicam o seu nível de habilidade, cujos resultados devem ser analisados para orientação futura tanto do indivíduo ou do grupo, como dos responsáveis pelas atividades acadêmicas oferecidas (Munhoz, 2004).

No contexto educacional, a avaliação do desempenho acadêmico afere se os objetivos de determinado curso foram atingidos, sendo função da avaliação aprimorar a aprendizagem do estudante (Souza, 1997). É também uma avaliação de qualificação, uma declaração sobre os resultados de trabalho concretamente alcançados podendo ser utilizada para esclarecer capacidades, interesses, motivações, conhecimentos e facilidades através de listas de qualificação e questionários (Dorsch, Hacker & Stapf, 2001).

O desempenho acadêmico envolve a relação entre as habilidades cognitivas e o domínio de conteúdos específicos. Para se obter um bom desempenho em qualquer área é preciso que se

realize um processo mental de informação relativamente específico (Flanagan, Ortiz, Alfonso & Mascolo, 2002). O desempenho acadêmico envolve a ação, sendo apresentado na forma do rendimento acadêmico por meio das notas obtidas em atividades específicas, ou seja, é o resultado da realização de atividades acadêmicas que envolvem as habilidades cognitivas e o domínio acadêmico específico mostrado pelo rendimento médio do estudante (Munhoz, 2004).

Autores como Kulieke, Bakker, Collins, Fennimore, Fine, Herman, Jones, Raack, e Tinzmann (1990), apontaram que a avaliação do desempenho acadêmico dos estudantes está baseada em perspectivas cognitivas filosóficas e multiculturais, sugerindo que a aprendizagem ocorre quando o estudante tem uma base de conhecimento que pode ser usada com fluência para resolver problemas e tomar decisões. Nesse contexto, uma prova deveria aferir o desempenho do estudante, e as análises de seus resultados mostrarem se o mesmo desenvolveu ou não a sua aprendizagem, como ocorre no ENADE. Segundo Limana e Brito (2005), o ENADE além de cumprir esse objetivo também avalia se o estudante é ou não capaz de trabalhar os conteúdos aprendidos durante a graduação.

As provas utilizadas nesses processos de avaliação do estudante devem possuir certas características que lhes confirmam credibilidade e precisão dos resultados, como fidedignidade e validade. Além dessas características, é necessário estudar a modalidade, a saturação, a dificuldade, a discriminação e o viés de resposta dos itens que compõem uma prova. Esses estudos oferecem uma base científica às provas e aos resultados conseguidos, pondo em evidências possíveis falhas metodológicas na construção desses instrumentos e garantindo instrumentos fidedignos para serem utilizados num processo de ensino.

As metodologias utilizadas para esses estudos baseiam-se principalmente em dois modelos matemáticos, a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI).

TEORIAS PSICOMÉTRICAS

TEORIA CLÁSSICA DOS TESTES (TCT)

A TCT surgiu com os primeiros trabalhos de Charles Spearman por volta de 1904, por meio da caracterização e avaliação da inteligência, tendo como principal objetivo encontrar um modelo estatístico que fundamentasse as pontuações nos testes e permitisse uma avaliação dos erros associados a todo processo de mensuração (Embretson & Reise, 2000). A Teoria Clássica dos Testes psicológicos leva à implicação de que os escores totais nos testes são compostos por um valor que representa a dimensão psicológica que se pretende medir e um valor que representa o erro cometido no processo de medida, fazendo com que seu objetivo seja encontrar um modelo matemático que possibilite estimar precisamente os erros cometidos neste processo, levando a uma interpretação mais adequada dos escores obtidos nos testes (Cronbach, 1996; Pasquali, 1997; Anastasi & Urbina, 2000).

A tarefa da TCT é elaborar estratégias estatísticas para controlar ou avaliar a magnitude do erro, que pode ser devido a uma grande variedade de fatores estranhos como defeitos do próprio teste e vieses do sujeito e como também fatores históricos e ambientais aleatórios (Pasquali, 2003). Autores como Muñiz (1994) e Pasquali (1997, 2003), apresentam três postulados para esse modelo: o valor verdadeiro é a expectativa do valor empírico; a correlação entre o valor verdadeiro e o erro é zero; e não há correlação entre os erros cometidos num teste j e em um teste i paralelo a ele.

No modelo da Teoria Clássica, as pontuações empíricas em um teste são amostras de comportamento que reúnem certos requisitos de medida que estão regidos por pressupostos que permitem fazer inferências probabilísticas bastante acertadas sobre os comportamentos. As estatísticas baseadas em amostras suficientemente grandes constituem estimativas adequadas para a descrição de características psicológicas dessa população (Cronbach, 1996; Pasquali, 2003).

A Teoria Clássica traz conceitos específicos que devem ser observados na construção dos instrumentos de medida. São os conceitos de parâmetros do teste (fidedignidade e validade) e parâmetros dos itens (modalidade, saturação, dificuldade, discriminação e viés de resposta) que estabelecerão a qualidade e a utilidade dos instrumentos de medida.

A fidedignidade ou precisão dos testes refere-se à estabilidade ou ao grau de consistência dos resultados, quanto mais próxima for a correlação entre os escores de um teste medido em situações diferentes, maior será sua precisão (Cronbach, 1996; Anastasi & Urbina 2000).

O conceito de precisão como apresentado por Cronbach (1996) é interpretado por uma relação estabelecida com o erro de medida, uma vez que o resultado de um teste ou prova pode não ser o mesmo em duas situações diferentes, então se diz que a diferença pode ser em virtude do erro. Em decorrência disso o escore verdadeiro, aquele que representa a melhor descrição da característica avaliada, se encontra em algum lugar em torno da média dos escores observados. O conceito de precisão está relacionado com a possibilidade de se estimar os erros aleatórios do processo de medir e não com outros tipos de erros que se podem controlar através de estratégias experimentais.

O coeficiente de precisão de um teste é expresso sob a forma de coeficientes de correlação, ou seja, é estabelecida uma correlação entre os escores de um teste obtidos em duas estimações diferentes, quanto mais próxima essa correlação estiver da unidade, mais precisos são os escores obtidos por aquela medida, e quanto mais distantes essa correlação estiver da unidade

e mais próxima de zero menor precisão terá a medida (Cronbach, 1996; Anastasi & Urbina, 2000). Segundo Cronbach (1996), os coeficientes de correlação podem ser calculados de várias maneiras, porém o mais comum é o coeficiente correlação produto-momento de Pearson. Anastasi e Urbina (2000) apresentam alguns métodos para se estimar os coeficientes de precisão por meio de coeficientes de correlação, descrevendo suas principais fontes de variância de erro.

Outro conceito de parâmetro do teste é a validade de uma prova ou teste e refere-se ao grau com que um teste mede aquilo que se propõe, permitindo inferências sobre os resultados encontrados nos testes e suas particularidades. É encontrado na literatura (Cronbach, 1996; Anastasi & Urbina 2000) três descrições acerca do tipo de validade: validade de conteúdo; validade de critério; e validade de construto que podiam ser escolhidos conforme as particularidades do instrumento. No entanto, não se deve identificar um tipo de validade como uma característica específica atribuída a cada tipo de instrumento, o desejável seria que um mesmo instrumento fosse submetido aos vários processos de verificação de validade.

Segundo a *American Educational Research Association-AERA*, *American Psychological Association-APA*, *National Council on Measurement in Education-NCME* (1999) atualmente procura-se nos testes por evidências de validade e não por tipo de validade, já que validade é um conceito único. Não se afirma que um teste ou prova é válido, mas sim que ele possui evidências de validade para a interpretação de um tipo de resultado, como descritor do construto psicológico que o teste pretende avaliar.

A proposta de validade, enquanto conceito ou propriedade psicométrica que tem como função confrontar as interpretações que são feitas a partir dos resultados dos testes psicológicos, leva a uma definição de validade como sendo o grau com que evidências teóricas ou empíricas sustentam as interpretações que são feitas dos escores de testes. Dessa forma, pode-se dizer que um teste psicológico pode ser válido em determinadas circunstâncias que descrevem um tipo

específico de uso, para as quais os dados teóricos ou empíricos encontraram sustentação para suas interpretações (AERA, APA & NCME, 1999).

Os procedimentos de pesquisa utilizados para a investigação da validade dos testes são divididos em cinco categorias: *evidências baseadas no conteúdo*, que buscam a representatividade dos itens do teste, investigando se eles se constituem em amostras abrangentes do domínio que se pretende avaliar com o teste; *evidências baseadas nas relações com outras variáveis*, que buscam os padrões de correlação entre os escores do teste e outras variáveis que meçam o mesmo construto ou construtos relacionados e com variáveis que meçam construtos diferentes; *evidências baseadas na estrutura interna*, um procedimento que se constitui em fonte de validade para as interpretações de um teste; *evidências baseadas no processo de resposta*, que buscam os processos mentais envolvidos na realização das tarefas propostas pelo teste e atribui significado psicológico para a realização correta do item a partir das relações entre seus componentes cognitivos; *evidências baseadas nas conseqüências da testagem*, que examinam as conseqüências sociais intencionais e não intencionais do uso do teste para verificar se sua utilização está tendo os efeitos desejados de acordo com os objetivos para o qual foi criado (AERA, APA & NCME, 1999).

Conforme Cronbach (1996), para se atribuir aos testes e provas seus parâmetros fundamentais, precisão e validade, é necessário verificar os parâmetros dos itens que o compõe, a saber, modalidade, saturação, dificuldade, discriminação e viés de resposta. A análise de itens para a estimação desses parâmetros torna-se parte fundamental do processo de construção de testes e provas, pois os decompõem nos seus aspectos mais elementares permitindo um julgamento mais acertado sobre quais itens são melhores descritores do construto ao qual o instrumento pretende avaliar, e quais podem ser descartados.

A TCT baseia-se na obtenção de resultados por escores brutos ou padronizados e suas análises e interpretações estão sempre associadas à prova como um todo, possuindo várias limitações como a dependência dos parâmetros dos itens (dificuldade e discriminação) da amostra de sujeitos utilizados, a existência de variação nas medidas fazendo com que a avaliação das aptidões seja dependente do teste utilizado e o fato de proporcionar uma análise orientada para o teste total e não para o item individual (Andrade, Tavares & Vale, 2000; Pasquali, 2003). As limitações apresentadas pelo modelo clássico suscitaram o desenvolvimento de novas abordagens provenientes da psicologia matemática, atualmente conhecida como Teoria de Resposta ao Item (TRI), cujas principais características são apresentadas a seguir.

TEORIA DE RESPOSTA AO ITEM (TRI)

A Teoria de Resposta ao Item teve suas origens nos trabalhos de Richardson (1936), Lawley (1943), Tucker (1946). A TRI começou a ser formalizada com os trabalhos de Lord em 1952 e 53 nos Estados Unidos e teve uma rápida expansão a partir de 1960 com o trabalho de Rasch na Dinamarca, sendo utilizada em testes de desempenho e de aptidão. Sua principal contribuição do ponto de vista teórico é a invariância dos parâmetros de medida, além de apresentar inovações técnicas como as funções de informação dos itens e do teste, medida dos Erros Padrões de Medida (EPM) mais refinados, apresentando suas variações ao longo da escala, além da possibilidade de novas formas de atribuição de significado psicológico para interpretação de escalas, baseada nas respostas aos itens (Andrade, Tavares & Vale, 2000; Muñiz, 1990).

Conforme Andrade, Tavares e Vale, (2000) a TRI vem sendo progressivamente introduzida na avaliação psicológica e educacional por permitir a comparação entre populações

submetidas a provas com itens comuns e a comparação entre os indivíduos da mesma população submetidos a provas diferentes. A TRI tem como elementos centrais os itens e não a prova como um todo. Diferentemente da Teoria Clássica, a TRI trabalha com traços latentes e adota dois axiomas fundamentais descritos por Pasquali (2003, p.82 - 83):

- 1) o desempenho do sujeito numa tarefa (item do teste) se explica em função de um conjunto de fatores ou traços latentes (aptidões, habilidades, etc.). O desempenho é o efeito e os traços latentes são a causa;
- 2) a relação entre o desempenho na tarefa e o conjunto dos traços latentes pode ser descrita por uma equação monotônica crescente chamada de CCI (Função Característica do Item ou Curva Característica do Item), onde se observa que sujeitos com aptidão maior terão maior probabilidade de responder corretamente ao item e vice-versa.

A TRI considera a interação entre sujeito e item reproduzindo a probabilidade do indivíduo acertar a questão considerando o processo de solução, a dificuldade do item e o fato de que um item não depende do outro. A pontuação dada em função das respostas do sujeito a cada item, possibilita afirmar se as pessoas são mais ou menos hábeis e se os itens são mais ou menos difíceis, pois itens e pessoas são colocados na mesma escala de desempenho e na mesma escala de uma característica latente subjacente (Fletcher, 1994).

Dentre as suposições feitas pela TRI, duas delas são essenciais para o seu entendimento, e delas depende a possibilidade de aplicação dos modelos da TRI aos dados: a *unidimensionalidade* e a *independência local* como descritas por vários autores (Almeida & Primi, 2001; Andrade, 2001; Andrade, Tavares & Vale, 2000; Andriola, 1998; e Pasquali, 1996, 2000, 2003). O critério de *unidimensionalidade* diz respeito a homogeneidade do conjunto de itens que devem medir um

único traço latente, ou seja, o grupo de itens deve medir uma mesma variável. Ao assumir que existe uma relação entre as respostas dos sujeitos e a dimensão do traço latente a ser avaliada, uma exigência para utilização do modelo é que a probabilidade de acerto dependa unicamente do valor do traço latente. Para satisfazer essa condição deve-se verificar se os itens usados para avaliar o traço latente medem apenas a dimensão descrita por ele e também se existe independência nas respostas entre os itens.

Na teoria, a unidimensionalidade perfeita pode ser encontrada em circunstâncias nas qual a resposta correta aos itens reunidos para avaliar uma determinada característica psicológica é determinada apenas por essa característica. Se a resposta correta aos itens depende ou sofre influência de outras dimensões psicológicas então não existe unidimensionalidade. Mas, como o desempenho humano é sempre multideterminado ou multimotivado, para satisfazer o critério de unidimensionalidade basta admitir que haja uma aptidão, fator ou habilidade dominante e responsável pelo conjunto de itens (Almeida & Primi, 2001; Andrade, 2001; Andrade, Tavares & Vale, 2000; Andriola, 1998; Pasquali, 1996, 2000 e 2003).

Dentre os métodos para se verificar a unidimensionalidade, a análise fatorial é o método mais difundido, e possibilita verificar qual o número mínimo de fatores que pode explicar a quantidade da variância total observada. Como não existe unidimensionalidade perfeita, ela é tratada nos modelos da TRI como uma questão de graus, sendo seu índice expresso pelo quociente entre a variância explicada pelo primeiro fator e a explicada pelo segundo fator (Muñiz, 1990).

O critério de *independência local* ou independência condicional supõe que a resposta de um sujeito a um item não sofra influências das respostas dadas a outros itens. Para níveis fixos de habilidade, quando mantidas constantes as aptidões que afetem o teste, as respostas dos sujeitos aos itens são estatisticamente independentes. Esse critério está diretamente ligado à

unidimensionalidade. Se todos os itens medem uma mesma dimensão, toda a variância do teste é explicada em função dos intervalos de representação de cada nível de dificuldade dos itens que medem essa dimensão presente no teste, além de acrescentar ao seu valor os erros aleatórios. E, portanto, os acertos ou erros dos sujeitos em cada item são independentes entre si, pois refletem a magnitude da dimensão avaliada pelo teste presente em cada um deles (Almeida & Primi, 2001; Andrade, 2001; Andrade, Tavares & Vale, 2000; Andriola, 1998; Pasquali, 1996, 2000 e 2003).

