

EXAME NACIONAL DE DESENVOLVIMENTO DE ESTUDANTES DE ESTATÍSTICA - DESAFIOS E PERSPECTIVAS PELA TRI

Edy Célia Coelho¹, Paulo Justiniano Ribeiro Junior²,
Wagner Hugo Bonat³

Resumo: *Este estudo analisa o Exame Nacional de Desempenho do Estudantes (Enade) na área de Estatística em 2009, provas objetivas, pela Teoria de Resposta ao Item (TRI). O Enade não aplica a TRI, mas a perspectiva dessa abordagem é de fornecer uma visão geral da utilização através dos modelos de teoria de resposta ao item unidimensional e multidimensional. Com o objetivo de proporcionar diagnósticos e subsídios para a implementação ou manutenção de novos procedimentos metodológicos educacionais na elaboração dos itens que compõem parte do processo do Enade. O conjunto de dados abordado é composto por 436 concluintes da graduação de Estatística. A prova é composta por 32 itens, dentre dos quais 8 são de formação geral e 24 de formação específica. Os estudos realizados mostraram que cada modelo proposto para avaliar a capacidade dos alunos tem sua particularidade. Nos resultados, foram indicados modelos multidimensionais, com dois traços latentes e, o modelo unidimensional, de três parâmetros. Através dessas análises, objetiva-se mostrar a importância dos modelos da TRI em avaliação, capaz de considerar que uma prova seja composta por múltiplas ou uma habilidade dominante e, consequentemente, respondida por elas.*

Palavras-chave: *Enade, Habilidades, Teoria de resposta ao item uni e multidimensional.*

Abstract: *The study examines objective evidence of the National Examination Performance of Students (Enade) in Estatística/2009, by Item Response Theory (IRT). The Enade not apply TRI. The data have 436 graduates and 32 items, 8 specifies and 24 general training. The analyzes indicate uni and multidimensional models.*

Keywords: *Enade, Abilities, Uni and multidimensional item response theory.*

1 Introdução

O estudo insere-se em uma área interdisciplinar composta pela Estatística e pela Educação, tratando da Teoria de Resposta ao Item e aplicando-a no Exame Nacional de Desempenho de Estudantes (Enade).

O Enade é realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), autarquia vinculada ao Ministério da Educação (MEC), é componente curricular obrigatório aos cursos de graduação, conforme determina a Lei nº 10.861/2004. Com o objetivo

¹PPGMNE-UFPR/SEED-PR. e-mail: edyceliacoelho@gmail.com

²LEG - PPGMNE/UFPR. e-mail: paulojus@leg.ufpr.br

³LEG - UFPR. e-mail: wbonat@gmail.com.

de aferir e acompanhar o desempenho acadêmico dos estudantes em relação aos conteúdos programáticos, suas competências e habilidades necessárias ao aprofundamento da formação geral e profissional.

Para tentar chegar mais próximo de uma conclusão real acerca do verdadeiro conhecimento de um indivíduo avaliado, foi desenvolvida, Teoria Clássica dos Testes (TCT), a Teoria da Resposta ao Item (TRI), que usa modelos estatísticos construídos para representar a probabilidade de um indivíduo responder corretamente a um item em determinado teste.

O problema que motiva este estudo do Enade é decorrente da aplicação de uma prova multidisciplinar, onde se supõe a priori que existem diversas habilidades/latentes sendo avaliadas - multidimensionalidade. A aplicação de modelos da TRI em provas multidisciplinares é ainda pouco explorada na análise de avaliações. Ao buscar fundamentos na literatura para aplicar a TRI no Enade, observa-se que o material disponível ainda é muito recente em se tratando de avaliação educacional. Cita-se alguns autores que abordaram a TRI na área de avaliação relatando as potencialidades da teoria na validade de testes, tais como: Nojosa (2001); Francisco (2005); Vendramini et al. (2005); Andriola (2008); Quaresma et al. (2012); Piton (2012).

O Enade, onde a grande escala de respondentes se faz presente por ser uma avaliação externa, e hoje é visto pelas instituições como ranqueamento tanto interno como externo de desempenho, cuja a nota é composta de forma a levar em consideração aspectos, como o corpo docente e a infra-estrutura, que não será abordado neste estudo, mas o questionamento é: será que esta avaliação considerando somente as questões objetivas compostas por itens gerais e específicos, estão contribuindo para a qualidade do ensino? Objetiva-se abordar a TRI, no sentido de possuir mais uma ferramenta para verificar a concordância ou não entre os métodos para análise dos itens, além de destacar os pontos em que a TRI tem mais a contribuir do que a TCT para o processo avaliativo.

O modelo clássico supõe que todas as questões estão avaliando a dimensão latente, conhecimento geral do graduado, porém, como ficam as habilidades específicas? Não são avaliadas? Será que esta suposição é razoável? Como testá-la? São perguntas coerentes com a realidade e que o modelo da Teoria de Resposta ao Item Multidimensional (TRIM) é capaz de responder. Os questionamentos são pertinentes e é possível indicar acerca do verdadeiro conhecimento de um indivíduo avaliado se a base está nas questões que compõem a avaliação. Assim, investiga-se métodos estatísticos dos modelos da TRI para contribuição neste processo de avaliação, almejando uma análise mais condizente com a realidade.

Conforme Reckase (2009), o modelo clássico de TRI impõe a restrição de que existe apenas uma habilidade latente sendo mensurada, e que essa é a principal responsável pela resposta do candidato a todas as questões. Essa suposição pode ser adequada em diversas situações, porém não existe um teste efetivo para o que é chamado de unidimensionalidade do modelo.

Espera-se que a análise da dimensionalidade da TRI possibilite uma reflexão metodológica na elaboração da avaliação, fornecendo subsídios para direcionar o esforço empreendido no processo de ensino e aprendizagem, de forma a contemplar a melhor abordagem pedagógica e o mais pertinente método didático adequado a cada disciplina. Segundo, a maioria das avaliações atuais leva em consideração não apenas uma habilidade latente, mas várias. Em termos de provas multidisciplinares, é razoável pensar que múltiplas habilidades estejam sendo avaliadas, portanto, utilizadas pelos respondentes em cada uma das questões. Porém, a metodologia clássica de TRI assume que uma habilidade geral é predominante, mas é necessário abordar, verificar os itens que compõem a prova e que habilidades a prova exige do respondente.

A intenção desse trabalho é analisar os modelos uni e multidimensionais da TRI na prova objetiva do Enade/2009 na área de Estatística, em que uma avaliação seja composta e, consequentemente, respondida por múltiplas habilidades latentes como também por um fator dominante. Como o modelo deve ser realista, pelas características da prova e dos avaliados procurou-se ajustar os modelos da TRI uni/multidimensional, verificando os fatores dominantes e através dos resultados das estatísticas diagnósticas da qualidade. Respeitando a quantidade de parâmetros estimados e os padrões de resposta dos itens.

Com a aplicação da TRI nas provas objetivas do curso avaliado do Enade/2009, busca-se um melhor entendimento no exame, levando à compreensão de como o conhecimento foi ou está sendo avaliado. Automaticamente, verifica-se o desempenho do graduado, além de se fornecer diagnósticos e subsídios para a implementação ou manutenção de novos procedimentos metodológicos educacionais na elaboração dos itens que compõem parte do processo do Enade.

O presente trabalho encontra-se dividido em cinco seções: esta primeira busca introduzir o problema de análise motivando o uso dos métodos da TRI; a segunda aborda os modelos da TRI utilizados nesta pesquisa; a terceira apresenta os principais resultados, a discussão; a quarta com conclusões; e a última, a bibliografia utilizada para os estudos.

