

Utilização da Teoria da Resposta ao Item na Análise dos Dados da Autoavaliação Institucional da Universidade Federal de Santa Maria: Vantagens e Oportunidades

Universidade Federal de Santa Maria (UFSM)

Eixo II - Indicadores e instrumentos de autoavaliação

Fernando de Jesus Moreira Junior (Universidade Federal de Santa Maria)

Charles Rogério Pavaglio Szinvelski (Universidade Federal de Santa Maria)

Resumo

A análise dos dados da autoavaliação institucional da Universidade Federal de Santa Maria (UFSM) é realizada por meio de estatísticas descritivas, tabelas de frequência e gráficos apresentados anualmente no relatório de autoavaliação institucional. Esse trabalho tem como objetivo apresentar as vantagens e as oportunidades que a utilização da técnica estatística conhecida como Teoria da Resposta ao Item (TRI) pode fornecer no sentido de complementar a análise que tradicionalmente é realizada. A TRI é uma metodologia muito utilizada nas avaliações educacionais internacionais em larga escala e, inclusive, vem sendo utilizada recentemente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) na avaliação do Exame Nacional do Ensino Médio (ENEM). Embora a maioria das aplicações esteja relacionada com as avaliações educacionais, a TRI também tem sido utilizada em diversas outras áreas, tais como, nas ciências administrativas, nas ciências da saúde, na psicologia, nas engenharias, etc. A TRI utiliza modelos de probabilidade para avaliar traços latentes, por exemplo, nível proficiência ou conhecimento, satisfação de clientes, grau de depressão, etc. A TRI permite analisar os itens (conjuntamente e individualmente) por meio das Curvas Características e das Curvas de Informação de cada Item e do Instrumento além de criar uma escala interpretável, onde os itens e os respondentes são posicionados. Para executar a análise desse trabalho, foram utilizados os bancos de dados que continham as respostas dos indivíduos que participaram da autoavaliação institucional realizada em 2012. As respostas foram dicotomizadas para fins de utilização do Modelo Logístico Unidimensional de Dois Parâmetros (ML2). Os dados foram analisados no Software R. Esse estudo caracteriza-se como pesquisa descritiva e de análise quantitativa. Os resultados permitiram verificar a qualidade dos itens que compõem o instrumento de autoavaliação institucional, que em geral possuíam um bom poder discriminatório. O próprio instrumento como um todo também pode ser analisado, no qual constatou-se uma concentração maior de informação à esquerda da escala criada. Nessa escala, os itens e indivíduos foram posicionados, o que permitiu a identificação de níveis na escala e a interpretação desses níveis. Os resultados mostraram que a TRI pode contribuir nas análises dos dados da autoavaliação institucional, fornecendo informações adicionais relacionadas com a qualidade do instrumento como um todo, dos itens individualmente e da avaliação dos respondentes.

Palavras-chave: Teoria da Resposta ao Item, autoavaliação institucional.

Introdução

A avaliação representa um meio para que a Instituição reconheça suas potencialidades e fragilidades, auxiliando na definição das prioridades que contribuirão para seu efetivo desenvolvimento institucional (UFSM, 2013). Podem participar da avaliação os Gestores, os Professores, os Técnico-Administrativos em Educação, os Alunos de Pós-Graduação, Graduação, da Educação Básica, Técnica e Tecnológica e Egressos que ingressaram na Instituição a partir de 1996. A análise dos dados da autoavaliação institucional da Universidade Federal de Santa Maria (UFSM) é realizada por meio de estatísticas descritivas, tabelas de frequência e gráficos apresentados anualmente no relatório de autoavaliação institucional.

Esse trabalho propõe apresentar as vantagens e as oportunidades que a utilização da técnica estatística conhecida como Teoria da Resposta ao Item (TRI) pode fornecer no sentido de complementar a análise que tradicionalmente é realizada. Segundo Andrade, Tavares e Valle (2000), a TRI é uma metodologia que sugere formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta a um item (questão), os traços latentes do indivíduo e as características dos itens, por meio de modelos matemáticos.

A Universidade Federal de Santa Maria, idealizada e fundada pelo Prof. Dr. José Mariano da Rocha Filho, foi criada pela Lei N. 3.834-C, de 14 de dezembro de 1960, com a denominação de Universidade de Santa Maria, instalada solenemente em 18 de março de 1961 e federalizada em 20 de agosto de 1965, pela Lei N. 4.759/65. Atualmente possui cinco Campus, onde oferece mais de 300 Cursos, atendendo quase 30.000 alunos com quase 2.000 docentes e 3.000 técnicos administrativos.

Os professores autores do presente trabalho fazem parte da Comissão de Avaliação Institucional do Centro de Ciências Naturais e Exatas (CCNE) da Universidade Federal de Santa Maria (UFSM), sendo que o professor Fernando está na comissão há cinco meses, enquanto que o professor Charles está na comissão há quase dois anos e, atualmente, é o Presidente da mesma. Quanto à técnica estatística conhecida como Teoria da Resposta ao Item (TRI) empregada na análise do presente estudo, salienta-se que o professor Fernando estuda e aplica essa técnica há seis anos, tendo inclusive empregado-a na sua Tese de Doutorado, e também lidera um Grupo de Pesquisa na UFSM com foco nessa técnica, denominado “Grupo de Estudos e Pesquisas em Teoria da Resposta ao Item – GEPTRI”.