De acordo com Primi (2004), a TRI propõe um modelo matemático contendo a relação entre os elementos essenciais da situação na qual uma pessoa responde a um item; quanto maior a habilidade da pessoa para resolver o item, maior será a probabilidade de que ela o responda corretamente, mas se a habilidade for constante, quanto maior for a dificuldade do problema, menor será a probabilidade da pessoa o acertar. Para se encontrar o valor da habilidade é considerada a probabilidade de acerto ou de erro e a dificuldade do item. Uma vez conhecida a habilidade da pessoa, pode-se estabelecer as expectativas de acerto nos itens que avaliem aquela habilidade. Conhecendo o seu escore, pode-se prever quais itens uma pessoa terá mais chances de acertar ou errar, obtendo assim, o domínio que a pessoa possui do que está sendo avaliado.

Embora, seja ilimitado o número de modelos matemáticos que expressem a relação da probabilidade de acerto em um item e a habilidade medida pelo teste, na prática eles se distinguem pelo número de parâmetros que se utiliza para descrever o item (Pasquali, 2003). Os parâmetros utilizados para descrever os itens são: a *dificuldade* (parâmetro *b*), a *discriminação* (parâmetro *a*) e a *probabilidade de acerto por acaso* (parâmetro *c*).

A *dificuldade* do item (parâmetro *b*) é dada na mesma escala da habilidade e se refere à habilidade necessária para uma dada probabilidade de acertar o item, calculada a partir da probabilidade de acertar o item por acaso. Seu valor é dado pelo ponto, na reta, perpendicular ao eixo das abscissas, e representa o valor do traço latente quando a probabilidade de o sujeito

acertar ao item é de 50% (quando $c=0$). Ele também mantém uma relação com o seu correspondente do modelo clássico.

A *discriminação* (parâmetro a) corresponde à inclinação da Curva Característica do Item (CCI) em relação ao eixo das abscissas, o valor do ângulo formado por essa inclinação é proporcional ao valor de a , que será tanto maior quanto maior for o ângulo e indica quanto indivíduos de diferentes habilidades diferem quanto à probabilidade de acertar um item. O índice de discriminação, na TRI, tem uma relação com seu correspondente na Teoria Clássica.

A *probabilidade de acerto ao acaso* (parâmetro c) seu valor dado pela probabilidade de um indivíduo com uma certa habilidade responder corretamente a um item quando a habilidade do sujeito tende à menor representação possível dessa habilidade e, ainda assim, a probabilidade de acerto se mantém diferente de zero. Isso significa que o sujeito tem uma habilidade menor que a exigida para a execução adequada do item e, mesmo assim, sua probabilidade de acerto é diferente de zero.

Ao se especificar os parâmetros dos itens que se quer avaliar e optar por uma função matemática que expressará as curvas dos itens, se define conceitualmente a curva característica do item (CCI) como uma propriedade típica da TRI. A representação gráfica da CCI tem forma de “S” (Figura 1), com inclinação e deslocamento na escala de habilidade definidos pelos parâmetros dos itens. No eixo das abscissas está indicado o nível observado de habilidade (traço latente) do indivíduo, designado por θ (*theta*), cujo valor pode variar de $-\infty$ a $+\infty$, e no eixo das ordenadas a probabilidade de responder corretamente ao item, designado por $P_i(\theta)$, podendo variando de 0 a 1.

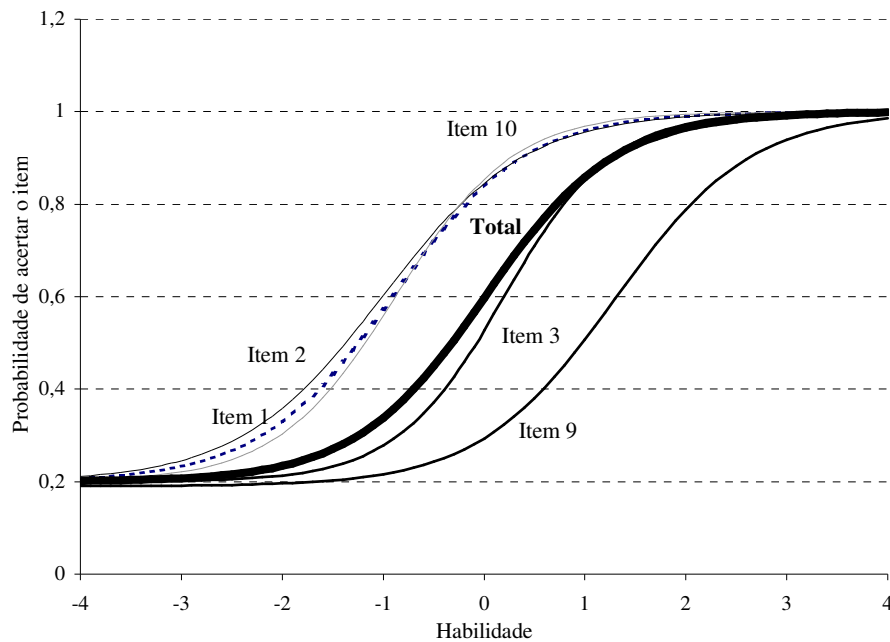


Figura 1 - Curvas características de alguns itens de uma prova (Vendramini, Silva & Canale, 2004)

Assim, os modelos matemáticos apresentados nas três expressões a seguir, representam para cada item i a função de probabilidade de indivíduos com habilidade θ acertarem esse item (Vendramini, 2001; Vendramini, Silva & Canale, 2004).

Existem vários modelos do TRI dependendo do tipo de função matemática que se use e dos parâmetros dos itens que se queira investigar. Os dois tipos de função mais frequente são as funções logísticas e as de curva normal acumulada o que produziriam pelo menos seis modelos, sendo possível avaliar até três parâmetros para cada função. Serão abordados neste trabalho apenas os modelos de função logística, pois permitem melhor tratamento matemático e são também mais frequentes na literatura especializada (Muñiz, 1990).

O modelo de três parâmetros é o modelo teórico mais completo. Considera os três parâmetros como variáveis que influenciam a probabilidade do indivíduo acertar o item e é expresso pela função matemática a seguir:

$$P(X_{ij} = 1 | \theta_j) = P_i(\theta) = c_i + \frac{(1 - c_i)}{1 + e^{-Da_i(\theta - b_i)}}$$

X_{ij} uma variável dicotômica que assume os valores 1 ou 0, conforme o indivíduo j acerte ou não o item i , respectivamente;

θ_j o valor que representa a variável latente (aptidão ou habilidade) que permite explicar a resposta do j -ésimo indivíduo aos itens;

$P_i(\theta)$ a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i ;

c_i a probabilidade de acerto ao acaso;

b_i índice de dificuldade do item i , medido na mesma escala da habilidade θ . Corresponde à habilidade necessária para a probabilidade de acerto igual a $(1 + c_i) / 2$;

a_i o índice de discriminação do item i , com valor proporcional à inclinação da CCI no ponto b_i ;

D constante de valor 1,7 (quando se deseja que a função logística se aproxime da curva normal padronizada);

e base dos logaritmos neperianos, cujo valor é aproximadamente 2,718;

O modelo de dois parâmetros foi desenvolvido por Birnbaum em 1968 e serve para avaliar o índice de dificuldade b_i e o de discriminação a_i , esse modelo pode ser entendido como um modelo de três parâmetros com o valor $c_i = 0$. Uma relação importante entre os parâmetros θ

e a_i é que a discriminação varia em função da habilidade do sujeito, ou seja, os itens discriminam melhor quando a habilidade coincidir com o ponto de inflexão da curva (Muñiz, 1990).

$$P(X_{ij} = 1 | \theta_j) = P_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}}$$

O modelo de um parâmetro, criado por *Rasch* em 1960 e expresso como modelo de ogiva possui em sua expressão o índice de dificuldade b_i , que se relaciona com a probabilidade de acertar o item i por acaso. Esse modelo pode ser entendido como um modelo de três parâmetros com o valor $c_i = 0$ e mesmo valor a para todos os a_i 's. Trata-se do modelo mais difundido devido a sua parcimônia de medida e da simplicidade de sua lógica, esse modelo é representado por uma função logística de um parâmetro no qual leva em conta que as respostas de um sujeito a um conjunto de itens dependem apenas de sua habilidade e da dificuldade dos respectivos itens (Baker, 2001).

$$P(X_{ij} = 1 | \theta_j) = P_i(\theta) = \frac{1}{1 + e^{-Da(\theta - b_i)}}$$

Um questionamento levantado por Baker (2001) acerca dos principais modelos da TRI é qual deles o pesquisador deve usar para analisar seus dados e que deve ser respondido de acordo com a adequação dos dados ou com o “ajuste” do modelo aos dados. Esse ajuste do modelo escolhido aos dados deve ser comprovado pelos parâmetros estimados.

O primeiro passo do ajuste do modelo constitui-se da definição rigorosa da dimensão que se pretende avaliar, logo que se defina qual a dimensão a ser avaliada passa-se a elaboração de itens que representem adequadamente essa dimensão. O próximo passo é a aplicação dos itens a uma amostra representativa da população para qual o instrumento foi construído, e uma análise fatorial na qual se verifica a unidimensionalidade dos itens. Acredita-se que nessa fase já se torna

possível vislumbrar qual dos modelos da TRI melhor se ajusta aos dados, pois se a discriminação dos itens não for constante pode-se supor que os dados se ajustam melhor ao modelo de dois parâmetros, ou se houver probabilidade de acerto ao acaso os dados se ajustam melhor ao modelo de três parâmetros. A partir das suposições que são possíveis nessa fase, o árbitro do processo, aquele que dá a palavra final, deve ser o ajuste do modelo aos dados (Baker, 2001).

Existem vários procedimentos estatísticos usados para a comprovação do ajuste do modelo aos dados, como o χ^2 (qui-quadrado), a máxima verossimilhança e a análise de resíduos, mas nenhum é totalmente satisfatório o que acaba por se tornar o ponto fraco da TRI no seu momento atual de desenvolvimento (Muñiz, 1990; Baker, 2001).

A TEORIA DE RESPOSTA AO ITEM NA AVALIAÇÃO EDUCACIONAL

Alguns testes e programas de testagem foram desenvolvidos para o uso na admissão, colocação e no aconselhamento de estudantes universitários. O mais conhecido, e o que mais se destaca, é o Programa de Testes de Avaliação Escolar (*Scholastic Aptitude Test – SAT*), do *College Board*, que é composto por dois subtestes, o SAT-I, que é o teste de raciocínio; e, o SAT-II, que se refere ao teste de conhecimento específico (Anastasi & Urbina, 2000).

O SAT-I consiste em perguntas de múltipla escolha medindo habilidades verbais e matemáticas, sendo utilizado na avaliação da prontidão de um estudante para realizar um trabalho de nível universitário. Já o SAT-II, tem como objetivo, avaliar o conhecimento em áreas específicas, como Literatura, Química. Pode ser usado para colocação, assim como admissão.

Um outro programa de avaliação é o *American College Testing Program (ACT)*, que inclui quatro testes, sendo eles: Inglês, Matemática, Leitura e Raciocínio Científico. Desta forma, algumas amostras do trabalho universitário, em que, sobrepõe testes de aptidão e realização tradicionais, estão centradas nas habilidades intelectuais básicas necessárias para um desempenho satisfatório na universidade.

No Brasil, várias formas governamentais de avaliar o estudante em sua vida acadêmica foram criadas, uma delas foi a avaliação do *Exame Nacional do Ensino Médio (ENEM)*, realizado com estudantes do Ensino Médio, e cuja pontuação é utilizada no processo seletivo de ingresso a várias instituições de Ensino Superior. Uma pesquisa realizada por Brito e cols. (2000), buscou verificar as relações existentes entre a parte matemática do ENEM, o raciocínio e o desempenho escolar em Matemática. Participaram deste estudo, 338 estudantes do segundo ano do ensino médio, sendo 166 de uma escola da rede estadual de ensino e 172 de uma escola da

rede particular do Estado de São Paulo. Foram aplicados, a Bateria de raciocínio BPR-5 e uma prova de Matemática, com 12 questões referentes à disciplina em questão, do ENEM de 1999. Os resultados apontaram o ENEM como um instrumento viável para a avaliação do estudante egresso do ensino médio, contudo, apresenta vários pontos a serem melhorados, como uma revisão a ser feita com relação à transformação dos resultados e às definições de habilidades e competências assim como a implementação de estudos que procurem identificar os fatores que influenciem o desempenho no ENEM.

Um outro instrumento criado com a finalidade de avaliar o estudante em sua vida acadêmica foi o Exame Nacional de Cursos (ENC) como já explicitado anteriormente. Landeira-Fernandez e Primi (2002) realizaram uma pesquisa na qual foi investigada a possível fonte de ameaças à validade em relação ao Exame Nacional de Cursos (ENC) de Psicologia realizados no ano de 2000, sendo verificado também, a possibilidade desse instrumento de avaliação de conhecimento específico de Psicologia ser sensível a diferenças impostas pelo vestibular. Participaram deste estudo, 386 calouros de psicologia, distribuídos entre quatro cursos selecionados de acordo com o seu conceito no ENC 2000 (A, B, C e E). Os estudantes realizaram a parte da prova constituída por questões de múltipla escolha do ENC de Psicologia 2000. Os resultados revelaram que calouros provenientes de diferentes cursos apresentaram desempenhos altamente diferenciados e proporcionais aos conceitos dos seus cursos. No entanto, o aumento do desempenho dos formandos em relação aos calouros foi diferenciado de acordo com o conceito do curso. Cursos com conceitos A e B no ENC 2000 apresentaram um maior incremento em relação aos cursos com conceitos C e E. Uma análise de itens revelou diferentes graus de validade entre as 39 questões de múltipla escolha que compuseram o ENC 2000. Uma re-análise dos resultados, considerando-se apenas 20 itens que apresentaram boa validade, indicou uma ausência de diferença no desempenho entre os calouros provenientes dos diferentes cursos. Por

outro lado, os formandos apresentaram um aumento nos seus desempenhos em relação aos calouros. Esse aumento no desempenho observado entre os formandos foi diferenciado e proporcional ao conceito atribuído ao curso. Finalmente, uma análise dos conteúdos específicos de cada um dos 39 itens do ENC 2000 revelou uma deficiência na formação acadêmica dos formandos na área básica da Psicologia relacionada com processos psicológicos básicos, metodologia científica e medidas e técnicas de avaliação psicológica.

Uma outra pesquisa realizada com a finalidade de verificar o desempenho de estudantes ingressantes no Ensino Superior foi descrita por Primi, Vendramini, Santos e Filho (1999). Os autores investigaram as diferenças de desempenho de candidatos ao processo seletivo de ingresso numa universidade particular, em que procurou-se estimar a magnitude do nível de instrução educacional de ambos os pais, dependência administrativa (privada ou pública) da escola em que o candidato cursou o ensino fundamental e médio, período em que cursou (noturno, diurno/integral) e se exercia atividade remunerada enquanto estudava. Assim como, medida indireta da condição financeira, em 9545 candidatos que responderam às provas do processo seletivo de uma universidade da rede particular do Estado de São Paulo. Utilizaram-se os modelos gerais lineares construindo-se contrastes para cada variável sócio-econômica e efetuando-se uma regressão linear múltipla, procurando prever o desempenho a partir delas. As variáveis estudadas contribuíram significativamente para explicar 25,9% das diferenças de desempenho. Com esse estudo, concluiu-se que as diferenças de desempenho de maior magnitude ocorreram entre os candidatos que estudaram em escolas privadas no período diurno ou integral e os que estudaram em escolas públicas em períodos noturnos, em que favoreceram os primeiros.