2 Material e métodos

Visando as avaliações externas, disponibilizadas pelo INEP, abordou-se, neste trabalho, um conjunto de dados resultante do Enade/2009, devido à gama de possibilidades existentes de análises com diferentes áreas de conhecimento e por serem avaliações em largas escalas, em que os dados são tratados como microdados. Optou-se pela área de conhecimento de Estatística, já que a abordagem deste estudo se desenvolve nesta área. Ressalta-se que esse exame acontece de três em três anos. Os dados de 2012, até o presente momento, não estão disponíveis no portal do INEP. Para mais detalhes, vale consultar o Manual do Enade (2004 a 2013).

As provas são elaboradas e aplicadas por instituição, ou consórcio de instituições, contratada por processo licitatório que, comprove capacidade técnica em elaboração de prova, preparo de instrumentos, aplicação e avaliação, e que atenda aos requisitos estabelecidos no Projeto Básico, conforme detalhes no *site* do INEP. As provas possuem questões de avaliação da formação geral, comuns aos cursos de todas as áreas, e questões específicas para cada curso. Nas duas partes, as questões foram discursivas e de múltipla escolha, envolvendo situações-problema e estudos de casos. Contudo, será abordado apenas as questões de resposta objetiva, pois as mesmas contemplam diferentes momentos da vida acadêmica do estudante, dividindo o nível de complexidade entre baixo, médio e alto. Os resultados do Enade são divulgados por meio da página na internet, conforme as referências Brasil/INEP-MEC, sendo de responsabilidade dos dirigentes e coordenadores de cursos de graduação divulgar amplamente, junto ao seu corpo discente e a todo o centro acadêmico, os resultados do exame.

Conforme o Relatório de Síntese Estatística do Enade (2009), a prova foi aplicada em 25 cursos de Estatística, sendo que 22 deles, o equivalente a 88,0% desse total, eram vinculados a universidades. As faculdades participaram com 2 cursos (8,0%), e os centros universitários, com 1 curso (4,0%). Não houve participação de instituições municipais de ensino em nenhuma região brasileira.

No Exame de 2009, em todo o Brasil, a população era de 1557 estudantes, dos quais participaram da avaliação 1153, sendo 710 ingressantes que se encontravam no final do primeiro ano, e 443 concluintes. Conforme o Relatório de Síntese do mesmo ano, a maioria dos estudantes era oriunda de instituições federais (73,3%). A avaliação teve duração total de 4 (quarto) horas e foi composta de 40 questões no total: 10 de formação geral, e 30 de formação específica da área. As duas partes têm questões discursivas e de múltipla escolha.

O conjunto de dado do Enade analisado é os respondentes presentes concluintes que entregaram o cartão resposta totalizando 443 sendo que sete dentre esses entregaram em branco, os que rasuraram alguma questão ou deixaram o cartão ou a questão em branco, utilizou a sigla (N.a) nenhuma alternativa. O conjunto em análise, por esses motivos, possui 436 respondentes e 32 questões considerando a habilidade geral.

Baseado no relatório do Enade (2009), para o estudo das análises da TRI, das 40 questões, considerou-se somente a prova objetiva no total: 8 questões da formação geral (depois das análises da comissão do Enade, decidiu-se por anular a questão 1 pelo critério do coeficiente de correlação ponto-bisserial); 27 de formação específica, sendo que 3 questões foram anuladas (14, 22, 35) pela comissão assim que divulgaram o gabarito. Na formação específica, o item 23

também foi excluída depois do relatório geral pelo critério do coeficiente de correlação ponto-biserial. Portanto, nesse estudo serão analisadas as 32 questões objetivas do Enade e as 30 questões objetivas excluindo os itens Q1 e G23, verificando os motivos dos dois itens terem sido cancelados pelo coeficiente de correlação.

A parte de avaliação de Formação Geral do Enade/2009 foi composta de questões com abordagens de estudos de caso, interpretação de textos, análise de gráficos e charges, simulações, dentre outros, além de verificar o domínio de conhecimentos e de diferentes níveis de habilidades e competências, a compreensão de temas que poderiam transcender ao seu ambiente próprio de formação e fossem importantes para a realidade contemporânea, bem como essenciais na formação de todos os estudantes de qualquer área da Educação Superior. Para o componente específico da área de Estatística, foi elaborada a partir das Diretrizes Curriculares Nacionais do Curso, aprovada e instituída pelo Conselho Nacional de Educação (CNE) do Ministério da Educação (MEC).

A finalidade da avaliação externa do Enade é tencionar, promover o aprofundamento dos compromissos e das responsabilidades sociais do cidadão em conjunto com as instituições de educação superior de formação, na busca pela valorização de sua missão pública, da promoção dos valores democráticos, do respeito à diferença e à diversidade, da afirmação da autonomia e da identidade institucional. A fim de melhorar a compreensão das análises desenvolvidas neste estudo, será abordado, brevemente, os modelos da TRI na próxima seção.

2.1 Modelos da Teoria da Resposta ao Item

O principal objetivo desta subseção é apresentar, resumidamente, os métodos estatísticos que serão utilizados. Dessa forma, as reflexões sólidas alicerçadas em conhecimento já existentes serão citadas no decorrer do texto referenciando-se alguns autores, para concretizar novos propósitos que tragam aprofundamento, nos estudos, tanto na área de Educação como na Estatística.

Os modelos com relação à quantidade de traços latentes ou habilidades podem ser classificados em: modelos unidimensionais, quando estes consideram uma habilidade, e multidimensionais, quando se examinam duas ou mais habilidades em um mesmo instrumento. Os modelos unidimensionais são os mais utilizados na prática, e toda a teoria matemática da TRI foi fundamentada com este enfoque, porém, pode-se encontrar, por exemplo, em Van Linden e Hambleton (1996) referências sobre os modelos multidimensionais e aplicações recentes em artigos de Chalmes R. Phili e Yanyan Sheng et al, (2012), onde o modelo unidimensional faz parte do modelo multidimensional, bem como na grande maioria das bibliográficas, estuda-se a unidimensionalidade, e subsequente, a multidimensionalidade, inclusive de Reckase (2009).

O método para tratar o Enade pela TRI será baseada nos modelos de k -dimensionais. As respostas são dicotômicas e a dimensionalidade será feita através da análise de componentes principais sobre a matriz de correlação tetracórica, análise paralela e, principalmente, através da abordagem da análise fatorial de informação completa, que se baseia em ajustes de modelos da TRI multidimensionais (TRIM). Nessa última abordagem, estuda-se a capacidade do instrumento de produzir novas dimensões interpretáveis, possibilitando conduzir e analisar algumas relações estruturais nos itens da prova multidisciplinar.

Mais detalhes sobre os modelos matemáticos utilizados na TRI podem ser encontrados em Baker (2001), Baker e Kim (2004), Boomsma, Embretson e Reise (2000), Lord (1980), Reckase (1985) e (2009) e Van Der Linden e Hambleton (1997).

O modelo TRIM, conforme Reckase (2009), são casos especiais da TRI, construídos sobre a premissa de que a função matemática inclui como parâmetros um vetor de características de habilidades e os conhecimentos que a pessoa traz para um teste, além de um vetor de características do item que descreve a dificuldade e a sensibilidade para diferenciar as características dos respondentes. A representação básica de um modelo é resumida pela teoria de resposta ao item multidimensional, que é uma consequência tanto da análise fatorial quanto da teoria de Resposta ao Item Unidimensional (TRI).

Portanto o procedimento é identificar quando e quanto vale a pena para um teste multidisciplinar, utilizar os modelos da TRI e TRIM. Tal como acontece com todos os modelos científicos de fenômenos observados, os modelos são úteis apenas na medida em que proporcionam aproximações razoáveis para relações do mundo real. Além disso, o uso de modelos mais complexos só se justifica quando eles proporcionam uma maior precisão ou novas percepções. Deste modo, iniciaremos descrevendo o modelo de k -dimensão permeando a TRI e a TRIM.