Com relação ao estudo das habilidades em universitários podemos citar um estudo realizado por Santos, Primi, Vendramini, Taxa, Lukjanenko, Muller, Sampaio, Andraus Jr., Kuse e Bueno (2000) com o objetivo de avaliar as habilidades básicas do estudante ingressante em

cursos de graduação, caracterizando-os, em função de seu potencial e limites, no que se refere à competência para leitura/escrita, conhecimentos gerais e atualidades, raciocínio lógico e juízo moral. Participaram 720 estudantes dos cursos de Administração, Medicina, Odontologia e Psicologia. Do total, 364 estudantes responderam a todas as provas. Foi utilizado um teste de compreensão de leitura (Santos e cols., 1998), um teste de conhecimentos gerais (Santos, Primi, Vendramini, Taxa, Lukjanenko, Muller, Sampaio, Andraus Jr., Kuse & Bueno, 2000), um teste de julgamento de situações (TJS) traduzido por Bzuneck (1989), os sub-testes de raciocínio abstrato e de raciocínio espacial da Bateria de Provas de Raciocínio (BPR-5) de Almeida e Primi (1998), e o teste de raciocínio lógico-dedutivo (Primi, Muller e Vendramini, 1997). Os autores procederam, então, a aplicação da estatística descritiva cujos resultados apontaram para a prova de raciocínio lógico-dedutivo como a mais difícil. A análise fatorial das seis provas apontou a presença de três fatores: a) juízo moral pré-convencional e convencional; b) inteligência cristalizada e juízo pós-convencional; c) inteligência fluída. Segundo os autores, a análise das provas com maior carga nos fatores dois e três está em consonância com a proposta de Ackerman (1996), de inteligência baseada na capacidade de processamento e de inteligência baseada no conhecimento, que seria na proposta de Carroll (1993) associada respectivamente à inteligência fluída e cristalizada. A prova de raciocínio lógico-dedutivo apresentou cargas nos dois fatores apontando que utiliza tanto de inteligência fluída, no enfrentamento do problema, como de inteligência cristalizada, pois este é mediado pela linguagem. Os resultados também apontam diferenças de desempenho de acordo com os cursos. Para a análise destas diferenças, os autores utilizaram somente os estudantes que haviam respondido a todas as provas. Observou-se que, excetuando-se a prova de raciocínio abstrato, Medicina obteve o melhor desempenho nas provas de raciocínio, seguido de odontologia. As diferenças foram significativas, entre Medicina e odontologia, nas provas de compreensão de leitura e raciocínio lógico-dedutivo. Odontologia

apresentou o melhor desempenho em raciocínio abstrato diferindo significativamente dos estudantes da Medicina. Os desempenhos dos estudantes de administração e psicologia ficaram próximos. Os autores também procuraram verificar a relação entre o desempenho e o número de candidatos/vaga. Percebeu-se que os cursos mais concorridos apresentavam melhor desempenho nas provas, o que apóia a idéia da relação entre conhecimento e raciocínio. Além dessas análises, os autores também procederam a uma verificação das diferenças de desempenho entre gêneros. Os homens apresentaram desempenho superior nas provas de raciocínio lógico-dedutivo e espacial enquanto que as mulheres apresentaram desempenho superior na prova de compreensão de leitura. Os autores também verificaram o valor preditivo dos instrumentos, correlacionando-os às médias do primeiro semestre dos estudantes e obtiveram correlações positivas significativas entre o desempenho acadêmico e as provas de raciocínio.

Dentre as várias pesquisas nacionais que utilizaram a TRI na avaliação educacional encontra-se o trabalho de Andriola (1998), que realizou uma análise comparativa entre a Teoria de Resposta ao Item e a Teoria Clássica dos Testes (TCT) a partir de um estudo de casos considerando o desempenho dos respondentes e os parâmetros métricos dos itens. Os dados que resultaram este estudo são resultantes da Avaliação da Qualidade do Ensino das Escolas Públicas do Estado do Ceará realizada em 1996. A amostra foi constituída por oito estudantes da 8ª série com idades entre 14 e 21 anos, tendo como critério de escolha quatro respondentes com dois desvios padrões acima da média (grupo superior) e quatro respondentes com dois desvios padrões abaixo da média (grupo inferior). O instrumento utilizado foi um teste de Português, composto por 25 questões fechadas e com tempo de resolução de 90 minutos. O autor interpretou os resultados dos respondentes pelas duas metodologias propostas, a TCT e a TRI, chegando às seguintes conclusões: o modelo de três parâmetros foi o que melhor se ajustou à distribuição dos

resultados, o que permitiu uma análise qualitativa mais acurada do nível de desempenho, ampliando as análises do estudo de casos proposto inicialmente.

Em outro estudo Andrade, Tavares e Valle (2000) discutiram uma aplicação prática, interpretada segundo a TRI, para exemplificar algumas contribuições que esta teoria traz para a área de Avaliação da Aprendizagem. Os autores utilizaram neste estudo as provas do SARESP – Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo, aplicadas em estudantes de 3ª a 8ª Série do Ensino Fundamental e 1ª série do Ensino Médio, nos anos de 1996 a 1998. Foi observado neste estudo que as avaliações não permitiam comparações sobre o desempenho dos estudantes ao longo do tempo, por não apresentarem itens comuns, sendo necessário a criação de uma prova adicional que permitiu equalizar as informações em escalas comparáveis de conhecimento. Com a utilização da TRI os autores puderam analisar as diferenças de desempenho dos estudantes de uma série para outra, o que permitiu fazer apontar indicadores sobre o quanto as intervenções no ensino público afetou tal desempenho.

Santos, Primi, Taxa e Vendramini (2002), realizaram um estudo que procurou investigar as propriedades psicométricas de um instrumento de avaliação da compreensão em leitura utilizando a Técnica de Close. Participaram deste estudo 612 estudantes ingressantes de quatro cursos de uma universidade particular cuja tarefa era preencher a lacuna do texto com a palavra que julgasse mais apropriada. Observou-se um bom ajuste ao modelo de dois parâmetros da teoria de Resposta ao Item. A escala formada pela soma do número de lacunas corretamente completadas provou possuir consistência interna alta. Os itens formaram dois grupos em razão das dificuldades: um grupo com itens muito fáceis e outro com itens muito difíceis. Os autores concluíram que os itens se agruparam formando uma escala precisa, reafirmando a viabilidade da utilização da Técnica de Close como instrumento de boa qualidade para a avaliação da compreensão em leitura.

Outro estudo utilizando a TRI na avaliação educacional foi realizado por Vendramini e Dias (2005) que teve como objetivo aplicar a Teoria de Resposta ao Item na análise de uma prova de Estatística. Participaram deste estudo 413 universitários de duas instituições particulares do Ensino Superior, predominantemente do curso de Psicologia. O instrumento utilizado consiste numa prova de 15 questões de múltipla escolha apresentada na forma de gráficos ou tabelas estatísticas. As autoras concluíram que a prova é predominantemente unidimensional e os itens são mais bem ajustados ao modelo de três parâmetros. Os índices de discriminação, dificuldade e correlação bisserial apresentaram valores aceitáveis.

Um estudo realizado com o ENADE foi feito por Primi (2005) com o objetivo de buscar evidências de validade das principais interpretações para a prova, tentando verificar as associações entre o desempenho dos estudantes e as variáveis: momento de formação e nível de desempenho do curso. Foram considerados nessa análise 106.936 estudantes, que responderam a pelo menos uma das partes da prova. Foi utilizado para análise dos dados duas Anovas Fatoriais, uma considerando o escore global no componente de formação geral e outra considerando o escore global no componente específico. O autor concluiu que é possível verificar nos dados das provas específicas que as provas foram sensíveis em detectar as diferenças de desempenho entre formandos e ingressantes e também detectar variações na magnitude dessas diferenças que poderiam ser atribuíveis aos cursos.

MÉTODO

FONTE DE DADOS

Foi utilizado um banco de dados que contém dados acadêmicos e os resultados do ENADE de estudantes ingressantes e concluintes, do ano de 2004 que participaram do exame. Neste estudo foi avaliado o curso de Medicina ($n = 8.124$).

A amostra desse banco de dados é composta por estudantes selecionados pelo INEP a partir da listagem enviada pelas IES contendo a relação dos ingressantes e dos concluintes. A IES apenas envia a lista dos estudantes que cumprem o critério para realizar o exame, sendo que todo procedimento de amostragem é feito pelo INEP, de tal forma que é impossível para qualquer IES escolher quem vai ou não participar do exame e após o processo de amostragem, cada IES recebe a relação dos estudantes selecionados.

No ano de 2004 foram inscritos 250.931 estudantes de todos os cursos selecionados para serem avaliados e, destes, foram selecionados 155.654. Dos estudantes selecionados, 140.340 estudantes realizaram a prova (entre ingressantes e concluintes). No total, compareceram para realizar o exame 143.170 estudantes e a diferença de 2.830 sujeitos inclui os inscritos posteriormente à seleção da amostra (voluntários e aqueles que conseguiram mandato de segurança).

A filtragem dos sujeitos do presente estudo foi feita a partir dos 140.340 estudantes da amostra que realizaram a prova e as análises apresentadas a seguir estão de acordo com o critério previamente estabelecido para a inclusão de estudantes nesta base de dados. O critério de filtragem foram estudantes que realizaram, pelo menos, uma das partes da prova. Após a

filtragem foi obtido o total de $N = 135.631$ estudantes, do qual foram selecionados somente os estudantes de Medicina ($n = 8.124$) e todas as análises são baseadas neste total.

No ano de 2004 foram considerados ingressantes os estudantes que tivessem cumprido, até a data inicial do período de inscrição, 7 a 22% da carga didática mínima do currículo do curso da IES e concluintes, os estudantes que tivessem cumprido, até a data inicial do período de inscrição, pelo menos 80% da carga mínima da grade curricular do curso da IES. Esse não é um critério fixo e pode ser alterado a cada ano visando a inclusão de um número maior ou menor de estudantes, de forma a atender as especificidades de duração de cada curso. Quando um curso tem um número baixo de estudantes, todos são incluídos. Os estudantes que não são sorteados ou ficam de fora da listagem por qualquer razão podem participar do ENADE, mas os resultados não são computados com o grupo.

Distribuição dos Cursos e Estudantes

Participaram desta pesquisa 8.124 estudantes do curso de Medicina, que responderam até 15% das questões analisadas, provenientes de 117 Instituições de Ensino Superior do país. Destas Instituições 52,9% são particulares e 47,1% são públicas divididas entre federais (33,2%), estaduais (11,4%) e municipais (2,5%). A maioria destas instituições concentra-se na região sudeste (50,7%).

Tabela 1. Distribuição dos estudantes de Medicina por região

Regiões do País	Número de estudantes		
	Nº	%	% Acumulada
Note	402	4,9	4,9
Nordeste	1251	15,4	20,3
Sudeste	4120	50,7	71,1
Sul	1829	22,5	93,6
Centro-Oeste	522	6,4	100,0
Total	8124	100,0	

Dos estudantes que compõem a amostra, 50,2% são do sexo masculino e 49,8% do sexo feminino. A idade varia de 15 anos a 63 anos ($M=23,16$; $DP=3,31$). Destes estudantes 56,8% são estudantes ingressantes e 43,2% são concluintes.

Tabela 2. Distribuição dos estudantes de acordo com a situação no curso

Situação	Número de estudantes		
	Nº	%	% Válida
Concluintes	3.509	43,2	43,2
Ingressantes	4.615	56,8	56,8
Total	8.124	100,0	100,0

MATERIAL

A prova do ENADE é composta de 40 (quarenta) questões, sendo 10 (dez) questões de avaliação da formação geral, oito de múltipla escolha e duas discursivas, comuns aos cursos de todas as áreas e 30 (trinta) questões específicas para cada área, compostas por vinte e oito

questões discursivas e duas de múltipla escolha. As questões são elaboradas em forma de estudos de caso, situações e alternativas problematizadoras, dentre outras. Com relação ao grau de dificuldade, as questões devem distribuir-se entre questões com baixo, médio e elevado graus de dificuldade.

De acordo com a legislação referente ao ENADE, a primeira parte da prova é comum a todos os cursos participantes do ENADE com o objetivo de aferir a aquisição de competências, o desenvolvimento de habilidades e conhecimentos considerados essenciais na formação geral de qualquer estudante de Educação Superior, permitindo assim a comparabilidade entre os cursos de diferentes Instituições. E a segunda parte, específica para cada área, trata de conteúdos específicos e busca verificar o domínio da área, aferindo o potencial de aprendizagem dos ingressantes e as competências que os concluintes desenvolveram ao longo da trajetória na IES. Estas questões devem contemplar as especificidades de cada área e de suas modalidades, tanto no domínio dos conhecimentos da área quanto nas habilidades esperadas para o perfil profissional, investigando o que o estudante é capaz de realizar a partir dos conhecimentos e atividades disponibilizados pelo curso, ela é elaborada com base nas Diretrizes Curriculares, aprovadas pelo Conselho Nacional de Educação e também no perfil profissional de cada curso, contemplando os saberes fundamentais exigidos em cada área profissional.

Neste estudo foram utilizados os componentes de Formação Geral e Formação Específica do ENADE aplicadas nos cursos de Medicina de todo país. Entraram nas análises descritas a seguir somente as questões objetivas de cada parte da prova, totalizando 28 itens.

PROCEDIMENTO

Após a aprovação do comitê de ética da Universidade de São Francisco para a realização da pesquisa, foi formalmente solicitada a autorização para a utilização do banco de dados selecionado para este estudo, comprometendo-se a manter em sigilo os dados individuais e o nome da instituição envolvida.

PLANO DE ANÁLISE DE DADOS

Tanto para a análise da TCT quanto a TRI foram consideradas as respostas dos estudantes nas cinco alternativas de cada item, indicadas as respostas corretas, para possibilitar que os itens assumam escores do tipo certo/errado (itens dicotômicos).

Os dados foram analisados por programas computacionais específicos: o programa TESTFACT para a análise da dimensionalidade da prova (Wilson, Wood & Gibbons, 1998); o programa WINSTEPS para análise do ajuste do modelo de um parâmetro de *Rasch*; e o XCALIBRE para o ajuste dos modelos de dois e três parâmetros pela estimação marginal de máxima verossimilhança (*Assessment System Corporation, 1995*).

Desenvolveram-se análises da prova do ENADE (curso de Medicina) para a estimação dos parâmetros usando os modelos de Rasch, de 2 e de 3 parâmetros. Os critérios para avaliação do ajuste do modelo aos dados são: para o modelo de Rasch os valores dos resíduos chamados de *Outfit* e *Infit* devem manter-se entre 0,70 e 1,30; para os demais modelos o valor dos resíduos deve ser inferiores a 2,0, sendo que valores menores significam melhor ajuste que valores maiores. Os parâmetros dos itens devem manter valores para a discriminação do item (parâmetro *a*) maiores que 0,30; para a dificuldade do item (parâmetro *b*) entre -2,95 e 2,95 e para a

probabilidade de acerto ao acaso (parâmetro c) valores menores que 0,40 (Linacre & Wright, 1994; Microcat, 1989).

Verificaram-se os coeficientes de fidelidade das provas estimadas utilizando a fórmula de Kuder-Richardson e as curvas de informação do teste estimado pelo modelo de 3-parâmetros.

Será apresentada inicialmente a análise pela TCT para, a seguir, serem apresentados e aplicados os três modelos matemáticos da TRI (um, dois e três parâmetros), selecionando-se aquele que se ajustar a um maior número de itens e que melhor represente as respostas dos sujeitos à prova.

RESULTADOS

A análise clássica dos itens de uma prova baseia-se em parâmetros descritivos dos itens, que auxiliam na interpretação da distribuição das respostas em cada alternativa do item. As propriedades psicométricas dos itens da prova (Tabelas 3 e 4) correspondem aos seguintes parâmetros: (1) *média do total de acertos* dos participantes que acertaram um determinado item; (2) *índice de facilidade* - proporção de participantes que responderam ao item corretamente; (3) *índice de dificuldade*, que corresponde ao índice de facilidade calculado em uma escala delta (Δ) com média 13 e desvio padrão 4; (4) *correlação ponto bisserial*, índice escolhido para ser utilizado pelo ENADE que mede o poder de discriminação das questões; (5) *média e desvio-padrão do número total de acertos*, considerando-se as respostas deixadas em branco como erradas em cada item.