Inicialmente, para descrever o modelo da TRI, considere:

- Y : um vetor aleatório;
- i : item, $i = 1, 2, 3, \dots, I$;
- j : indivíduo, $j = 1, 2, 3, \dots, J$;
- Y_{ij} : variável aleatória correspondente à resposta de um indivíduo ao item em um teste qualquer;
- y_{ij} : valor real da resposta dada a probabilidade condicionada;
- Y_j : $(Y_{1j}, Y_{2j}, \dots, Y_{Ij})$ vetor total de respostas do respondente submetido ao um teste e composto por itens;
- θ : vetor do traço latente para TRI;
- k : dimensionalidade do instrumento de medida;
- θ_j : um vetor k -dimensional dos traços latentes dos indivíduos; logo, $\theta_j = (\theta_{1j}, \theta_{2j}, \dots, \theta_{Kj})$ para TRIM;
- η_i : vetor de parâmetros correspondentes às características do item;
- α_i : é o parâmetro relacionado à discriminação do item-TRI;
- α'_i : é o vetor de parâmetros relacionados à discriminação do item i , $\alpha'_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ik})$ -TRIM;
- β_i : é o parâmetro que representa a dificuldade do item;
- c_i : é o parâmetro que representa a probabilidade de um indivíduo com baixa habilidade responder corretamente ao item i ;
- d_i : é o parâmetro relacionado à dificuldade do item, medido na mesma escala da habilidade, ou seja, é o escalar que representa a dificuldade de cada item $d_i = \alpha_i \beta_i$
- D : é um fator escala, introduzido para tornar a função logística mais próxima possível da função da ogiva normal ao utilizar $D = 1,702$, caso contrário, utiliza a constante $D = 1$.

Portanto, sendo o modelo da TRI, considere:

$$P(Y_{ij} = y_{ij} | \theta_j, \eta_i) = f(\theta_j, \eta_i, y_{ij})$$

Assumindo a independência condicional sob a forma geral do modelo da TRI, temos:

$$P(Y_{ij} = y_{ij} | \theta_j, \eta_i) = \prod_{i=1}^I P(Y_{ij} = y_{ij} | \theta_j, \eta_i)$$

Considerar ϕ a função de distribuição acumulada da normal padrão e que o modelo completo descrito na forma hierárquica é:

$$Y_{ij} | \theta_j \sim \text{Bernoulli}(P_{ij})$$

considerando a dimensão onde $\theta_{kj} \sim N(0, 1)$, temos para $k = 1$ o traço latentes conforme: $\theta_{1j} \sim N(0, 1)$,

Assumindo a forma logística com um traço latente/unidimensional é:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-D(\alpha_{i1}\theta_{1j} + d_i)}} \quad (1)$$

Considerando a dimensão onde $\theta_{kj} \sim N(0, 1)$, para $k = 2$ os traços latentes são independentes entre si, conforme: $\theta_{1j} \sim N(0, 1)$, $\theta_{2j} \sim N(0, 1)$, logo, o modelo logístico com dois traços latente/multidimensional é:

$$P_{ij} = P(Y_{ij}|\theta_j, \alpha_i, d_i, c_i) = c_i + (1 - c_i) \cdot \frac{1}{1 + e^{-D(\alpha_{i1}\theta_{1j} + \alpha_{i2}\theta_{2j} + d_i)}} \quad (2)$$

Portanto:

$$P_{ij} = P(Y_{ij}|\theta_j, \eta_i) = \phi(\theta_j' \alpha_i + d_i) \quad (3)$$

Com o modelo descrito, aborda-se verificando inicialmente os resultados da equação (1), em seguida a (2) e assim sucessivamente quando há indicação para outras dimensões/habilidades pela análise fatorial. Os modelos descritos a partir do Modelo Logístico de três parâmetros (ML3), o parâmetro de acerto casual (c_i) refere-se à probabilidade de um respondente com baixa habilidade acertar o item (ao acaso). Quando não se necessita desse parâmetro na análise, ele pode ser retirado e tem-se o Modelo Logístico de dois parâmetros (ML2). Ainda, se retirarmos o parâmetro de discriminação (α_i), tem-se o Modelo Logístico de um parâmetro (ML1), chamado de Modelo de Rasch para a unidimensionalidade, e para a multidimensionalidade, o valor é considerado constante e aponta que todos os itens discriminam da mesma forma, podendo-se, dentre esses modelos, identificar casos particulares seguindo a hierarquia com modelos encaixados, conforme as quantidades de parâmetros propostos, isto é, possível fixar o acerto ao acaso (ML2f), e essa análise compõem os resultados deste estudo.

O objetivo desse estudo é estimar os modelos conforme a dimensionalidade indicada e compará-los a fim de encontrar o que melhor descreve a prova em análise. A comparação entre os modelos, pode tanto ser feita pelo teste de razão de verossimilhanças, quanto pelo Critério de Akaike (AIC) ou pelo Critério Baysiano (BIC). os quais foram contemplados nas análises seguintes.

Outro aspecto importante da amostra e quantidade de parâmetro estimado, ou seja, o grau de liberdade (df). Conforme Chalmers (2013) e Reckse (2009), o número de padrões de resposta original, que é, o número de padrões de resposta únicos sem repetições é um dos fatores da escolha do modelo ao comparar com o df . de modo a evitar graus de liberdade totais negativos para o modelo. Exemplificando: não deve-se estimar muito parâmetros em pequenos conjuntos de dados, principalmente quando os dados não variam muito os padrões de resposta no teste, pois o conjunto de dados possuir vetores de resposta repetidas e diminui a efetividade do df total. Além do mais, é inviável ter mais parâmetros para estimar do que o df total.

Contudo, o número de parâmetros calculados é geralmente apenas uma contagem de matriz de parâmetros livres. Por exemplo, para 32 itens, deve haver em torno de 128 parâmetros estimados, ou seja, $n + (n \cdot m)$, onde m representa os 3 parâmetros da TRI. Um dos motivos de abordar a df dessa forma é que se houver mais parâmetros para estimar do que graus de liberdade, haverá na estatística modelos indefinidos. Lembrando que quanto mais parâmetros para estimar, maior serão as dificuldades computacionais. Logo, deve valer a pena a escolha de um modelo, para compensar a troca de um modelo mais simples para um mais complexo, com mais df , verifica-se e testa-se a razão do qui-quadrado.

Para fazer inferência sobre os parâmetros deste modelo, é necessário a obtenção da verossimilhança marginal, adquirida após a integração dos efeitos aleatórios, neste caso, as habilidades latentes θ_j . O integrando desta verossimilhança marginal é o produto de uma binomial por

uma gaussiana padrão e não tem solução analítica. Desta forma, é necessário usar métodos para integração numérica. Detalhes da inferência sobre os parâmetros deste modelo podem ser consultados em Bonat et al (2012). Mais detalhes sobre os fundamentos da TRI e os modelos matemáticos utilizados podem ser encontrados em Van der Linden e Hambleton (1997), Baker (2001) e Reckase (2009).

O uso dos métodos computacionais em cálculo numérico cresce cada vez mais em intensivo métodos e técnicas na resolução de problemas, para os quais as soluções manuais são impraticáveis, impossíveis ou imprecisas. Conforme a complexidade do problema proposto, torna-se impossível o avanço pela via analítica, e surge a necessidade de diversas ferramentas computacionais como *softwares* a fim de automatizar os cálculos. Portanto, para a estimação dos parâmetros dos itens, utilizou-se o *software R*, através dos pacotes: *ltm* de Rizoupoulos (2006) versão 0.9-9, atualizado em setembro de 2012 - modelos de variáveis latentes para dados dicotômicos, para análise exploratória; *psych* de Revelle (2013) versão 1.3.10.12, de outubro de 2013 - análise fatorial das correlações tetracóricas e policóricas; *mirt* de Chalmers (2012) versão 1.1 dezembro de 2013 - análise dos modelos uni/multidimensionais para os modelos exploratórios e de confirmação usando métodos de máxima verossimilhança.