O índice de facilidade é obtido por meio do cálculo da proporção de acertos na questão. Dessa forma um índice de facilidade igual a 0,90 significa que 90% dos estudantes responderam corretamente uma determinada questão. O intervalo desse índice varia de 0 (nenhum estudante acerta) a 1 (todos os estudantes acertam). Quanto mais o índice estiver próximo de 1 maior a facilidade dos estudantes para resolver a questão (Cronbach, 1994; MEC, 2004e).

Os resultados mostram que o componente de Formação Geral (Tabela 3) apresenta itens muito fáceis (com valores acima de 80% de acerto – item 1), itens fáceis (com valores entre 60% e 80% - itens 3, 5 e 7) e itens de nível médio (com valores entre 40% e 60% - itens 2, 4, 6 e 8). Pode ser observado que nenhuma questão foi considerada muito difícil (com valores abaixo de

20%) e nem difícil (com valores entre 20% e 40%) mostrando que não há uma distribuição equilibrada das questões em relação à sua facilidade.

O componente de Formação Específica (Tabela 4) apresenta itens muito fáceis (itens 9 e 10), fáceis (itens 15, 16, 24 e 28), itens médios (itens 19, 20, 22 e 25), difíceis (itens 11, 12, 13, 14, 17, 18, 23, 26 e 27) e muito difícil (item 21). O que significa uma distribuição razoavelmente equilibrada, com nove das vinte questões classificadas como difíceis.

Tabela 3. Parâmetros descritivos dos itens do componente de Formação Geral

Item	Média de acerto dos que acertaram o item	Índice de facilidade	Dificuldade (DIFF)	Correlação Ponto-bisserial
1	5,14	0,911	7,61	0,350
2	5,72	0,558	12,42	0,534
3	5,47	0,699	10,92	0,491
4	5,57	0,542	12,58	0,417
5	5,30	0,798	9,66	0,426
6	5,79	0,407	13,94	0,432
7	5,41	0,644	11,52	0,379
8	5,74	0,416	13,84	0,414
Total	4,9			
DP	1,6			

O índice de discriminação (estimado pela correlação ponto bisserial) mede o grau com que uma questão é capaz de separar os estudantes com melhor desempenho na prova daqueles com pior desempenho. Questões classificadas como muito fáceis (que todos ou quase todos acertam) ou muito difíceis (que todos ou quase todos erram), em geral, tem baixo índice de discriminação (MEC, 2004e).

Ainda na Tabela 3, observa-se que seis questões são classificadas como tendo índice de discriminação muito bom (correlação ponto bisserial maior que 0,40 – itens 2, 3, 4, 5, 6, 8) e duas questões classificadas como tendo índice de discriminação bom (correlação ponto bisserial entre 0,30 e 0,40 – itens 1 e 7), indicando que as questões foram capazes de separar os estudantes com melhor desempenho daqueles com pior desempenho. O item 1 embora seja o item mais fácil da prova e com menor correlação ponto bisserial, ainda pode ser considerado como um item de boa discriminação, capaz de separar os estudantes por seu desempenho.

As questões do componente de Formação Específica (Tabela 4) apresentam índices de discriminação muito bons (correlação ponto bisserial maior que 0,40 – itens 13, 16, 20, 22, 23, 25, 27 e 28), bom (correlação ponto bisserial entre 0,30 e 0,40 – itens 10, 11, 15, 17, 18, e 26), médios (correlação ponto bisserial entre 0,20 e 0,30 – itens 9, 19 e 24) e fracos (correlação ponto bisserial menor que 0,20 – itens 12, 14 e 21). As questões em sua grande maioria, 14 das 20 questões, foram capazes de separar os estudantes com melhor desempenho daqueles com desempenho mais baixos, sendo preciso rever os itens com fraca discriminação. Os itens 11, 13, 17 e 26 apesar de serem considerados difíceis, são capazes de separar os estudantes pelo desempenho por apresentarem um bom índice de discriminação.

Esses resultados não diferem em sua maioria dos apresentados no Relatório Síntese do curso de Medicina publicado pelo INEP, considerando que no presente estudo foram eliminados todos os estudantes que entregaram a prova em branco, ou que não responderam a pelo menos 15% das questões (MEC, 2004e).

Tabela 4. Parâmetros descritivos dos itens do componente de Formação Específica

Item	Média de acerto dos que acertaram o item	Índice de facilidade	Dificuldade (DIFF)	Correlação Ponto-bisserial
9	9,69	0,903	7,81	0,276
10	9,72	0,902	7,84	0,303
11	11,24	0,292	15,19	0,356
12	10,01	0,329	14,77	0,131
13	12,06	0,266	15,51	0,479
14	10,05	0,273	15,42	0,118
15	10,10	0,731	10,53	0,346
16	10,52	0,634	11,63	0,439
17	11,73	0,205	16,29	0,352
18	10,92	0,394	14,08	0,369
19	10,24	0,552	12,47	0,272
20	11,77	0,431	13,70	0,610
21	10,57	0,192	16,48	0,168
22	11,32	0,427	13,73	0,487
23	11,92	0,348	14,57	0,543
24	9,99	0,722	10,65	0,263
25	11,58	0,484	13,16	0,625
26	11,18	0,245	15,76	0,300
27	11,49	0,354	14,50	0,455
28	10,19	0,778	9,94	0,415
Total	9,4			
DP	3,4			

Em média os estudantes acertaram mais que 50% das questões do componente de Formação Geral ($M = 4,9$; $DP = 1,6$). Das oito questões deste componente, a média de acertos para os participantes que acertaram um determinado item variou de 5,14 a 5,79. Com relação às 20 questões do componente de Formação Específica, a média de acertos foi inferior a 50% das questões ($M = 9,4$; $DP = 3,4$), variando de 9,69 a 12,06, indicando uma variabilidade de

desempenho em torno da média (*coeficiente de variação* = 35,6%) ligeiramente superior ao componente de Formação Geral (*coeficiente de variação* = 32,7%).

Os itens do componente de Formação Geral foram considerados de dificuldade mediana e consistência interna razoável pela técnica de *Kuder-Richardson*, que permite verificar a fidedignidade do teste ($KR-20 = 0,391$). Os itens do componente de Formação Específica foram considerados de dificuldade mediana e consistência interna boa pela técnica de *Kuder-Richardson* ($KR-20 = 0,681$). Observa-se que no componente de Formação Específica a precisão é bem superior ao componente de Formação Geral, isso pode ser explicado pelo conteúdo que compõe cada um dos componentes, sendo, o componente de Formação Geral bem mais abrangente em termos de conhecimentos avaliados.

Um das suposições da TRI é que a prova seja unidimensional, ou pelo menos que se possa assumir um fator predominante para se utilizarem modelos unidimensionais. O programa TESTFACT efetua a análise considerando questões do tipo certo/errado (dados dicotômicos) a partir das respostas dos participantes (em vez da matriz de correlação). Esta análise é denominada Análise Fatorial com Informação Completa (*Full Information Factor Analysis*) e inclui progressivamente fatores que indicam a contribuição do fator incluído para a explicação das correlações entre os itens, possibilitando fazer previsões das respostas dos participantes aos itens a partir das curvas dos itens. Diferentes padrões de resposta podem ser esperados quando os itens são completamente independentes, ou quando medem um único fator ou mais de um fator. A adequação de um modelo unidimensional ou multidimensional aos padrões de respostas dos participantes é verificada pelo teste Qui-quadrado. É imprescindível verificar, por este mesmo teste, se a inclusão sucessiva de fatores nos modelos tem um efeito significativo (Bock, Gibbons & Muraki, 1988). Nas tabelas 5 e 6 estão apresentadas as cargas fatoriais dos itens nos três fatores considerados para a análise fatorial dos dados.

Os resultados da análise fatorial com informação completa pelo método de rotação PROMAX para os itens do componente de Formação Geral, considerando o modelo unidimensional, indicou que 16,2% da variância total pode ser explicada por esse modelo. Incluindo-se um segundo fator, modelo bidimensional, a variância explicada aumenta para 23,5%.

Para se obter uma prova com maior consistência interna poderiam ser excluídos da prova os itens com cargas fatoriais inferiores a 0,30 em todos os fatores, no componente de Formação Geral o item 4 conforme dados apresentados na Tabela 5, a seguir.

Tabela 5. Cargas não rotacionadas dos fatores principais do componente de Formação Geral

Item	Comunalidade	Fator 1	Fator 2	Fator 3
1	0,210	0,394	-0,030	0,232
2	0,289	0,534	-0,033	-0,048
3	0,999	0,901	-0,103	-0,421
4	0,089	0,189	-0,018	0,230
5	0,264	0,351	-0,019	0,375
6	0,176	0,216	-0,032	0,359
7	1,000	0,145	0,989	-0,005
8	0,153	0,177	-0,029	0,348

A análise fatorial com informação completa pelo método de rotação PROMAX, considerando-se 3 fatores na análise (Tabela 5) indicou que os mesmos são correlacionados entre si ($r_{12} = 0,55$; $r_{13} = 0,32$; $r_{23} = 0,35$). Além disso, pouca explicação é acrescentada por cada uma das dimensões, resultado que implicou em considerar neste trabalho a predominância do primeiro fator, ou seja, a unidimensionalidade do componente de Formação Geral.

O modelo unidimensional para os dados do componente de Formação Específica, pode explicar 20,4% da variância entre os itens e incluindo um segundo fator, modelo bidimensional, a variância explicada aumenta para 23,5%.

Ao se considerar um modelo tridimensional, a variância total explicada aumenta muito pouco, para ambos os tipos de componentes, a magnitude da diferença entre os padrões de resposta observados e os reproduzidos pelos modelos, quando se acrescenta um terceiro fator, não é significativa, podendo supor que ele não seja necessário.

Também para o componente de Formação Específica os itens com cargas fatoriais inferiores a 0,30 em todos os fatores deveriam ser excluídos da prova, caso se desejasse obter uma prova com melhor consistência interna entre os itens. A análise fatorial com informação completa pelo método de rotação PROMAX indicou que os 3 fatores considerados na análise (Tabela 6) são correlacionados entre si ($r_{12}=0,538$; $r_{13}=0,465$; $r_{23}=0,410$) embasando a unidimensionalidade do componente, dado que a cada dimensão pouca variância explicada foi acrescida.

Tabela 6. Cargas não rotacionadas dos fatores principais do componente de Formação Específica

Item	Comunalidade	Fator 1	Fator 2	Fator 3
9	0,317	0,406	-0,363	0,140
10	0,294	0,499	-0,199	0,072
11	0,215	0,322	-0,007	-0,334
12	0,058	-0,029	-0,199	-0,134
13	0,476	0,539	-0,123	-0,413
14	0,061	-0,045	-0,184	-0,159
15	0,183	0,356	-0,235	-0,030
16	0,249	0,474	-0,140	-0,071
17	0,139	0,369	0,048	0,012
18	0,126	0,348	0,037	0,062
19	0,063	0,178	-0,141	0,108
20	0,605	0,754	0,184	0,051
21	0,030	0,070	0,108	-0,114
22	0,317	0,532	0,159	-0,095
23	0,439	0,638	0,181	0,015
24	0,166	0,212	-0,305	0,168
25	0,759	0,825	0,230	0,162
26	0,077	0,167	0,069	0,032
27	0,233	0,477	0,063	-0,038
28	0,344	0,554	-0,134	0,138

Foram analisados os ajustes dos dados aos modelos da TRI considerando as 8 questões objetivas do componente de Formação Geral e as 20 questões objetivas do componente de Formação Específica. Os resultados das estatísticas gerais dos itens indicam a adequabilidade ou não dos itens aos modelos propostos. Essas estatísticas são calculadas a partir do agrupamento dos escores de habilidades semelhantes, da probabilidade de acerto teórica para cada subgrupo, da curva característica do item e dos resíduos relativos à probabilidade real observada.

Ao se proceder a estimação dos parâmetros dos itens do ENADE por meio dos modelos logísticos de um, modelo de Rasch, de 2 e de 3 parâmetros da TRI, os resultados para os componentes de Formação Geral e Formação Específica, estimados para a população de estudantes do curso de Medicina são descritos nas Tabelas 7 e 8 apresentadas a seguir. São apresentados os resultados da estimação dos parâmetros dos itens calculados pelos programas específicos *WINSTEPS* e *XCALIBRE*. Em cada uma das tabelas a primeira coluna apresenta os itens do componente analisado, componente de Formação Geral e componente de Formação Específica; a segunda coluna apresenta o índice de dificuldade b calculado usando o modelo de Rasch; a terceira e quarta, os índices de ajuste do modelo de Rasch (*infit* e *outfit*); a quinta e sexta coluna, apresentam os parâmetros de discriminação a e dificuldade b calculados usando o modelo de 2-parâmetros; a sétima coluna apresenta o índice de ajuste ao modelo de 2-parâmetros (Resíduo); a oitava, a nona e a décima colunas apresentam os parâmetros de discriminação a , dificuldade b e probabilidade de acerto ao acaso c calculados usando o modelo de 3-parâmetros; a décima primeira coluna apresenta o índice de ajuste ao modelo de 3-parâmetros (Resíduo); a décima segunda coluna mostra a proporção de estudantes na amostra que responderam ao item corretamente (PC); a décima terceira coluna apresenta a correlação item-total pelo produto momento de Pearson (PBs) e a décima quarta coluna apresenta uma medida analógica, a correlação item-total na qual se substitui o escore total pelo traço latente (θ) subjacente (PBt).

Os resultados apresentados nas Tabelas 7 e 8 indicam que os dados dos componentes do ENADE para o curso de Medicina se ajustam melhor ao modelo de Rasch (1 parâmetro). Apesar de nos modelos de 2 e 3-parâmetros serem observados alguns itens desajustados quando comparados os resíduos para os parâmetros dos itens calculados pelos dois modelos (colunas 6 e 10). Os valores calculados pelo modelo de 3-parâmetros (coluna 10) são quase sempre menores que os valores encontrados para os outros dois modelos, o que corresponde ao esperado, uma vez

que quando há aumento do número de parâmetros analisados se flexibiliza o procedimento de cálculo e o ajuste do modelo é melhor (Baker, 2001).

Nem todos os itens do componente de Formação Geral (Tabela 7) apresentam ajuste adequado aos modelos de dois e três parâmetros, uma vez que o item 1 apresenta resíduo maior que 2,0 e os itens 1 (para ambos os modelos) e 8 apresentam valores de b abaixo do limite crítico -2,95 ou acima de 2,95. Na estimação dos parâmetros pelo modelo de Rasch observou-se que todos os itens se ajustam ao modelo, pois a quantidade de itens é pequena e nos casos apresentados o ajuste não viola os critérios assumidos. A análise do *infit* e *outfit* apresenta resultados dentro do valor crítico esperado, entre 0,70 e 1,30, permitindo a conclusão de que o modelo de 1 parâmetro (modelo de Rasch) se ajusta melhor aos dados do que os demais modelos utilizados para o ajuste do componente de Formação Geral.

No componente de Formação Geral foram encontrados 5 itens desajustados quando o modelo usado para estimação foi o de 2 parâmetros, eles aparecem em negrito na quarta, quinta e sexta coluna e correspondem aos itens 1, 4, 6, 7 e 8 e 4 itens desajustados quando o modelo usado foi o de 3 parâmetros, que aparecem em negrito nas colunas sétima e oitava e correspondem aos itens 1, 4, 7 e 8. Quando usado o modelo de Rasch observou-se todos os itens estão ajustados.

Ainda na Tabela 7 pode-se observar na décima primeira coluna a proporção de estudantes que responderam corretamente o item, permitindo concluir que o item 1 foi o item mais fácil para essa amostra com 91% de estudantes que acertaram esse item e o item 6 o que os estudantes mais erraram com 40% de estudantes acertando o item.

Tabela 7. Parâmetros dos Itens do componente de Formação Geral, modelos de 1, 2 e 3 parâmetros

Item	Modelo de Rasch			Modelo de 2 parâmetros			Modelo de 3 parâmetros							
	b	Infit	Outfit	a	b	Resíduo	a	b	c	Resíduo	PC	PBs	PBt	N
1	-1,85	0,97	0,86	0,49	-3,00	2,74	0,45	-3,00	0,25	1,13	0,91	0,34	0,36	8107
2	0,39	0,91	0,89	0,57	-0,32	0,47	0,93	0,18	0,18	0,75	0,55	0,53	0,75	8107
3	-0,26	0,93	0,89	0,58	-1,14	0,57	0,91	-0,48	0,22	0,55	0,69	0,50	0,73	8107
4	0,47	1,04	1,05	0,22	-0,43	1,19	0,25	1,42	0,27	0,65	0,54	0,43	0,24	8107
5	-0,81	0,97	0,95	0,31	-2,82	0,54	0,31	-2,00	0,25	0,45	0,79	0,43	0,34	8107
6	1,09	1,03	1,06	0,20	1,24	0,98	0,36	2,85	0,27	0,55	0,40	0,43	0,23	8107
7	-0,36	1,08	1,12	0,20	-2,06	1,71	0,20	-0,21	0,27	0,89	0,65	0,38	0,21	8107
8	1,01	1,04	1,09	0,20	1,08	1,33	0,32	3,00	0,27	1,09	0,42	0,43	0,20	8107

Os resultados apresentados na Tabela 7 podem ser mais bem visualizados pela Curvas Características dos Itens (CCI) apresentadas nas Figuras 2 e 3, para os modelos de um e três parâmetros, respectivamente.

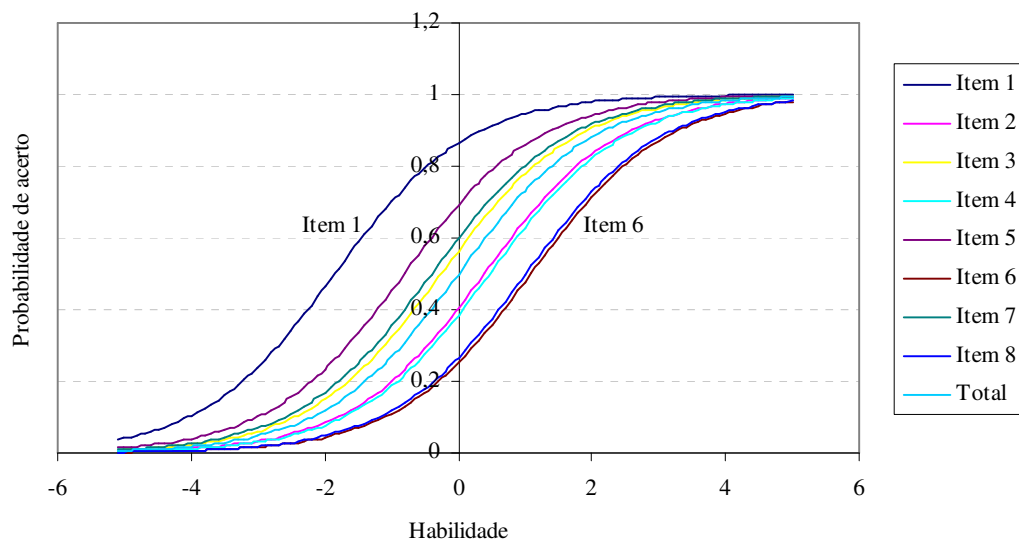


Figura 2. Curvas características dos oito itens do componente de Formação Geral, modelo de um parâmetro

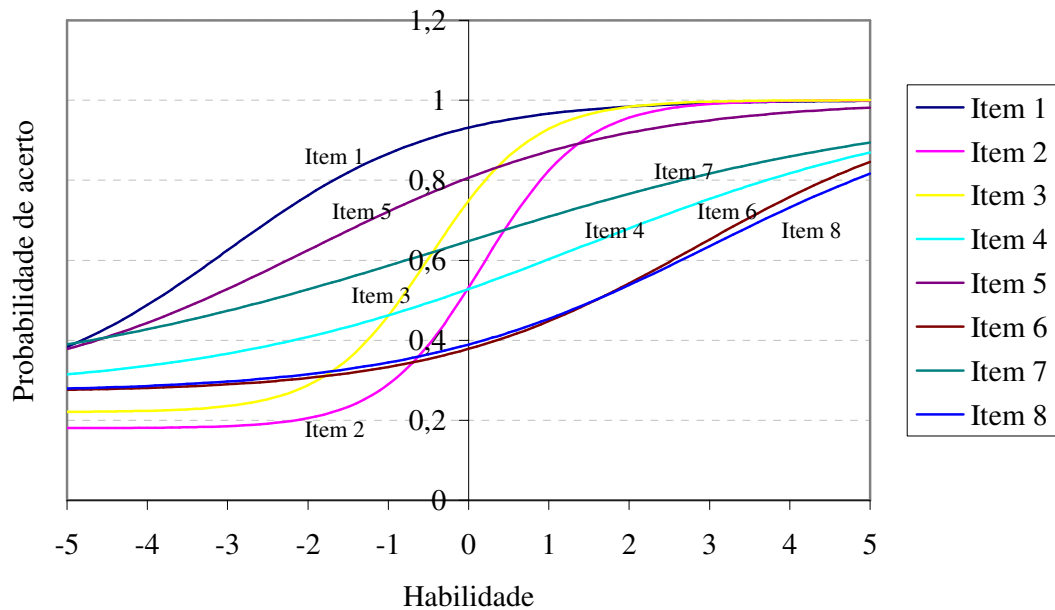


Figura 3. Curvas características dos oito itens do componente de Formação Geral, modelo de três parâmetros

Os dados do componente de Formação Específica (Tabela 8) também não apresentam ajuste adequado ao modelo de dois parâmetros, uma vez que os itens 11, 12, 13, 14, 20, 21 e 22 apresentam resíduos maiores que 2,0 e os itens 14 e 21 valores de b acima de 2,95. No modelo de 3 parâmetros os itens 12 e 14 apresentam resíduos maiores que 2,0 e os itens 12, 14 e 21 valores de b acima de 2,95, apresentando um ajuste melhor que o modelo de 2 parâmetros mas ainda não adequado aos dados, se comparado às demais análises realizadas. Na estimação dos parâmetros pelo modelo de Rasch encontra-se maior número de itens ajustados podendo considerar que o modelo de 1 parâmetro é o que apresenta o ajuste mais adequado.

Tabela 8. Parâmetros dos Itens do componente de Formação Específica, modelos de 1, 2 e 3 parâmetros

Item	Modelo de Rasch			Modelo de 2 parâmetros			Modelo de 3 parâmetros							
	b	infit	outfit	a	b	resíduo	a	b	c	Resíduo	PC	PBs	PBt	N
9	-2,53	0,96	0,86	0,49	-2,94	0,85	0,56	-2,51	0,17	1,31	0,90	0,29	0,32	8107
10	-2,52	0,95	0,76	0,57	-2,65	1,24	0,69	-2,18	0,17	0,58	0,90	0,31	0,38	8107
11	0,90	1,02	1,09	0,35	1,61	2,76	1,26	1,76	0,20	0,57	0,29	0,36	0,28	8107
12	0,70	1,26	1,33	0,20	2,85	7,58	1,30	3,00	0,27	4,73	0,33	0,13	-0,01	8107
13	1,05	0,90	0,90	0,63	1,22	2,16	1,23	1,43	0,13	0,59	0,26	0,48	0,44	8107
14	1,01	1,26	1,41	0,20	3,00	5,40	1,27	3,00	0,26	2,81	0,27	0,12	-0,02	8107
15	-1,22	0,99	0,97	0,39	-1,68	1,84	0,42	-1,11	0,18	1,57	0,73	0,36	0,33	8107
16	-0,72	0,94	0,91	0,52	-0,74	1,52	0,60	-0,20	0,19	0,59	0,63	0,45	0,44	8107
17	1,43	1,00	1,07	0,43	2,14	0,91	0,78	2,12	0,11	1,33	0,20	0,35	0,31	8107
18	0,40	1,03	1,06	0,39	0,77	1,18	0,57	1,33	0,17	0,81	0,39	0,38	0,34	8107
19	-0,33	1,12	1,16	0,22	-0,51	1,43	0,28	0,62	0,20	1,31	0,55	0,28	0,20	8107
20	0,23	0,80	0,76	1,11	0,25	2,07	1,94	0,53	0,13	1,05	0,42	0,61	0,69	8107
21	1,52	1,16	1,44	0,25	3,00	4,58	0,96	3,00	0,17	1,07	0,19	0,17	0,08	8107
22	0,25	0,92	0,92	0,64	0,37	2,24	1,15	0,86	0,19	0,79	0,42	0,49	0,49	8107
23	0,63	0,86	0,85	0,83	0,65	1,93	1,36	0,90	0,12	1,56	0,34	0,55	0,58	8107
24	-1,13	1,06	1,19	0,26	-2,11	1,50	0,30	-1,30	0,17	1,87	0,71	0,29	0,24	8107
25	-0,01	0,78	0,74	1,24	0,05	1,91	2,18	0,32	0,13	1,11	0,48	0,63	0,73	8107
26	1,18	1,08	1,14	0,31	2,31	0,65	0,59	2,45	0,14	0,80	0,24	0,30	0,24	8107
27	0,60	0,95	0,95	0,57	0,80	1,49	0,88	1,19	0,15	0,68	0,35	0,46	0,44	8107
28	-1,45	0,92	0,81	0,67	-1,36	1,47	0,82	-0,95	0,18	0,51	0,77	0,44	0,52	8107

No componente de Formação Específica foram encontrados 9 itens desajustados quando o modelo usado para estimação foi o de 2 parâmetros e 5 itens desajustados quando o modelo usado para estimação foi o modelo de 3 parâmetros. Quando usado o modelo de Rasch observou-se apenas 3 itens desajustados, eles aparecem em negrito na quarta coluna e correspondem aos itens 12, 14, e 21. A análise dos outfit para esses itens apresenta resultados fora do valor crítico, de 0,70 a 1,30. Os resultados apresentados podem ser melhor visualizados nas figuras a seguir.

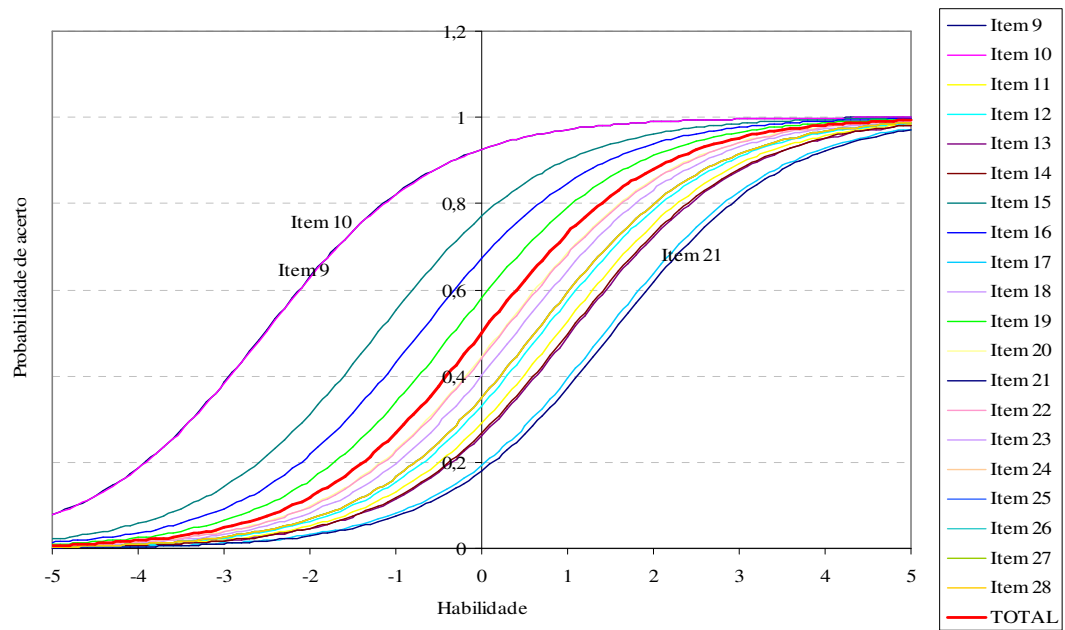


Figura 4. Curvas características dos 20 itens do componente de Formação Específica, modelo de um parâmetro

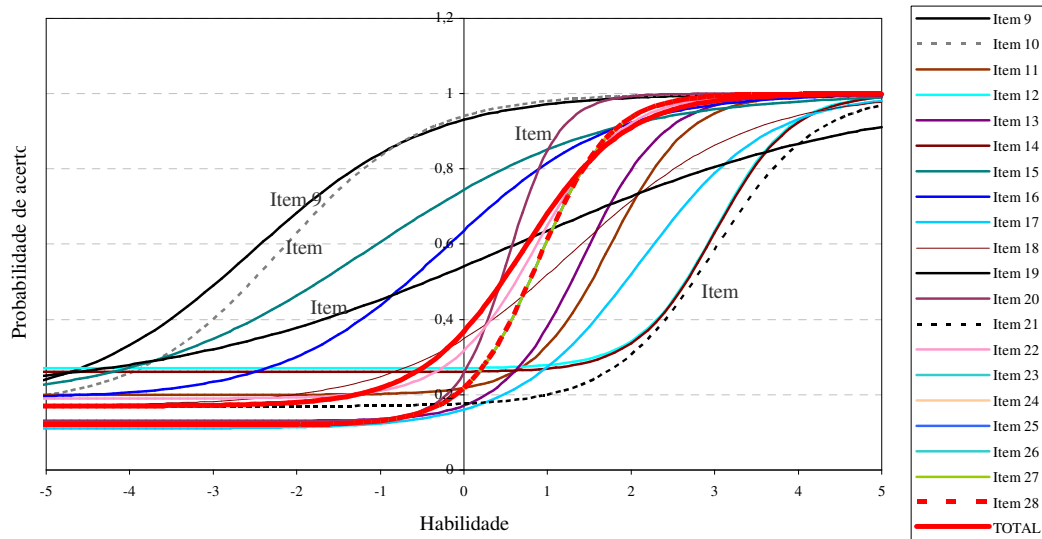


Figura 5. Curvas características dos 20 itens do componente de Formação Específica, modelo de três parâmetros

Em seguida foi realizado o exame dos índices de fidedignidade para cada componente do ENADE estimado pelos modelos de 1, 2 e 3 parâmetros. Pode-se observar que em modo geral os componentes apresentam uma quantidade máxima de informação em um valor Theta. Por outro lado a quantidade máxima de informação tanto no componente de Formação Geral quanto no de Formação Específica é maior no modelo de Rasch, de acordo com as Tabelas 9 e 10.

A Tabela 9 apresenta informações mais gerais do componente de Formação Geral sobre os valores da fidedignidade de cada componente calculados pela fórmula *KR-21*, juntamente com valores máximo de informação estimada para os estudantes. Pode-se observar que de modo geral a fidedignidade é mais alta para o modelo de Rasch, assim como a informação esperada e a média de informação.

Tabela 9. Características do componente de Formação Geral

Formação Geral			
Modelo	Fidedignidade KR-21	Informação Esperada	Média de Informação
Rasch	0,586	1,477	1,146
2 parâmetros	0,352	0,571	0,495
3 parâmetros	0,352	0,667	0,445

A Tabela 10 apresenta informações sobre os valores da fidedignidade do componente de Formação Específica calculados pela fórmula *KR-21*, juntamente com valores máximo de informação estimada para os estudantes. Pode-se observar que aqui também a fidedignidade é mais alta para o modelo de Rasch, apesar de a informação esperada e a média de informação ser maior para o modelo de 3 parâmetros a diferença é mínima em relação às outras análises realizadas.