As informações apresentadas aqui apenas abrangerão as tendências gerais, pois o pacote *mirt* ainda está sendo executado por Chalmers, devido as datas recentes de atualizações das versões. A questão aqui é não tornarmos obsoletos apenas em um *software*, visto que ainda não estão previamente estabelecidos qual estimador foi e é utilizado na obtenção do melhores resultados, mas, sim, fazer das comparações ferramentas benéficas para o usuário.

3 Resultados e discussões

O conjunto em análise, possui 436 respondentes e 32 questões no total. Segue na TABELA 1, a primeira análise descritiva, pelo pacote *ltm*, apresenta os resultados de oito questões de formação geral (G) e de 21 de formação específica (Q), da proporção de acertos para cada uma das questões, ponto bisserial e *Cronbach's alpha*.

A Tabela 1, também contempla o Relatório Síntese do Enade (2009) e detalhes podem ser consultados, em que as questões foram classificadas segundo índice de facilidade (IF) e índice de discriminação (ID), para todas as áreas e inclusive para o curso em estudo.

No índice de facilidade duas das oito questões de formação geral foram difíceis, 1 foi muito difícil, 3 foram médias e, por fim, apenas 1 fácil e 1 muito fácil. Utilizando-se a mesma classificação para o índice de discriminação, o ponto bisserial informa que ao incluir as questões, todas apresentaram uma boa distribuição, com ressalva à G1, que apresentou um índice de discriminação fraco.

Com os índices obtidos da proporção de acertos, as questões objetivas da prova foram assim avaliadas para a área de Estatística: das oito questões, somente a G1 teve o percentual abaixo de 15%, ou seja, somente uma questão foi muito difícil na avaliação. As questões G7 e G8 foram fácil e muito fácil. Quanto aos acertos na formação específica, dentre as 24 questões, 17 apresentaram valores em torno de $0,40 > IF \geq 0,15$, e somente 7 entre $0,60 > IF \geq 0,40$. Logo, na formação geral, a prova foi composta por questões difíceis e médias.

O coeficiente *Cronbach's alpha* (AC) na formação geral com o valor de 0,484 para as oito questões indica uma média consistente, mas ao retirar as duas questões em negrito, a G1 e a G4, a consistência interna entre as questões aumenta. Observando o AC na formação específica, as 21 questões possuem uma consistência interna de 0,741 e, excluindo-se a G23, aumentaria a confiabilidade entre as questões de 0,744. Detalhes no Relatório Síntese de 2009, no INEP.

Recomenda-se que sejam verificadas as questões G1, G4 e Q23 da prova do Enade/2009 do curso de Estatística e analisar pedagógica e metodologicamente, juntamente com profissionais da área, os motivos que levaram a estatística descritiva a apresentar esses valores. Vale ressaltar que a proporção de acerto não é baixa como a G1 na análise feita na Tabela 1. Para esgotar as análises descritivas, mas não apresenta-se neste resumo os resultados, as 32 questões, não

Tabela 1: Índices de facilidade conforme a proporção de acertos, índices de discriminação de acordo com o ponto bisserial e o Cronbach's de alpha para as questões da formação geral e específica

436 respondentes						
Questões	Proporção % acertos	Ponto Bisserial		Cronbach's alpha *Todas as questões e Excluindo a questão	Índice de Facilidade	Índice de Discriminação
		incluir	excluir			
Formação Geral				0,484*		
G1	10,30%	0,24	0,04	0,497	Muito difícil	Fraco
G2	58,30%	0,43	0,12	0,476	Médio	Bom
G3	54,50%	0,51	0,22	0,435	Médio	Muito Bom
G4	34,90%	0,33	0,03	0,521	Difícil	Médio
G5	39,90%	0,49	0,2	0,456	Difícil	Bom
G6	53,50%	0,51	0,31	0,413	Médio	Bom
G7	84,90%	0,51	0,31	0,394	Fácil	Muito Bom
G8	87,90%	0,49	0,31	0,398	Muito fácil	Muito Bom
Formação Específica				0,741*		
Q11	51,80%	0,36	0,25	0,736	Médio	Bom
Q12	56,40%	0,29	0,17	0,74	Médio	Médio
Q13	17,30%	0,44	0,37	0,73	Difícil	Muito bom
Q15	38,50%	0,36	0,25	0,734	Difícil	Bom
Q16	36,90%	0,43	0,33	0,73	Difícil	Muito bom
Q17	27,00%	0,38	0,28	0,733	Difícil	Bom
Q18	15,50%	0,32	0,24	0,735	Difícil	Bom
Q19	27,20%	0,43	0,34	0,729	Difícil	Muito bom
Q20	16,20%	0,28	0,19	0,739	Difícil	Médio
Q21	29,20%	0,45	0,35	0,729	Difícil	Muito Bom
Q23	26,80%	0,18	0,08	0,744	Difícil	Fraco
Q24	30,60%	0,29	0,18	0,739	Difícil	Bom
Q25	18,50%	0,41	0,32	0,731	Difícil	Muito bom
Q26	23,20%	0,45	0,37	0,727	Difícil	Muito bom
Q27	48,20%	0,49	0,39	0,725	Médio	Muito bom
Q28	31,80%	0,38	0,28	0,731	Difícil	Bom
Q29	30,20%	0,28	0,17	0,74	Difícil	Médio
Q30	49,60%	0,3	0,19	0,738	Médio	Bom
Q31	41,50%	0,47	0,37	0,727	Médio	Muito bom
Q32	51,50%	0,27	0,16	0,739	Médio	Médio
Q33	32,40%	0,5	0,41	0,724	Difícil	Muito bom
Q34	19,10%	0,43	0,34	0,73	Difícil	Muito bom
Q36	20,80%	0,42	0,34	0,732	Difícil	Muito bom
Q37	44,40%	0,46	0,35	0,726	Médio	Muito bom

FONTE: O autor (2013)

NOTA: Questão (Q). No Cronbach's alpha, para verificar a confiabilidade dos itens, a notação (*) significa que todas as questões são consideradas, excluindo as que não apresentam asterisco. As questões em negrito são as questões que apresentaram valores inadequados para a proporção de acerto, ponto bisserial e Cronbach's alpha.

fazendo-se a divisão das formações gerais e específicas, confirmando a indicação da G1 e a Q23 para exclusão pelo AC.

Para verificar o desempenho de acertos nas questões dos alunos, em relação à probabilidade de acerto ao acaso a Figura 1 mostra a frequência de acertos das questões. Assim, é possível verificar qual seria, o comportamento dos resultados observados no exame em comparação com o esperado, sob a hipótese de que os alunos estão respondendo aleatoriamente a todos os 32 itens, considerando a habilidade geral em responder toda a prova.

Na Figura 1, percebe-se que nenhum aluno zerou a prova e nem gabaritou as questões. Por esse motivo, não temos inflações em zero. Mesmo considerando-se que os NAs fossem substituídos por zero, não foram observados valores com irregularidade.

Ainda, na Figura 1, observou-se uma distribuição assimétrica positiva para acerto das questões, indicando que os alunos não tiveram valores observados acima do esperado, ou seja, mesmo havendo um boa distribuição de acertos observados, o esperado possuía valores acima do acerto casual, o desempenho dos respondentes foi maior quando os itens foram acima de 10 acertos. Os outros itens mantiveram padrão de acerto maior do que os observados. Logo, para os concluintes de Estatística, sobre a frequência de acertos observados nos resultados da prova

objetiva - até nove acertos - foi abaixo do que o número esperado de acerto, sob a hipótese de resposta aleatória.

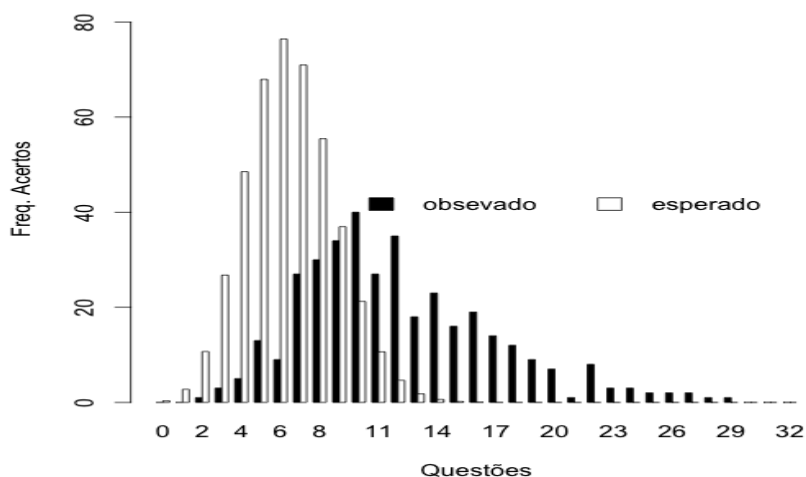


Figura 1: Frequência de acertos do valor observado e o número esperado de acerto casual dos 436 respondentes do ENADE/2009

Como a intenção é analisar a prova e verificar quantas habilidades o aluno utilizou para responder cada item, detalha-se a prova geral, ou seja, sem dividir a prova em formação geral ou específica para verificar se a TRI relata e capta essa observação no modelo da TRI.

Para desenvolver estudos e avançar nas perspectivas de avaliar os resultados no Enade, segue uma visão geral das análises dos modelos exploratórios da TRI. Primeiramente, identifica-se a dimensionalidade deste exame pelo pacote *psych*. Em segundo lugar, analisa-se os resultados das estatísticas de diagnósticos da qualidade (logLik, AIC e BIC). Então, aplica-se o pacote *mirt* para validar, comparar os modelos e, assim, verificar os ajustes conforme os critérios respeitando os df , indicar os melhores modelos e, conforme a importância dos assuntos das análises da TRI, a necessidade de subdividi-los em subseções, dando continuidades e ordenando os estudos. Em terceiro lugar, explora-se e verifica-se, conforme os modelos escolhidos, utilizando o pacote *mirt*, a técnica das cargas fatoriais e quais itens se relacionam a cada fator com a rotação oblíqua, e a partir desses passos indicar qual o modelo que melhor representa o exame pela TRIM. Para o quarto caso, o modelo indicado não é unidimensional, é possível, ainda, verificar o modelo confirmatório para confirmar a dimensão dos itens, pelo pacote *mirt*.

O pacote *psych* Revelle (2013), baseado na análise paralela sobre a matriz de correlação tetracórica, auxilia na identificação da dimensionalidade dos dados em estudo, adota a junção de dois testes (componentes principais e análise fatorial). Espera-se encontrar pelo menos dois traços latentes, visto que a prova objetiva possui questões gerais e questões específicas. A Figura 2 apresenta o resultado da técnica aplicada aos dados deste trabalho, considerando todos os 32 itens e também excluído as duas questões (G1 e Q23), determinada pela comissão do INEP.

O instrumento verifica as dimensões interpretáveis e observa pelos autovalores que existe um fator bastante expressivo; na seqüência, outro fator bem menor, mas ainda destacado dos demais, um indicativo de 2 traços latentes significativos, configurando bidimensionalidade. Os demais fatores associados aos dados têm contribuições ínfimas, e assumir mais traços latentes na modelagem não traria melhorias significativas ao ajuste do modelo e entendimento do comportamento dos dados. Pensando nas competências do Enade, pode-se supor que o fator mais relevante poderia ser a interpretação, bastante útil nas duas habilidades avaliadas, e o segundo

fator a competência estatística. Essa constatação pode indicar que os dois testes estejam avaliando basicamente as mesmas coisas. Diante disso, optou-se por verificar o ajuste a modelos uni/multidimensionais de 2 fatores, de resposta dicotômica.

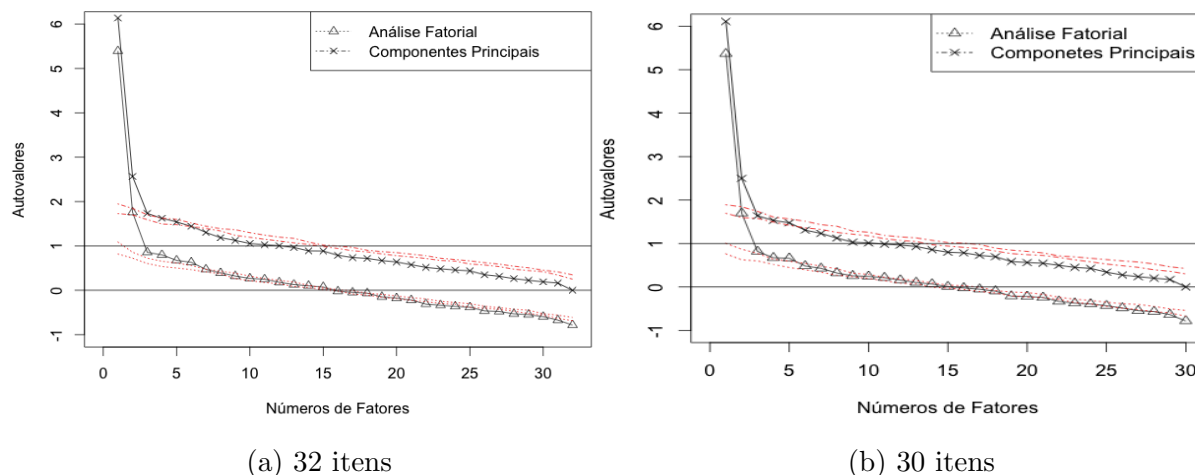


Figura 2: Gráficos para diagnosticar a dimensão dos 436 respondentes com: (a) 32 questões e (b) 30 questões excluindo a G1 e Q23.

Será empregado no modelo logístico da TRIM o modelo logístico com 1 traço latente (MLD1) e com 2 traços latentes (MLD2), visto que ambos contemplam a TRIU e a TRIM, cujo os parâmetros logísticos dos itens são abordados em: estimação da discriminância e o escalar que representa a dificuldade (2PL); estimação do 2PL, mas considerando que o acerto ao acaso é fixo em um único valor (2PLf); e os três parâmetros dos itens sendo estimados (3PL). A diferença é que conforme aumenta-se a dimensão, o modelo também adapta-se às dimensões propostas, mas o questionamento é: quanto vale a pena abordar mais que uma dimensão? Com esse questionamento, segue a Tabela 2 considerando para essas análises as 32 questões e excluindo a G1 e Q23.

Para os resultados da Tabela 2, utilizou-se o pacote *mirt*, com o procedimento iterativo de estimação do parâmetro abordando a máxima verossimilhança marginal e a integração pelo método de Gauss - Hermite utilizando a Estimativa de Máxima verossimilhança tradicional (EM). Os modelos exploratórios e de confirmação são estimadas por um algoritmo estocástico. Mais detalhes em Cai (2010 a, b).