Tabela 10. Características do componente – Formação Específica

Formação Específico			
Modelo	Fidedignidade KR-21	Informação Esperada	Média de Informação
Rasch	0,770	3,459	2,782
2 parâmetros	0,673	3,116	2,246
3 parâmetros	0,673	3,575	2,796

DISCUSSÃO E CONSIDERAÇÕES FINAIS

A aplicação da TRI no estudo, construção e validação de testes psicológicos e educacionais tem sido muito utilizada por vários pesquisadores e instituições, como no SAT-I, II e III, no ACT, no SARESP e outros (Anastasi & Urbina, 2000; Andrade, Tavares e Valle, 2000; Pasquali, 2003; Primi & Moraes; Vendramini, 2005; dentre outros). Esse método de análise muito tem contribuído para enriquecer as análises feitas referentes aos itens de instrumentos psicológicos ou educacionais, razão pela qual se optou por essa teoria para a análise dos itens dos componentes do ENADE.

As análises realizadas com todos os itens confirmaram o não-ajuste do reduzido número dos itens que compunham a segunda e terceira dimensão, optando-se apenas pela análise da dimensão predominante. Segundo Embretson (1999), as diferentes dimensões podem ser analisadas separadamente, na falta de programas específicos para uma análise multidimensional, embora já existam programas computacionais específicos para se proceder a essa análise.

Assim, os dois pressupostos para a aplicação da TRI foram comprovados, a unidimensionalidade com 16,2% de explicação para o componente de FG e 20,4% de explicação para o componente de FE com os fatores correlacionando-se entre si. E a independência local, que segundo Hambleton, Swaminathan e Rogers (1991), se a unidimensionalidade é cumprida, disto deriva, matematicamente, a independência local entre os itens, dado que os dois conceitos são equivalentes. Resultados estes que vão ao encontro aos estudos feitos por Andriola (1998) e Vendramini e Dias (2005).

O componente de Formação Específica apresentou índice de fidedignidade (KR-20=0,681) mais satisfatório que o de FG (KR-20=0,391). Em um estudo de validade para o

ENADE realizado por MEC (2006), resultados semelhantes foram apresentados, no qual foram encontrados um bom índice de confiabilidade para o componente de FE (*alfa de cronbach* = 0,834) e um índice mais baixo para o componente de FG (*alfa de cronbach* = 0,462), indicando que o componente de FE difere mais do que o de FG os alunos ingressantes dos concluintes, o que confirma os resultados encontrados no presente estudo.

Constatou-se que, segundo a TCT, duas das 8 questões propostas no componente de Formação Geral apresentaram índices de discriminação baixos (menores ou iguais a 10,0). O índice de maior discriminação (13,94) foi o da questão 6. No componente de Formação Específica três dos 20 itens apresentaram índices de discriminação baixos (menores ou iguais a 10,0). O índice de maior discriminação (16,48) foi o da questão 21. De acordo com a TRI, verificou-se que apenas o item 1, para o componente de Formação Geral, apontados pela TCT como de baixa discriminação tiveram seus ajustes rejeitados segundo a interpretação do modelo de 2 e 3 parâmetros e o item 9, para o componente de Formação Específica, segundo a interpretação do modelo de Rasch.

Como descrito na sessão anterior, os parâmetros dos modelos da TRI, são estimados inicialmente com base na proporção de acertos e correlação bisserial, calculadas pela Teoria Clássica, e refinados segundo procedimentos estatísticos que buscam estimadores de máxima verossimilhança e de resíduos mínimos. Os índices de dificuldade b são transformados inicialmente em uma escala, que varia de -3 (itens extremamente fáceis) a 3 (itens extremamente difíceis); os de discriminação a , que variam de 0,5 (baixa discriminação) a 2,0 (alta discriminação); a probabilidade de acerto por acaso c , a partir do número de alternativas do item, neste estudo $c = 0,2$ (uma de cinco alternativas).

Quando comparados os índices de facilidade da TCT com os de dificuldade da TRI, observa-se que, quanto maior a proporção de acertos no item, menor tende a ser o seu índice de

dificuldade. A vantagem da TRI é que, conhecida a habilidade de um indivíduo, não necessariamente participante da amostra, é possível determinar a probabilidade de ele acertar um item. Isto não ocorre com a TCT, cujos resultados são dependentes da amostra. Os itens da prova do ENADE são de nível mediano, não são muito difíceis ($1,30 < b_i < 3,00$) e não muito fáceis ($-3,00 < b_i < -0,88$). O que vai ao encontro também aos estudos realizados por Primi (2005), no qual é possível verificar que as provas do ENADE são sensíveis em detectar as diferenças de desempenho entre os estudantes e também verificar as diferenças entre os cursos. Resultados estes que também se assemelham aos estudos apresentados por Santos e cols. (2000), no qual os resultados também apontam diferenças de desempenho de acordo com os cursos, ou seja, cursos mais concorridos apresentam melhor desempenho nas provas, apoiando a idéia da relação entre conhecimento e raciocínio.

Os índices de discriminação da TCT são calculados a partir da diferença entre a proporção de acertos dos 27% dos participantes com maior pontuação total e a proporção de acertos dos 27% de menor pontuação total. Na TRI, esses índices são estimados a partir da correlação bisserial item total, e diferem dos encontrados pela TCT (Muñiz, 1990).

A literatura especializada aponta algumas vantagens da análise psicométrica de itens utilizando o modelo de resposta ao item (Embretson, 1996). São elas, a independência dos parâmetros em relação à amostra da população utilizada em sua estimação, a independência dos resultados dos sujeitos do tipo de teste usado e a possibilidade de não se considerar o escore total como referência de medida (Muñiz, 1990; Baker, 2001). Também os estudos realizados por Andriola (1998) e Andrade, Tavares e Valle (2000), demonstram as vantagens de se utilizar a TRI na análise qualitativa do nível de desempenho, permitindo analisar as diferenças de desempenho dos estudantes de uma série para outra.

De um modo geral pode-se assumir que os dados se ajustam melhor ao modelo de Rasch nos dois componentes do ENADE, conforme as análises de resíduos e Qui-quadrado, embora os dados também se ajustem ao modelo de 3 parâmetros, mas com menor confiabilidade e com maior número de itens que não atendem aos critérios mínimos estabelecidos para os parâmetros deste modelo. A vantagem conseguida pela utilização do modelo de Rasch e o fato de que a precisão da demonstração da invariância dos parâmetros para os parâmetros a e c não acrescentam informação reforça a recomendação feita pela literatura especializada de que se deve adotar o modelo de Rasch (Muñiz, 1990). Neste estudo chegou-se à conclusão de que o melhor modelo para representar os dados é o de Rasch, o que vai ao encontro à recomendação desses autores.

Com relação à informação dada pela prova observou-se que de um modo geral os componentes do ENADE apresentam maior fidedignidade para os ajustes ao modelo de Rasch. As curvas de informação mostram que de um modo geral as provas apresentam maior informação para um nível de habilidade mais alta para o componente de Formação Específica que no componente de Formação Geral. De acordo com o estudo do MEC (2006), pode-se perceber que nos itens do componente de FE existe um padrão de interação que evidência a validade da prova, uma vez que a diferença ingressante concluinte é maior em cursos com níveis de desempenho mais elevados. Segundo Baker (2001) o máximo de informação alcançado por uma prova é obtido quando essa é aplicada a sujeitos que possuem uma habilidade equivalente ao índice de dificuldade de seus itens. As variações encontradas nos índices de informação podem estar ligadas a uma diferença na relação habilidade/dificuldade dos componentes do ENADE para estudantes ingressantes e concluintes.

Não se pretendeu com este estudo analisar qualitativamente as questões que compõem a prova de Medicina, mas fornecer subsídios para que pesquisadores desta área complementem

seus estudos com informações sobre a prova como um todo e sobre cada um de seus itens. Além disso, pela TRI é possível analisar conjuntamente a dificuldade dos itens e a habilidade dos estudantes subjacentes ao desempenho na prova.

Embora, neste estudo tenha-se optado por analisar a qualidade da prova para a população brasileira como um todo, outros estudos podem ser realizados para verificar se existiriam diferenças nos parâmetros ajustados, quando se considerassem outras variáveis tais como região, categoria administrativa, nível sócio-econômico dos estudantes, entre outras.

REFERÊNCIAS

- Almeida, L. S. (1994). *Inteligência: definição e medida*. Porto: Instituto Nacional de Investigação Científica.
- Almeida, L. S., & Primi, R. (1998). *Baterias de Provas de Raciocínio (BPR-5): manual técnico*. São Paulo: Casa do Psicólogo.
- Almeida, L. S., & Primi, R. (2001). Teoria de Resposta ao Item. Em E. M. Fernandes & L. S. Almeida (Eds.). *Métodos e Técnicas de Avaliação: contribuindo para a prática e investigação psicológicas* (pp. 205-232). Braga, Portugal: Centro de Estudos em Educação e Psicologia, Universidade do Minho.
- American Educational Research Association-AERA, American Psychological Association-APA, & National Council on Measurement in Education-NCTM (1999). *Standards of Educational and Psychological testing*. Washington, DC: American Educational Research Association.
- Anastasi, A., & Urbina, S. (2000). *Testagem Psicológica*. Porto Alegre: Artes Médicas.
- Andrade, D. F., Tavares, H. R., & Valle, R. C. (2000) *Teoria da Resposta ao Item: conceitos e aplicações*. São Paulo: Associação Brasileira de Estatística.
- Andrade, D. F. (2001). Comparando desempenhos de grupos de estudantes por intermédio da Teoria de Resposta ao Item. *Estudos em Avaliação Educacional*, 23(2), 31-69.
- Andriola, W. B. (1998). Avaliação da aprendizagem: uma análise descritiva segundo a teoria de resposta ao item (TRI). *Educação em Debate*, 20(36), 93-102.

Andriola, W. B. (1998). Utilização da Teoria de Resposta ao Item (TRI) para a organização de um banco de itens destinados a avaliação do raciocínio verbal. *Psicologia: Reflexão e Crítica*, 11(2), 295-308.

Araújo, C. M. M. (2004). O Desenvolvimento de Competências no ENADE: a mediação da avaliação nos processos de desenvolvimento psicológico e profissional. *Revista de Avaliação Institucional*, 09(4), 77-97.

Assessment System Corporation (1995). *User's manual for the XCALIBRE-Marginal Maximum-Likelihood IRT Parameter Estimation Program*. 2ed., Windows 3.x/95/NT, St. Paul, MN: Author.

Baker, F. B. (2001). *The Basics of Item Response Theory*. USA: ERIC Clearinghouse on Assessment and Evaluation.

Barreyro, G. B. (2004). Do Provão ao SINAES: o processo de construção de um novo modelo de avaliação da Educação Superior. *Avaliação Institucional*, 09(01), 37-49.

Bertolin, J. C. G. (2004). A Transformação do SINAES: da proposta emancipatória à lei híbrida. *Avaliação Institucional*, 09(04), 67-76.

Brito, M. R., Munhoz, A., Primi, R., Gonçalves M. H., Rezi, V., Neves, L. F., Sanches, M. H., & Marinheiro, F. B. (2000). Exames nacionais: Uma análise do ENEM aplicado à Matemática. *Revista Avaliação*, 5(4), 445-54.

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.

- Dorsch, F., Hacker, H., & Stapf K. H. (2001). *Dicionário de Psicologia Dorsch*. Rio de Janeiro: Vozes.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, 8(4), 341-349.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologist*. New Jersey: Lawrence Erlbaum.
- Ferrão, M. E. (2003). *Introdução aos modelos de regressão multinível em Educação*. Campinas, São Paulo: Komedi.
- Flanagan, D. P., McGrew, K. S. & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: a contemporary approach to interpretation*. Needham Heights: Allyn and Bacon.
- Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2002). *The achievement test desk reference (ATDR): Comprehensive Assessment and Learning Disabilities*. Needham Heights: Allyn and Bacon.
- Fletcher, P. R. (1994). A Teoria da Resposta ao Item: medidas invariantes do desempenho escolar. *Ensaio* 2(1), 21-28.
- Hambleton, R. K. (1990). Item Response Theory: introduction and bibliography. *Psicothema*, 2(1), 97-107.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage Publishers.
- Kulieke, M., Bakker, J., Collins, C., Fennimore, T., Fine, C., Herman, J., Jones, B. F., Raack, L., & Tinzmann, M. B., (1990). *Why Should Assessment Be Based on a Vision of Learning*. NCREL, Oak Brook.

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

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Needham Heights: Allyn & Bacon.

Ministério da Educação e Cultura (2004a). *Diretrizes para a Avaliação das Instituições de Educação Superior*. Brasília: Instituto Nacional de Estudos e pesquisas Educacionais Anísio Teixeira. 48 p.

Ministério da Educação e Cultura (2004b) – Instituto Nacional de Estudos e pesquisas Educacionais Anísio Teixeira (INEP). *LEI N° 10.861, de 14 de Abril de 2004*. Em: www.inep.gov.br/superior/enade.

Ministério da Educação e Cultura (2004c) – Instituto Nacional de Estudos e pesquisas Educacionais Anísio Teixeira (INEP). *PORTARIA N° 2.051, de 09 de Julho de 2004*. Em: www.inep.gov.br/superior/enade.

Ministério da Educação e Cultura (2004d). *SINAES – Sistema Nacional de Avaliação da Educação Superior: da concepção à regulamentação*. (2. ed. ampl.). Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. 155 p.

Ministério da Educação e Cultura (2004e). *Resumo Técnico – ENADE 2004*. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

- Munhoz, A. M. H. (2004). *Uma análise multidisciplinar da relação entre inteligência e desempenho acadêmico em universitários ingressantes*. Tese de Doutorado. Universidade Estadual de Campinas, Faculdade de Educação, Campinas.
- Muñiz, J. (1990). *Teoría de Respuesta a los Ítems: un nuevo enfoque en la evolución psicológica y educativa*. Madrid: Ediciones Pirámide, S. A.
- Muñiz, J. (1994). *Teoría Clássica de los Tests*. Madrid: Ediciones Pirâmides, S. A.
- Pasquali, L. (1996). *Teoria e Métodos de Medida em Ciência do Comportamento*. Brasília: INEP.
- Pasquali, L. (1997). *Psicometria: teoria e Aplicações – A Teoria Clássica dos Testes Psicológicos*. Brasília: Editora UnB.
- Pasquali, L. (2000). *Psicometria: Teoria dos Testes Psicológicos*. Brasília: LabPAM.
- Pasquali, L. (2003). *Psicometria: Teoria dos testes na psicologia e na educação*. Petrópolis, Rio de Janeiro: Vozes.
- Primi, R. (1998). *Desenvolvimento de um instrumento informatizado para a avaliação do raciocínio analítico*. Tese de Doutorado, Universidade de São Paulo, São Paulo.
- Primi, R., Vendramini, C. M. M., Santos, A. A. A., & Figueiredo Filho, N. (1999). Impacto de variáveis sócio econômicas no desempenho de candidatos ao Ensino Superior. Em A. P. Soares, S. Araújo, & S. Cairis (Eds.), *Avaliação Psicológica, Formação e Contextos* (pp. 195-202). Braga: 1999.
- 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.