Com os modelos aninhados pela Tabela 2, são comparadas os resultados através do valor da log-verossimilhança ($\log\text{Lik}$) e pelos critérios de AIC/BIC para a indicação do modelo. Observa que todos os modelos ajustaram, e o indicado com os 32 itens pelo $\log\text{Lik}$ é o MLD2 com 3PL, pelo critério de AIC o menor é o MLD2/2PLf e o BIC o MLD1/2PLf. É importante definir que o valor considerado fixo do acerto ao acaso do 2PLf foi determinado estimando-se e perfilhando os modelos pela máxima verossimilhança, conforme segue os valores do c_i que, neste caso, foram quase equivalentes entre os modelos dimensionais. Ao se retirar os itens G1 e Q23, o modelo pela $\log\text{Lik}$ é o MLD2/3PL, para o AIC, é o MLD1/3PL; para o BIC, é o MLD1/2PLf.

Como os resultados não foram unânimes, compara-se a quantidade estimada de parâmetros (df) com os padrões de resposta, neste caso, para todos os modelos propostos, existem mais padrões de resposta do que parâmetros estimados. Para esgotar e finalizar as análises desta tabela, testa-se a razão da verossimilhança entre os modelos aninhados, para verificar se compensa ir de um modelo mais simples para um mais complexo. Testou-se o MLD1/2PLf com

Tabela 2: Resultados das estatísticas de diagnostico da qualidade do ajuste dos modelos

MLD1	436 respondentes			
	df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0$)	64	-7727,53	15583,07	15844,04
2PLf ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0,097$)	65	-7699,55	15527,11	15789,1
3PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c_i$)	96	-7674,83	15541,65	15934,64
MLD2				
2PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0$)	96	-7665,22	15520,44	15909,33
2PLf ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0,097$)	97	-7649,13	15488,27	15877,16
3PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c_i$)	128	-7624,53	15503,07	16022,95
436 respondentes sem as questões G1 e Q23				
MLD1	df	logLik	AIC	BIC
2PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0$)	64	-7335,14	14790,28	15034,94
2PLf ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = 0,103$)	65	-7306,36	14732,74	14977,39
3PL ($\alpha_{i1} = \alpha_i, d_i = d_i, c_i = c_i$)	96	-7287,16	14754,33	15121,32
MLD2				
2PL ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0$)	96	-7275,09	14728,18	15091,09
2PLf ($\alpha_{i1} = \alpha_{i1}, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = 0,103$)	97	-7259,39	14696,79	15059,7
3PL ($\alpha_{i1} = \alpha_i, \alpha_{i2} = \alpha_{i2}, d_i = d_i, c_i = c_i$) ($c_i = c_i$)	128	-7239,05	14716,10	15201,34

FONTE:Autor (2014)

NOTA: Questões Gerais (G) e questões específicas (Q); diferença de parâmetros (df). Modelo logístico de um traço latente (MLD1); modelo logístico de dois traços latentes (MLD2); dois parâmetros logístico (2PL); dois parâmetros logístico estimado e o acerto ao acaso fixado (2PLf) e três parâmetros logístico (3PL) .

MLD2/2PLf e MLD1/3PL com MLD2/3PL a 95% de confiança e $df = 32$ para ambos os modelos, e também quando retira-se as questões, o teste compensa o modelo com mais parâmetros estimados e com duas dimensões.

Continuando nos modelos exploratórios, apresenta-se na Tabela 3, os resultados dos carregamentos dos fatores de explicação, e conforme Nosoja (2002), Reckase (2009) e Chalmers (2013), a técnica das cargas fatoriais possibilita informar e também averiguar quais itens se relacionam a cada fator. Quanto maior a carga fatorial, melhor é a relação entre item e o fator. Até porque, mesmo que com os testes já realizados e sendo o modelo indicado um mais complexos, vale ressaltar que computacionalmente e metodologicamente a relação entre custo benefício tem que valer a pena.

Segue a Tabela 3, os resultados indicados pelas análises dos modelos, após a rotação oblíqua das cargas fatoriais.

Na Tabela 3 para os 436 respondentes com as 32 questões, conforme os modelos indicados, é possível, analisar as cargas fatoriais dos itens nos modelos nas dimensões. O item G1 somente explica sua inclusão quando considera o acerto ao acaso fixo, no MLD2/2PLf, enquanto a Q23 explica MLD1 e MLD2 no 3PL. Observa-se, também, que incluindo ou não a G1 e a Q23, poucas questões carregaram explicando o F2. Para essa prova, a grande maioria das questões foi explicada no F1.

Observa no modelo dimensional MLD1 que 22 itens explicam melhor no 3PL, considerando todos os parâmetros estimados, do que fixando o acerto ao acaso no 2PLf. Outro fator importante é que devido ao resumo, não foi incluído o MLD1 das 30 questões na tabela, porque como é visto até no MLD2, há poucas diferenças numéricas entre os carregamentos.

Na tabela existe a indicação de excluir a G2, já que não carrega em nenhuma modelo e foi indicada também na análise descritiva pelo ponto bisserial na prova. Como o Enade envolve análise de itens, seria necessário um especialista em questões, para verificar os itens G1, G2 e Q23 antes de excluí-los devido os impactos das análises ao retirá-los. Portanto, seria conveniente um modelo unidimensional e multidimensional, mas com benefício, e como são poucas questões que carregaram na segunda dimensão, é vantajoso indicar o modelo MLD1/3PL. Como a indicação é do modelo unidimensional, não aborda-se o modelo confirmatório para confirmar a segunda dimensão.

Tabela 3: Cargas fatoriais dos modelos para as 32 questões e retirando a G1 e Q23