- Primi, R., Santos, A. A. A., & Vendramini, C. M. M. (2002). Habilidades básicas e desempenho acadêmico em universitários ingressantes. *Estudos de Psicologia*, 7(1), 47-55.
- 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., & Moraes, I. F. (2005). Validade de um Instrumento de Desempenho de Docentes do Ensino Superior. *Revista de Avaliação Institucional*, 10(3), 41-59.
- Santos, A. A. A., Primi, R., Vendramini, C. M. M., Taxa, F., Lukjanenko, M. F., Muller, F., Sampaio, I., Andraus Jr., S., Kuse, F. K., & Bueno, C. H. (2000) Habilidades básicas de ingressantes universitários. *Revista de Avaliação Institucional*, 2(16), 33-45.
- Santos, A. A. A., Primi, R., Vendramini, C. M. M., Taxa, F. (2002). O Teste de Cloze na avaliação da compreensão em leitura. *Psicologia Reflexão e Crítica*. 15(3), 549-560.
- Souza, C. P. (1997). Avaliação do rendimento escolar: sedimentação de significados. Em C. P. Souza (Org). *Avaliação do rendimento escolar* (6ª ed). Campinas: Papyrus.
- Souza, N. A. (2005). Avaliação de Competencias: o aperfeiçoamento profissional na área de enfermagem. *Estudos em Avaliação Educacional*, 16(32), 57-80.
- Sternberg, R. J. (1990). *Metaphors of mind: conceptions of the nature of intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R. J. (2000). *Psicologia cognitiva*. Trad. Maria Regina Borges Osório. Porto Alegre: Artes Médicas.
- Vendramini, C. M. M. (2000). *Implicações das atitudes e das habilidades matemáticas na aprendizagem dos conceitos de estatística*. Tese de Doutorado, Universidade Estadual de Campinas, Campinas.

Vendramini, C. M. M. (2001). Aplicação da Teoria de Resposta ao Item na Avaliação Educacional. Em Primi, R. (Org). *Temas em Avaliação Psicológica* (pp. 116-130). Campinas: Instituto Brasileiro de Avaliação Psicológica.

Vendramini, C. M. M., Silva, M. C., & Canale, M. (2004) Análise de itens de uma prova de raciocínio estatístico. *Psicologia em Estudo*, 9(3), 487-498.

Vendramini, C. M. M. (2005). O Uso de um Modelo Multidimensional para Análise do ENADE. *Revista de Avaliação Institucional*, 10(3), 27-40.

Vendramini, C. M. M. & Dias, A. S. (2005). Teoria de resposta ao item na análise de uma prova de estatística em universitários. *Psico-USF*, 10(2), 201-210.

Wilson, D. T., Wood, R., & Gibbons, R. (1998). *TESTFACT 2 – test scoring, item statistics, and item factor*. Chicago: Scientific Software International.

ANEXOS

Anexo 1. Prova do ENADE – Componente de Formação Geral

TEXTO I

“O homem se tornou lobo para o homem, porque a meta do desenvolvimento industrial está concentrada num objeto e não no ser humano. A tecnologia e a própria ciência não respeitaram os valores éticos e, por isso, não tiveram respeito algum para o humanismo. Para a convivência. Para o sentido mesmo da existência.

Na própria política, o que contou no pós-guerra foi o êxito econômico e, muito pouco, a justiça social e o cultivo da verdadeira imagem do homem. Fomos vítimas da ganância e da máquina. Das cifras. E, assim, perdemos o sentido autêntico da confiança, da fé, do amor. As máquinas andaram por cima da plantinha sempre tenra da esperança. E foi o caos”.

ARNS, Paulo Evaristo. **Em favor do homem**. Rio de Janeiro: Avenir, s/d. p.10.

Questão 1

De acordo com o texto, pode-se afirmar que

- (A) a industrialização, embora respeite os valores éticos, não visa ao homem.
- (B) a confiança, a fé, a ganância e o amor se impõem para uma convivência possível.
- (C) a política do pós-guerra eliminou totalmente a esperança entre os homens.
- (D) o sentido da existência encontra-se instalado no êxito econômico e no conforto.
- (E) o desenvolvimento tecnológico e científico não respeitou o humanismo.

TEXTO II



Questão 2

A charge de Millôr aponta para:

- (A) a fragilidade dos princípios morais.
- (B) a defesa das convicções políticas.

- (C) a persuasão como estratégia de convencimento.
- (D) o predomínio do econômico sobre o ético.
- (E) o desrespeito às relações profissionais.

Questão 3

A charge de Millôr e o texto de Dom Paulo Evaristo Arns tratam, em comum,

- (A) do total desrespeito às tradições religiosas e éticas.
- (B) da defesa das convicções morais diante da corrupção.
- (C) da ênfase no êxito econômico acima de qualquer coisa.
- (D) da perda dos valores éticos nos tempos modernos.
- (E) da perda da fé e da esperança num mundo globalizado.

Questão 4

“Os determinantes da globalização podem ser agrupados em três conjuntos de fatores: tecnológicos, institucionais e sistêmicos”.

GONÇALVES, Reinaldo. **Globalização e Desnacionalização**. São Paulo: Paz e Terra, 1999.

“A ortodoxia neoliberal não se verifica apenas no campo econômico. Infelizmente, no campo social, tanto no âmbito das idéias como no terreno das políticas, o neoliberalismo fez estragos (...).

SOARES, Laura T. **O Desastre Social**. Rio de Janeiro: Record, 2003.

“Junto com a globalização do grande capital, ocorre a fragmentação do mundo do trabalho, a exclusão de grupos humanos, o abandono de continentes e regiões, a concentração da riqueza em certas empresas e países, a fragilização da maioria dos Estados, e assim por diante (...). O primeiro passo para que o Brasil possa enfrentar esta situação é parar de mistificá-la”.

BENJAMIM, Cesar & outros. **A Opção Brasileira**. Rio de Janeiro: Contraponto, 1998.

Diante do conteúdo dos textos apresentados acima, algumas questões podem ser levantadas.

1 - A que está relacionado o conjunto de fatores de “ordem tecnológica”?

2 - Considerando que globalização e opção política neoliberal caminharam lado a lado nos últimos tempos, o que defendem os críticos do neoliberalismo?

3 - O que seria necessário fazer para o Brasil enfrentar a situação da globalização no sentido de “parar de mistificá-la”?

A alternativa que responde corretamente às três questões, em ordem, é:

(A) revolução da informática / reforma do Estado moderno com nacionalização de indústrias de bens de consumo / assumir que está em curso um mercado de trabalho globalmente unificado.

(B) revolução nas telecomunicações / concentração de investimentos no setor público com eliminação gradativa de subsídios nos setores da indústria básica / implementar políticas de desenvolvimento a médio e longo prazos que estimulem a competitividade das atividades negociáveis no mercado global.

(C) revolução tecnocientífica / reforço de políticas sociais com presença do Estado em setores produtivos estratégicos / garantir níveis de bem-estar das pessoas considerando que uma parcela de atividades econômicas e de recursos é inegociável no mercado internacional.

(D) revolução da biotecnologia / fortalecimento da base produtiva com subsídios à pesquisa tecnocientífica nas transnacionais / considerar que o aumento das barreiras ao deslocamento de pessoas, o mundo do trabalho e a questão social estão circunscritos aos espaços regionais.

(E) Terceira Revolução Industrial / auxílio do FMI com impulso para atração de investimentos estrangeiros / compreender que o desempenho de empresas brasileiras que não operam no mercado internacional não é decisivo para definir o grau de utilização do potencial produtivo, o volume de produção a ser alcançado, o nível de emprego e a oferta de produtos essenciais.

Questão 5

“Crime contra Índio Pataxó comove o país

(...) Em mais um triste “Dia do Índio”, Galdino saiu à noite com outros indígenas para uma confraternização na Funai. Ao voltar, perdeu-se nas ruas de Brasília (...). Cansado, sentou-se num banco de parada de ônibus e adormeceu. Às 5 horas da manhã, Galdino acordou ardendo numa grande labareda de fogo. Um grupo “insuspeito” de cinco jovens de classe média alta, entre eles um menor de idade, (...) parou o veículo na avenida W/2 Sul e, enquanto um manteve-se ao volante, os outros quatro dirigiram-se até a avenida W/3 Sul, local onde se encontrava a vítima. Logo após jogar combustível, atearam fogo no corpo. Foram flagrados por outros jovens corajosos, ocupantes de veículos que passavam no local e prestaram socorro à vítima. Os criminosos foram presos e conduzidos à 1ª Delegacia de Polícia do DF onde confessaram o ato monstruoso. Aí, a estupefação: ‘os jovens queriam apenas se divertir’ e ‘pensavam tratar-se de um mendigo, não de um índio,’ o homem a quem incendiaram. Levado ainda consciente para o Hospital Regional da Asa Norte - HRAN, Galdino, com 95% do corpo com queimaduras de 3º grau, faleceu às 2 horas da madrugada de hoje”.

Conselho Indigenista Missionário - Cimi, Brasília-DF, 21 abr. 1997.

A notícia sobre o crime contra o índio Galdino leva a reflexões a respeito dos diferentes aspectos da formação dos jovens.

Com relação às questões éticas, pode-se afirmar que elas devem:

- (A) manifestar os ideais de diversas classes econômicas.
- (B) seguir as atividades permitidas aos grupos sociais.
- (C) fornecer soluções por meio de força e autoridade.
- (D) expressar os interesses particulares da juventude.
- (E) estabelecer os rumos norteadores de comportamento.

Questão 6

Muitos países enfrentam sérios problemas com seu elevado crescimento populacional.

Em alguns destes países, foi proposta (e por vezes colocada em efeito) a proibição de as famílias terem mais de um filho. Algumas vezes, no entanto, esta política teve conseqüências trágicas (por exemplo, em alguns países houve registros de famílias de camponeses abandonarem suas filhas recém-nascidas para terem uma outra chance de ter um filho do sexo masculino). Por essa razão, outras leis menos restritivas foram consideradas. Uma delas foi: as famílias teriam o direito a um segundo (e último) filho, caso o primeiro fosse do sexo feminino.

Suponha que esta última regra fosse seguida por todas as famílias de um certo país (isto é, sempre que o primeiro filho fosse do sexo feminino, fariam uma segunda e última tentativa para ter um menino). Suponha ainda que, em cada nascimento, sejam iguais as chances de nascer menino ou menina.

Examinando os registros de nascimento, após alguns anos de a política ter sido colocada em prática, seria esperado que:

- (A) o número de nascimentos de meninos fosse aproximadamente o dobro do de meninas.
- (B) em média, cada família tivesse 1,25 filhos.

- (C) aproximadamente 25% das famílias não tivessem filhos do sexo masculino.
 (D) aproximadamente 50% dos meninos fossem filhos únicos.
 (E) aproximadamente 50% das famílias tivessem um filho de cada sexo.

Questão 7

A leitura do poema de Carlos Drummond de Andrade traz à lembrança alguns quadros de Cândido Portinari.

Portinari

De um baú de folhas-de-flandres no caminho da roça
 um baú que os pintores desprezaram
 mas que anjos vêm cobrir de flores namoradeiras
 salta João Cândido trajado de arco-íris
 saltam garimpeiros, mártires da liberdade, São João da Cruz
 salta o galo escarlata bicando o pranto de Jeremias
 saltam cavalos-marinhos em fila azul e ritmada
 saltam orquídeas humanas, seringais, poetas de e sem óculos, transfigurados
 saltam caprichos do nordeste – nosso tempo
 (nele estamos crucificados e nossos olhos dão testemunho)
 salta uma angústia purificada na alegria do volume justo e da cor autêntica
 salta o mundo de Portinari que fica lá no fundo
 maginando novas surpresas.

ANDRADE, Carlos Drummond de. **Obra completa**. Rio de Janeiro: Companhia Editora Aguilar, 1964. p.380-381.
 Uma análise cuidadosa dos quadros selecionados permite que se identifique a alusão feita a eles em trechos do poema.



Podem ser relacionados ao poema de Drummond os seguintes quadros de Portinari:

- (A) I, II, III e IV
 (B) I, II, III e V
 (C) I, II, IV e V
 (D) I, III, IV e V
 (E) II, III, IV e V

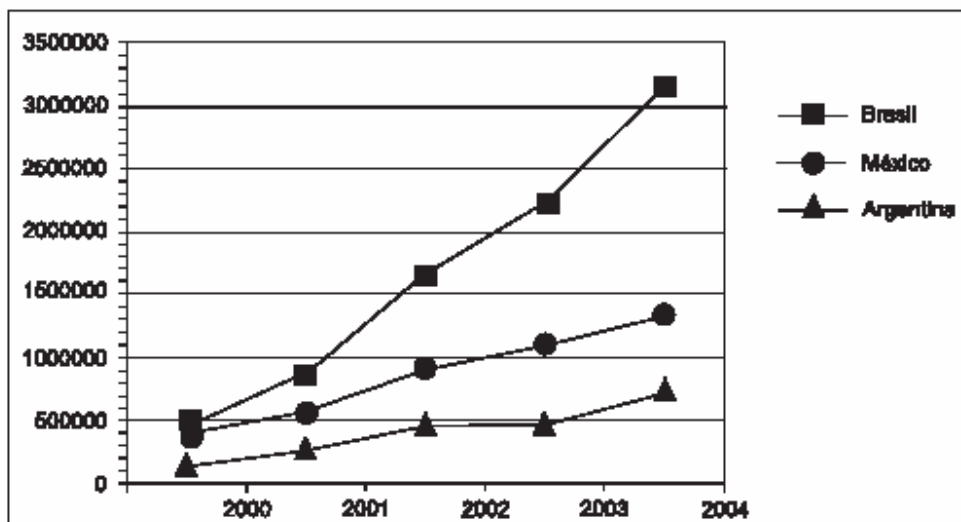
Questão 8

Os países em desenvolvimento fazem grandes esforços para promover a inclusão digital, ou seja, o acesso, por parte de seus cidadãos, às tecnologias da era da informação. Um dos indicadores empregados é o número de hosts, ou seja, número de computadores que estão conectados à Internet. A tabela e o gráfico abaixo mostram a evolução do número de hosts nos três países que lideram o setor na América Latina.

Número de hosts

	2002	2000	2001	2003	2004
Brasil	446444	876596	1644575	2237527	3163349
México	404873	559165	918288	1107795	1333406
Argentina	142470	270275	465359	495920	742358

Fonte: Internet Systems Consortium, 2004



Fonte: Internet Systems Consortium, 2004

Dos três países, os que apresentaram, respectivamente, o maior e o menor crescimento percentual no número de hosts no período 2000-2004 foram:

- (A) Brasil e México.
- (B) Brasil e Argentina.
- (C) Argentina e México.
- (D) Argentina e Brasil.
- (E) México e Argentina.

Anexo 2. Prova do ENADE – Componente de Formação Específica

Questão 9

Uma das diretrizes do SUS é a hierarquização das ações de saúde, com ênfase no papel das unidades básicas de saúde. Pode-se afirmar que a unidade básica de saúde deve

- (A) restringir-se às ações programáticas, como controle de hipertensão arterial e diabetes.
- (B) ofertar grande variedade de consultas especializadas e exames complementares.
- (C) ser a porta de entrada aos serviços de saúde, não se ocupando de ações curativas.
- (D) desenvolver ações que prescindam de exames complementares.
- (E) implementar as ações básicas de saúde de natureza preventiva e curativa.

Questão 10

Embora a expectativa de vida da população brasileira venha aumentando nas últimas décadas, preocupam as autoridades sanitárias os níveis elevados de mortalidade da população jovem, especialmente na faixa etária entre 15 e 29 anos, nos grandes e médios centros urbanos. As ações de maior impacto potencial para a diminuição da mortalidade da população adulta jovem brasileira devem estar centradas em

- (A) prevenção da AIDS.
- (B) prevenção das mortes violentas e por acidentes.
- (C) melhoria das condições sanitárias.
- (D) elevação da renda per capita da população.
- (E) combate à fome.

Questão 11

A estimativa da mortalidade infantil no Brasil para 2001 foi de 28,6 óbitos em menores de 1 ano para 1.000 nascidos vivos no mesmo período. Em 1986, estimou-se a mortalidade infantil no Brasil em 85,6 por 1.000 nascidos vivos. A principal contribuição para a queda desse coeficiente deve ser atribuída ao componente

- (A) neonatal precoce.
- (B) neonatal tardio.
- (C) pós-neonatal.
- (D) perinatal.
- (E) fetal.

Questão 12

Após um traumatismo grave, um paciente sem uso de suporte nutricional tem como maior fonte de fornecimento de calorias para a manutenção de sua homeostase a

- (A) proteólise.
- (B) aminoacidólise.
- (C) lipólise.
- (D) glicólise.
- (E) glicogenólise.