Q	436 respondentes e 32 questões									436 respondentes e 30 questões						
	MLD1				MLD2					MLD2						
	2PLf		3PL		2PLf			3PL		2PLf		3PL		3PL		
	F1	h2	F1	h2	F1	F2	h2	F1	F2	h2	F1	F2	h2	F1	F2	h2
G1	0,23	0,05	0,09	0,01	0,37	-0,67	0,59	0,07	-0,16	0,03	-	-	-	-	-	-
G2	0,12	0,01	0,12	0,01	0,08	-0,22	0,06	0,1	-0,18	0,04	0,08	-0,26	0,07	0,16	-0,26	0,09
G3	0,36	0,13	0,34	0,11	0,32	-0,28	0,18	0,29	-0,29	0,17	0,32	-0,28	0,18	0,29	-0,27	0,16
G4	0,28	0,08	0,53	0,28	0,31	0,12	0,11	0,25	0,08	0,07	0,32	0,12	0,12	0,25	0,09	0,07
G5	0,38	0,15	0,48	0,23	0,37	-0,2	0,18	0,54	-0,29	0,37	0,36	-0,24	0,19	0,52	-0,27	0,35
G6	0,45	0,2	0,44	0,19	0,44	-0,37	0,33	0,48	-0,42	0,41	0,43	-0,38	0,34	0,49	-0,41	0,4
G7	0,63	0,4	0,61	0,37	0,47	-0,78	0,84	0,41	-0,76	0,75	0,48	-0,81	0,88	0,46	-0,76	0,8
G8	0,66	0,44	0,68	0,46	0,5	-0,57	0,58	0,54	-0,51	0,55	0,51	-0,56	0,58	0,56	-0,49	0,56
Q11	0,39	0,15	0,38	0,14	0,38	-0,04	0,15	0,4	0,02	0,16	0,39	-0,02	0,15	0,4	0,05	0,16
Q12	0,35	0,13	0,36	0,13	0,31	-0,38	0,24	0,26	-0,39	0,22	0,32	-0,35	0,22	0,27	-0,38	0,22
Q13	0,81	0,66	0,82	0,68	0,73	0,57	0,86	0,63	0,43	0,58	0,76	0,51	0,84	0,61	0,45	0,57
Q15	0,42	0,18	0,8	0,63	0,43	0,07	0,19	0,85	-0,22	0,77	0,44	0,05	0,19	0,85	-0,16	0,75
Q16	0,57	0,32	0,68	0,46	0,55	-0,08	0,31	0,63	-0,08	0,41	0,55	-0,08	0,31	0,64	-0,05	0,41
Q17	0,52	0,27	0,86	0,73	0,53	0,05	0,29	0,91	-0,13	0,84	0,54	0,06	0,3	0,86	-0,05	0,75
Q18	0,94	0,88	0,97	0,94	0,89	0,41	0,96	0,86	0,49	0,99	0,9	0,39	0,96	0,85	0,5	0,98
Q19	0,7	0,49	0,83	0,69	0,67	0,13	0,47	0,82	0,13	0,68	0,69	0,16	0,51	0,81	0,17	0,68
Q20	0,76	0,57	0,81	0,66	0,68	0,16	0,48	0,71	0,01	0,5	0,77	-0,01	0,6	0,77	-0,06	0,6
Q21	0,65	0,43	0,66	0,44	0,65	0,08	0,44	0,71	0,12	0,52	0,66	0,13	0,46	0,68	0,16	0,49
Q23	0,23	0,05	0,92	0,85	0,28	0,22	0,13	0,93	0,19	0,9	-	-	-	-	-	-
Q24	0,36	0,13	0,37	0,14	0,34	-0,06	0,12	0,8	0,08	0,65	0,35	-0,02	0,12	0,72	0,12	0,53
Q25	0,76	0,58	0,7	0,49	0,8	-0,19	0,68	0,54	0,01	0,29	0,82	-0,21	0,72	0,71	-0,07	0,51
Q26	0,7	0,49	0,63	0,39	0,7	-0,1	0,5	0,64	-0,14	0,43	0,72	-0,12	0,53	0,66	-0,12	0,45
Q27	0,65	0,42	0,62	0,38	0,65	-0,15	0,44	0,61	-0,1	0,39	0,66	-0,14	0,45	0,62	-0,09	0,39
Q28	0,5	0,25	0,74	0,54	0,49	0,16	0,26	0,84	0,41	0,87	0,5	0,16	0,28	0,82	0,44	0,86
Q29	0,33	0,11	0,72	0,52	0,34	0,19	0,15	0,64	0,42	0,59	0,36	0,22	0,18	0,61	0,43	0,55
Q30	0,28	0,08	0,36	0,13	0,29	0,08	0,09	0,28	0,04	0,08	0,3	0,07	0,1	0,27	0,03	0,07
Q31	0,62	0,39	0,59	0,35	0,62	-0,09	0,39	0,59	-0,13	0,36	0,62	-0,09	0,39	0,58	-0,11	0,35
Q32	0,27	0,07	0,24	0,06	0,26	-0,09	0,08	0,91	-0,4	0,98	0,27	-0,09	0,08	0,92	-0,35	0,98
Q33	0,73	0,54	0,82	0,67	0,72	-0,03	0,53	0,85	-0,06	0,73	0,74	-0,03	0,55	0,85	-0,04	0,73
Q34	0,82	0,67	0,9	0,81	0,72	0,41	0,68	0,64	0,36	0,54	0,75	0,37	0,71	0,62	0,36	0,51
Q36	0,72	0,51	0,79	0,63	0,67	0,23	0,49	0,71	0,26	0,57	0,69	0,19	0,51	0,7	0,27	0,56
Q37	0,62	0,38	0,81	0,66	0,64	0,11	0,42	0,83	-0,05	0,7	0,64	0,09	0,42	0,84	-0,04	0,71

FONTE: Autor (2014).

NOTA: Questões Gerais (G) e questões específicas (Q); Modelo logístico de um traço latente (MLD1); modelo logístico de dois traços latentes (MLD2); dois parâmetros logístico (2PL); dois parâmetros logístico estimado e o acerto ao acaso fixado (2PLf) e três parâmetros logístico (3PL).

O resultado do ajuste do modelo indicado MLD1/3PL, pode ser visto através das Curvas Características do Itens (CCI), as quais apresentam todas as informações relevante proveniente dos 32 itens. Visto que poucos itens necessitaram mais do que uma habilidade, segue a Figura 3, através de um gráfico a habilidade dos indivíduos, com a probabilidade de estes responderem corretamente a cada um dos itens. Ou seja, é um gráfico onde no eixo X está a habilidade, e no eixo Y, a probabilidade de acerto dada a habilidade.

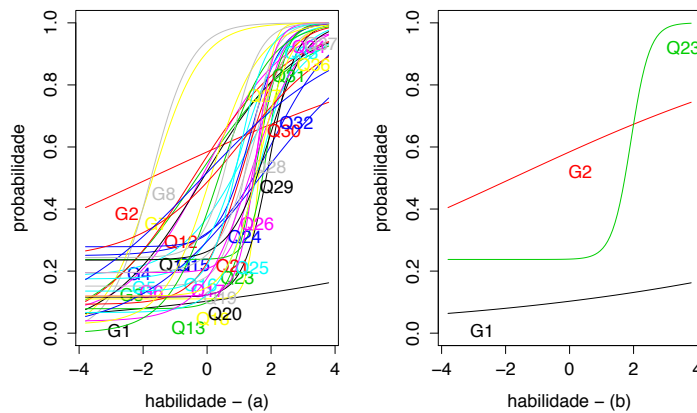


Figura 3: Curva Característica: (a) dos 32 itens e (b) dos itens 1, 2 e 23

A Figura 3 contempla os 32 itens na figura (a) e na figura (b) os itens G1, G2 e Q23, para não dificultar a visualizações dos itens. Assim, conforme as CCI das questões representando o MLD1/3PL, é possível identificar que a questão Q23 teve um bom comportamento, enquanto a G2 teve alta probabilidade de chute e pouca discriminância entre as questões, em comparação com G1 que teve pequena probabilidade de acerto ao acaso. Mais detalhes podem ser encontrados nos relatórios do Enade (2009), e recomenda-se olhar as questões para entender os motivos que levaram esse comportamento da área geral analisada.

Como ainda não temos no Enade a utilização da TRI, prossegue-se as análises verificando o escore dos respondentes, conforme a indicação unidimensional do MDL1/3PL e transformando o escore resultante do modelo MDL1/3PL para escala métrica entre 0 a 1. Segue a Figura 4 para visualização das notas da prova objetiva pela TRI.

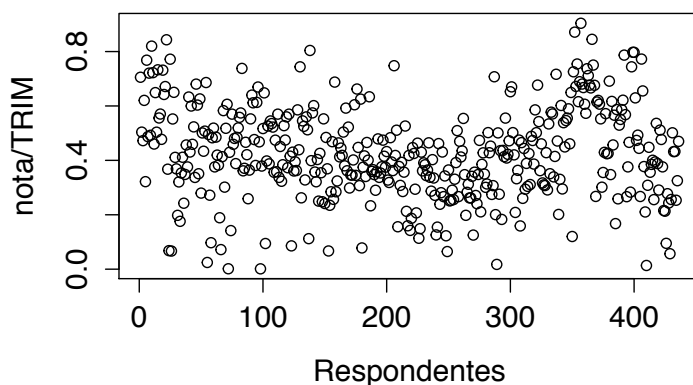


Figura 4: Escore dos respondentes pela da TRI

A Figura 4 proporciona uma visualização geral de como os alunos sucederam na avaliação objetiva do Enade/2009 com 32 questões, ou seja, poucos alunos tiveram nota acima 0,8. A grande maioria ficou entre 0,4 a 0,6. Com as análises descritas no resumo, é possível detalhar os itens do Enade/2009 e ter o apanhado geral da prova e o desempenho dos concluintes do curso de Estatística.