Questão 13

A Razão de Mortalidade Proporcional de 50 anos ou mais (ou Indicador de Swaroop-Uemura) mede a proporção de mortes a partir da idade mencionada em relação à totalidade de mortes. Em um determinado local onde o Indicador de Swaroop-Uemura é alto, pode-se afirmar que:

I - o coeficiente de mortalidade infantil desse local deve ser alto;

II - a desnutrição protéico-calórica nesse local deve ser pouco prevalente;

III - a expectativa de vida nesse local deve ser baixa;

IV - as afecções perinatais e as anomalias congênitas devem constituir as principais causas de morte entre menores de 1 ano de idade.

Está(ão) correta(s) somente a(s) afirmativa(s)

(A) II.

(B) I e III.

(C) I e IV.

(D) II e III.

(E) II e IV.

Questão 14

A produção elevada de proteínas de fase aguda é característica da síndrome da resposta inflamatória sistêmica que ocorre na resposta metabólica associada ao politrauma, grandes operações e em queimaduras envolvendo grande percentual da superfície corporal. A produção dessas proteínas é mediada pela elevação sérica concomitante de

(A) ACTH.

(B) hormônio do crescimento.

(C) catecolaminas.

(D) interleucina 6.

(E) proteína C reativa.

Questão 15

Considere uma paciente de 32 anos com amenorréia, tendo sido descartada gravidez, e que somente veio a menstruar com o uso de estrogênio e progestogênio. Com base nestes dados, a origem da amenorréia é

(A) hipotalâmica, hipofisária ou ovariana.

(B) hipofisária ou tubária.

(C) tubária, hipotalâmica ou ovariana.

(D) uterina ou ovariana.

(E) uterina, hipotalâmica ou hipofisária.

Questão 16

Considere as afirmativas abaixo, a respeito de diversas modalidades de câncer feminino.

I - O câncer de colo uterino deixou de ser responsável por grande parte da mortalidade na população feminina, em virtude da excelente cobertura do exame citopatológico em nosso país.

II - Com base na tentativa de aumentar o diagnóstico precoce e conseqüente redução da mortalidade por câncer de mama é que se tem justificado o rastreamento mamográfico anual ou, ao menos, bienal a partir dos 40 anos de idade.

III - A neoplasia intra-epitelial cervical de alto grau não tratada evolui para o câncer invasor em um percentual significativo de casos e, por este motivo, deve ser devidamente diagnosticada e tratada para reduzir a morbimortalidade do câncer de colo uterino.

IV - O diagnóstico em estágios iniciais faz do câncer de ovário o de menor taxa de mortalidade entre os tumores genitais.

Estão corretas somente as afirmativas

- (A) I e II.
- (B) I e III.
- (C) II e III.
- (D) II e IV.
- (E) III e IV.

Questão 17

Em relação ao parto pré-termo, considere as afirmativas a seguir.

I - Apesar de se ter agregado maior conhecimento sobre o assunto e de se dispor de recursos para a inibição, a sua incidência não tem diminuído nas últimas décadas.

II - Os agentes tocolíticos parecem proporcionar tempo importante para a ação de drogas que aceleram a maturidade pulmonar.

III - É responsável por 10% da morbidade e mortalidade perinatal.

IV - O feto pré-termo tolera melhor a hipóxia durante o trabalho de parto que o feto a termo.

Estão corretas somente as afirmativas

- (A) I e II.
- (B) I e III.
- (C) II e III.
- (D) II e IV.
- (E) III e IV.

Questão 18

Primigesta com 40 semanas está internada na maternidade em trabalho de parto e há 3 horas permanece com 7 cm de dilatação cervical. A altura uterina mede 40 cm e o feto encontra-se em apresentação cefálica, na posição occipto-ilíaca esquerda anterior e a altura da apresentação no plano menos 2 de De Lee. As contrações uterinas são de forte intensidade e na frequência de três em 10 minutos. Os batimentos cardíacos fetais mantêm-se na frequência de 140 por minuto antes, durante e depois das contrações. A ruptura espontânea das membranas ocorreu há 4 horas quando se constatou líquido amniótico claro com grumos. Diante destes achados, o que pode estar ocorrendo?

- (A) Discinesia uterina.
- (B) Desproporção cefalopélvica.
- (C) Fase ativa prolongada.
- (D) Fase latente prolongada.
- (E) Distócia de posição.

Questão 19

Mulher de 32 anos de idade apresenta anti-HIV (ELISA) positivo, confirmado por Western-Blot. É natural e procedente da cidade de São Paulo. Relata que o marido é soropositivo para HIV. Não apresenta queixas. Nega antecedentes de contato com portadores de doença respiratória. O exame clínico cuidadoso revela que a paciente está

em boas condições de saúde física e mental. Qual a conduta neste momento, para esta paciente?

- (A) Encaminhá-la para um Centro de Referência para DST/Aids a fim de que sejam realizados exames obrigatórios como contagem de células T CD4/CD8; carga viral do HIV; intradermorreação para tuberculose, sorologia para toxoplasmose (IgG) e citomegalovírus.
- (B) Encaminhá-la para um Centro de Referência para DST/Aids a fim de que sejam realizados exames obrigatórios: contagem de células T CD4/CD8; carga viral do HIV; sorologia para toxoplasmose e citomegalovírus (IgG e IgM) e 3 exames protoparasitológicos de fezes.
- (C) Orientar a paciente para retornar em 6 meses, reforçando a necessidade de prática de sexo seguro, uso obrigatório de “camisinha” e anticoncepcional de baixa dosagem por via oral.
- (D) Não é necessário referenciar esta paciente para um centro especializado, devendo-se solicitar exames obrigatórios: contagem de células T CD4/CD8; PCR qualitativo para o HIV; intradermorreação para tuberculose, sorologia para citomegalovírus (IgG e IgM) e Chagas.
- (E) Não é necessário referenciar a paciente para centro especializado, devendo-se marcar retorno em 6 meses e pedir os exames obrigatórios: de genotipagem do vírus HIV, carga viral, sorologia para hepatite C (anti – HCV), hepatite B (agHBs e antiHBc total) e para herpes simplex (IgG e IgM).

Questão 20

Mulher de 72 anos vem ao ambulatório de clínica médica geral de um hospital secundário com queixa de desânimo e fraqueza há 2 meses. Relata inapetência e perda de 8% do peso corpóreo no período. Ao exame clínico P=88 bpm PA=124x58 mmHg, mucosa descorada, hidratada. Ausculta cardíaca e pulmonar normais, abdome flácido, plano sem visceromegalias ou massas palpáveis, membros sem edema. Realizou hemograma Hb=7,6 g/dl, Ht=22% VCM=64 HCM=21 8200 leucócitos com diferencial normal, plaquetas=200.000. Levando em consideração a hipótese diagnóstica principal para a anemia da paciente, qual a conduta diagnóstica mais adequada neste momento?

- (A) Índice de segmentação de neutrófilos, dosagem sérica de vitamina B12 e ácido fólico, eletroforese de hemoglobina.
- (B) Colonoscopia, ferro sérico, capacidade total de ligação de ferro e ferritina.
- (C) Ferro sérico, ferritina, mielograma e biópsia de medula óssea.
- (D) VHS, proteína C reativa, ferritina e protoparasitológico de fezes.
- (E) DHL, reticulócitos, haptoglobina, urina I e Teste de Coombs.

Questão 21

Mulher de 48 anos de idade, moradora de uma capital brasileira, diarista, é encaminhada ao ambulatório de clínica médica a fim de que se investigue quadro de ganho de peso. Relata que nos últimos 6 meses ganhou aproximadamente 8 kg. Conta também que vem apresentando outros sintomas neste período, como preocupação excessiva com os filhos e marido, desânimo, angústia, perda de prazer em suas atividades, e despertar mais cedo que o habitual (4h da manhã), o que a está prejudicando. Nega qualquer outro problema prévio. Faz acompanhamento ginecológico de rotina, sem anormalidades. Os dados de saúde da família são insignificantes. Ao exame clínico observa-se paciente tensa, pouco à vontade, em vias de chorar. Pressão arterial medida no braço esquerdo na posição deitada = 120X84 mmHg; pulso = frequência cardíaca = 68 batimentos/min; ausculta cardíaca = bulhas rítmicas, sem outras alterações. Não há outros dados de interesse no exame clínico. Diante das possibilidades clínicas, para o estabelecimento do diagnóstico e alívio dos sintomas, a conduta a ser adotada é solicitar

- (A) atendimento nos serviços de saúde mental, pois os dados são suficientes para estabelecimento de diagnóstico.
- (B) medida de FSH e LH, Papanicolau e mamografia.
- (C) medida de FSH e LH no sangue e iniciar reposição de estrógenos conjugados.
- (D) medida de TSH no sangue e prescrever antidepressivos tricíclicos até o retorno.
- (E) medida de TSH e T4 livre no sangue e encaminhar para psicoterapia.

Questão 22

Homem de 58 anos, diabético e hipertenso há 6 anos em uso de captopril, hidroclorotiazida e insulina, procura ambulatório de clínica médica geral de um hospital terciário por ter apresentado dois episódios de perda da força em braço e perna direitos com desvio da rima bucal para a esquerda com duração de aproximadamente 40 minutos e reversão completa, no último mês (último há 5 dias). É tabagista de 25 maços/ano há 35 anos e nega etilismo. Pai hipertenso e mãe diabética, sem outros antecedentes mórbidos. Ao exame: eupnéico, corado, hidratado, anictérico, acianótico, consciente e orientado. Aparelhos respiratório e cardíaco normais. PA=132x88mmHg. FC=84bpm. Abdome normal. Exame neurológico completamente normal. A conduta para o caso é

- (A) solicitar ressonância magnética de encéfalo, ultrasonografia com doppler de carótidas, ecocardiograma e retorno com resultados.
- (B) encaminhar ao pronto-socorro para avaliação por neurologista e realização de tomografia computadorizada de crânio de urgência.
- (C) iniciar AAS 200 mg/dia, solicitar tomografia computadorizada de crânio e retorno com resultados.
- (D) iniciar AAS 200 mg/dia, solicitar ultra-sonografia com doppler de carótidas, ecocardiograma e retorno com resultados.
- (E) internar o paciente para realização de angiorressonância cerebral e iniciar heparina endovenosa.

Questão 23

No último exame de pré-natal que D. Januária realizou, foi identificado um AgHbs positivo. Como ela entrou em trabalho de parto antes do previsto, não teve oportunidade de conversar com seu obstetra. No berçário, ao examinar a criança, de posse dos resultados dos exames maternos, o pediatra deve fazer a seguinte orientação para a criança

- (A) ficar em observação rigorosa nas primeiras 48 horas de vida para identificar surgimento de icterícia e realizar as provas sorológicas para evidenciar a infecção pelo vírus B da hepatite.
- (B) suspender amamentação até receber o resultado da sorologia do recém-nascido.
- (C) administrar a vacina contra hepatite B e a imunoglobulina específica para hepatite B no recém-nascido nas primeiras 6 horas de vida e manter a amamentação.
- (D) administrar no recém-nascido a imunoglobulina específica para hepatite B nas primeiras 12 horas de vida e a vacina contra hepatite B até o final do primeiro mês de vida.
- (E) prescrever aciclovir para o recém-nascido ainda no berçário e suspender a amamentação.

Questão 24

Homem, 70 anos, tabagista crônico (130 maços/ano), está sob acompanhamento clínico por apresentar falta de ar progressiva há 5 anos. Atualmente a falta de ar ocorre aos pequenos esforços. Apresenta cianose de extremidades, lobos de orelhas e língua. A frequência respiratória em repouso é de 24 movimentos/min, a frequência cardíaca é de 96 batimentos/min, a pressão

arterial (semelhante às medidas anteriores) = 118X78 mmHg. Observa-se engurgitamento jugular e desaparecimento da amplitude de pulso radial na inspiração. Bulhas cardíacas rítmicas e hipofonéticas, sem sopros; fígado percutível a partir do 6º espaço intercostal direito e palpável a 3 cm do Rebordo Costal, edema de membros inferiores de intensidade moderada. O conjunto de condutas e orientações para o caso, considerando qualidade de vida e sobrevida, é

- (A) cessação do tabagismo, realização de espirometria e avaliação com vistas à indicação de oxigenoterapia domiciliar.
- (B) cessação do tabagismo, pois a espirometria não modificará a conduta clínica e a oxigenoterapia é precoce para o caso.
- (C) avaliação com vistas à indicação de oxigenoterapia domiciliar e cessação do tabagismo, pois a oxigenoterapia não trará impacto na sobrevida.
- (D) realização de espirometria e avaliação com vistas à indicação de oxigenoterapia domiciliar, pois a cessação do tabagismo não trará impacto.
- (E) oxigenoterapia domiciliar, pois a cessação do tabagismo não trará impacto e a espirometria não modificará a conduta clínica.

Questão 25

Um paciente de dois anos chega ao consultório com relato de tosse e febre alta há 48 horas. Qual(ais) dos achados abaixo ele deve apresentar para que se suspeite de pneumonia?

- (A) Tosse seca e contínua com lassidão.
- (B) História anterior de pneumonia.
- (C) Febre acima de 39 °C, que responde pouco aos antitérmicos.
- (D) Taquipnéia, mesmo se afebril, associada à tiragem.
- (E) Secreção nasal esverdeada com tosse produtiva.

Questão 26

Consultan unos padres porque su hijo, de cinco años, lleva varias noches despertando-se agitado como sí hubiera soñado algo que le angustía. Cuando acuden a su lado por la noche, el niño les mira y dice palabras que no tienen ningún significado. Al cabo de un rato vuelve a dormirse y por la mañana no recuerda nada de lo ocurrido. El diagnóstico sería

- (A) pesadillas.
- (B) terrores nocturnos.
- (C) foco epiléptico.
- (D) disomnia.
- (E) sonambulismo.

Questão 27

Um hábil e competente cirurgião do interior tem uma dúvida sobre qual operação realizar para beneficiar um paciente de 55 anos com adenocarcinoma de reto médio (distante 8 cm da borda anal, estágio pré-operatório: pT3 N0, M0). Basicamente, ele tem que decidir entre duas opções cirúrgicas e quando fazer, ou não fazer, radioterapia e quimioterapia. Ele tem cinco informações diferentes e deve escolher a melhor para seu paciente. Uma das sugestões é de um cirurgião com larga experiência e as outras quatro vêm de artigos atuais (2003-2004) que leu. Os quatro artigos são: uma metaanálise, um relato de casos tratados, um estudo retrospectivo e um ensaio clínico não controlado.

Fonte de Informação	Operação Sugerida como a melhor	Radioterapia e Quimioterapia Sugeridas		
		Pré-operatória	Pós-operatória	Não realizar
Conselho de cirurgião experiente	Amputação abdominoperineal		X	
Meta-análise	Ressecção anterior	X		
Estudo retrospectivo	Amputação abdominoperineal	X		
Relato dos casos	Ressecção anterior		X	X
Ensaio clínico não controlado	Ressecção anterior			

Diante disso, a opção que ele deve escolher é

- (A) amputação abdominoperineal com rádio e quimioterapia pré-operatória.
- (B) amputação abdominoperineal com rádio e quimioterapia pós-operatória.
- (C) ressecção anterior com rádio e quimioterapia pré-operatória.
- (D) ressecção anterior com rádio e quimioterapia pós-operatória.
- (E) ressecção anterior sem rádio ou quimioterapia pré-operatória.

Questão 28

A 68-year-old man comes with a chief complaint urinary frequency, hesitancy, and a slow stream for 18 months. Rectal examination shows a firm, slightly enlarged prostate. After he voids, a Foley catheter is inserted and yields 600 mL of urine. Urinalysis is within normal limits. Which of the following is the most likely diagnosis?

- (A) Acute prostatitis.
- (B) Benign prostatic hypertrophy.
- (C) Neurogenic bladder.
- (D) Bladder cancer.
- (E) Urethral stricture.