4 Conclusões

Este resumo apresentou o emprego dos modelos de TRIU e TRIM ao Enade/2009 no curso de graduação de Estatística para 443 concluintes. Os resultados mostraram que, em geral, a prova apresenta boa coerência interna com a maioria dos itens, bem como boa aderência ao instrumento de medida proposto.

A principal contribuição da TRI é apresentar de forma sistemática a análise não apenas dos alunos, mas também da prova, dando uma visão crítica de sua construção e de sua capacidade em aferir o conhecimento dos alunos. Quando trata os itens, que foram excluídos tanto no estudo como pela comissão do INEP utilizando a análise descritiva, verifica-se que apresentaram, tanto pelo ponto bisserial como no coeficiente de *Cronbach's alpha*, as mesmas indicações. Vale ressaltar que ao se detalhar o modelo pela TRI considerando-se todas as questões ou excluindo-se a G1 e a Q23 e indicando o item G2 como item que requeria atenção, devido ao alto índice de acerto ao acaso e por ter, conforme a Figura 3, uma CCI fora do padrão de discriminância, como no caso do item G1.

A questão Q23 apresentou bom desempenho para o modelo da TRI e a CCI está adequada para a prova. Essas indicações são relevantes para apresentar informações e tomadas de decisões e, nesse caso, quanto à dificuldade de cada item, e quando os itens têm diferentes níveis de discriminância e probabilidade de acerto casual, apresentaram resultados concisos. Portanto, a TRI é mais um auxílio ao ENADE.

Com os resultados, o ENADE pode contar com mais um instrumento para a busca da melhoria da qualidade do ensino e desempenho do aluno. É também um aporte que pode auxiliar os professores e as instituições a identificar as disciplinas que geram mais dificuldades e que necessitam de mais atenção, para garantir o desenvolvimento de competências pelos alunos.

Outro fator importante foi a abordagem da TRIU e da TRIM, bem como verificar o comportamento dos parâmetros dos itens e os ajustes dos modelos, tanto na prova de conhecimentos gerais como na de conhecimentos específicos. Pode ser considerada pela análise dos paralelos com 2 dimensões, embora o modelo unidimensional tenha se ajustado bem aos padrões de resposta dos itens, devido às análises feitas pelo carregamento das cargas nos itens. A escolha entre o modelo uni e o multidimensional da TRI depende das características da avaliação e dos avaliados, especialmente, depende dos objetivos da análise.

Como proposta para futuras pesquisas, sugere-se a avaliação deste exame considerando modelos da TRI que seja capazes de contemplar a avaliação de mais de uma habilidade latente, bem como considerem a influência de várias habilidades na resposta de cada item.

Referências

- [1] ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da Resposta ao Item: conceitos e aplicações**. São Paulo: ABE Associação Brasileira de Estatística, 2000.
- [2] ANDRIOLA, W. B. **Uso da Teoria de Resposta Ao Item (TRI) para analisar a equidade do processo de avaliação do aprendizado discente**. Revista Iberoamericana de Avaliação Educacional, v. 1, p. 171-189, 2008
- [3] ——. **Avaliações Externas e Seu Uso na Gestão Educacional**. Congresso Internacional Educação: Uma Agenda Urgente. Textos para Consulta. pág. 6. Brasília, 2011.
- [4] BAKER, F. B. **The basics of Item Response Theory**. 2. ed. USA: ERIC Clearinghouse on Assessment and Evaluation, 2001.
- [5] BAKER, F. B.; KIM, S. **Item Response Theory: parameter estimation techniques**. 2. ed. revised and expanded. New York: Marcel Dekker, 2004.
- [6] BRASIL. MINISTERIO DA EDUCAÇÃO. INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS (INEP). **Mapa da educação superior**. Brasília, PF: MEC; INEP, 2004. 85p.
- [7] BRASIL. MEC/INEP. **Relatório IDEB**. Disponível por: <http://portalideb.inep.gov.br/>. Acesso em: 22 mar. 2012.
- [8] BOOMSMA A.; VAN DUIJN, M. A. J.; SNIJDERS, T. A. B. **Essays on item response theory**. Lecture Notes in Statistics (Springer-Verlag), n. 157. New York: Springer, 2000.
- [9] BONAT, H. W. et al. **Métodos Computacionais em Inferência Estatística**. ABE- Associação Brasileira de Estatística, SINAPE, 2012.
- [10] Cai L (2010a). **High-Dimensional Exploratory Item Factor Analysis by a Metropolis-Hastings Robbins-Monro Algorithm**. Psychometrika, 75(1), 33-57.
- [11] Cai L (2010b). **Metropolis-Hastings Robbins-Monro Algorithm for Confirmatory Item Factor Analysis**. Journal of Educational and Behavioral Statistics, 35(3), 307-335.
- [12] Chalmers Phili. **R. mirt: A Multidimensional Item Response Theory Package for the R Environment**. Vol. 48, Issue 6, May 2012.
- [13] EMBRETSON, S.; REISE, S. P. **Item Response Theory for Psychologists**. New Jersey: Lawrence Erlbaum Associates, Inc. Publishers, 2000.

- [14] QUARESMA, S.E.; DIAS, S. T. C.; SARTORIO, D.S. **Avaliação da aprendizagem e das provas do centro de formação interdisciplinar/UFOPA via Teoria da Resposta ao Item**. UFOPA. Disponível em: <http://www.sbec.org.br/evt2012/trab16.pdf>. Acesso em: 25 jul. 2013.
- [15] FRANCISCO, R. **Aplicação da Teoria da Resposta ao Item (TRI) no Exame Nacional de Cursos (ENC) da Unicentro**. 2005. 144 f. Dissertação (Mestrado em Ciências) - Pós- Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, 2005.
- [16] LORD F.M. **Applications of item response theory to practical testing problems**. Hillsdale: Erlbaum; 1980.
- [17] NOJOSA, Ronald T. **Modelos Multidimensionais para a Teoria de Resposta ao Item**. Pernambuco, UFPE, Tese de Mestrado, 2001.
- [18] R. DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2009. Disponível em: <http://www.lsw.uni-heidelberg.de/users/christlieb/teaching/UKStaSS10/R-refman.pdf>. Acesso em: 25 jul. 2013.
- [19] PITON-GONÇALVES, J. **Desafios e Perspectivas da Implementação Computacional de Testes Adaptativos Multidimensionais para Avaliações Educacionais**. Tese de Doutorado, - Universidade de São Paulo, 2012
- [20] RECKASE, M. D. **Multidimensional Item Response Theory: Statistical for social and behavioral sciences**. Springer Science Business Media: LLC, 2009.
- [21] REVELLE, W. **Procedures for Psychological, Psychometric, and Personality Research**. Acesso em 09 de 12 de 2013, disponível em The Comprehensive R Archive Network: <http://cran.r-project.org/web/packages/psych/psych.pdf>
- [22] RIZOPOULOS, D. **ltm: An r packages for latent variable modelling and item response theory analyses**. R package. Disponível em: <http://cran.R-project.org/package=ltm>
- [23] VAN DER LINDEN, W. J.; HAMBLETON, R. K. **Handbook of Modern Item Response Theory**. New York: SpringerVerlag, 1997.
- [24] VENDRAMINI, C. M. M.; DIAS, A. S. **Teoria de Resposta ao Item na análise de uma prova de estatística em universitários**. Psico-USF, v. 10, n. 2, p. 201-210, jul./dez. 2005.
- [25] YANYAN Sheng and Todd C. Headrick. **A Gibbs Sampler for the Multidimensional Item Response Model**. Section on Statistics and Measurement, Department of EPSE, Southern Illinois University Carbondale, Wham 223, MailCode 4618, Carbondale, IL 62901-4618, USA. Received 2 March 2012; Accepted 26 March 2012