Esse trabalho tem como objetivo apresentar as vantagens e as oportunidades que a utilização da Teoria da Resposta ao Item (TRI), por meio de uma análise inicial e exploratória.

Objetivos

O objetivo geral desse trabalho é, por meio de uma análise inicial e exploratória, apresentar as vantagens e as oportunidades que a utilização da Teoria da Resposta ao Item (TRI) pode fornecer para complementar a análise que tradicionalmente é realizada autoavaliação institucional.

Os objetivos específicos são:

- Analisar os itens do questionário por meio da TRI.
- Estimar o nível de satisfação dos respondentes por meio da TRI em relação aos itens avaliados.
- Criar uma escala interpretável para a avaliação com base na TRI.

Metodologia

Para executar a análise desse trabalho, foram utilizados os bancos de dados que continham as respostas dos indivíduos que participaram da autoavaliação institucional realizada em 2012, no período de 12 de novembro de 2012 a 11 de janeiro de 2013, cujos resultados das análises tradicionais encontram-se disponíveis em CPA (2012). Esse banco de dados foi disponibilizado pelo Centro de Processamento de Dados (CPD) da UFSM.

Pelo fato do questionário não ser exatamente o mesmo para os diferentes grupos de respondentes, optou-se trabalhar somente com os dados referentes à avaliação realizada pelos discentes dos Cursos de Graduação da UFSM. Esse grupo foi escolhido por ter sido o grupo com maior quantidade de respondentes, uma vez que a análise por meio da TRI necessita de uma amostra grande de respondentes para fornecer resultados consistentes. No total, 3.506 discentes dos Cursos de Graduação participaram da avaliação. O questionário de autoavaliação discentes dos Cursos de Graduação é composto por 46 questões, sendo 45 de caráter objetivo e uma de caráter subjetivo. Como na TRI só é possível analisar questões objetivas, a questão subjetiva foi ignorada. Dentre as 45 questões objetivas, uma perguntava se o respondente conhecia o Projeto Pedagógico de seu Curso (PPC), enquanto que as demais avaliavam a satisfação do aluno em relação a diversos itens e possuíam as mesmas categorias de respostas: (1) excelente, (2) muito bom, (3) bom, (4) regular e (5) insatisfatório. Como o traço latente avaliado aqui é a satisfação do aluno, a questão relacionada com o conhecimento do PPC não foi considerada na análise.

Os dados foram analisados no *Software R*, por ser um *software* livre e por ser um dos poucos *softwares* que realizam análise por meio da TRI. Há vários pacotes computacionais no R que trabalham com TRI. No entanto, foi escolhido o pacote *irtoys* (PARTCHEV, 2013) devido ao fato de trabalhar com Modelos dicotômicos e apresentar boas opções de estimação e de análises gráficas. Porém, antes da análise por meio da TRI, foi realizada uma análise tradicional com o pacote *ltm* (RIZOPOULOS, 2013).

No pacote *irtoys*, a estimação dos parâmetros dos itens é feita pelo pacote *ltm*, o qual utiliza o método da Máxima Verossimilhança. Por outro lado, para estimar os parâmetros dos indivíduos, o pacote *irtoys* oferece quatro métodos. Dentre esses, foi escolhido o método bayesiano da Média a Posteriori (EAP) por ser um método que obtém o menor erro padrão e a maior fidelidade (HONTANGAS; PONSODA; OLEA, 1999), além de não necessitar de métodos iterativos, já que são dados os pontos de quadratura e não é necessário calcular as complexas integrais (COSTA, 2009), motivo pelo qual é mais recomendada.

As respostas foram dicotomizadas para fins de utilização do Modelo Logístico Unidimensional de Dois Parâmetros (ML2), um dos modelos dicotômicos disponível no pacote *irtoys* do R. A dicotomização foi realizada da seguinte forma: as alternativas (1) excelente, (2) muito bom e (3) bom foram consideradas como uma única categoria denominada (1) satisfatório; enquanto que as alternativas (4) regular e (5) insatisfatório foram consideradas como uma única categoria denominada (0) insatisfatório. Muitas questões ainda possuíam a alternativa “desconheço”, que foi considerada como “não resposta” na presente análise. No entanto, o pacote *irtoys* possui uma limitação, pois não reconhece os dados de “não resposta”. Dessa forma, todos os alunos que responderam “desconheço”, ao menos uma vez, tiveram que ser desconsiderados na presente análise. Assim, a amostra de alunos com dados válidos para essa análise foi reduzida para 983, que ainda sim é uma amostra suficientemente grande para uma análise da TRI como o modelo proposto (NUNES; PRIMI; 2005; EMBRETSON; REISE, 2000; HAMBLETON; SWAMINATHAM, 1985).

Referencial Teórico

Em relação à avaliação tradicional da Teoria Clássica dos Testes (TCT) ou Teoria Clássica da Medida (TCM), onde o traço latente é mensurado por meio de um somatório ponderado, ou não, dos pontos atribuídos à cada questão do questionário, a TRI apresenta alguns avanços (EMBRETSON; REISE, 2000). Os principais avanços são os procedimentos de equalização e a criação de escalas interpretáveis. Os procedimentos de equalização foram um grande avanço para a avaliação educacional, uma vez que permite que indivíduos avaliados por instrumentos de avaliação (provas) parcialmente diferentes (com alguns itens em comum) sejam colocados numa mesma escala, o que permite compará-los e acompanhar a sua evolução ao longo do tempo (ANDRADE; TAVARES; VALLE, 2000; EMBRETSON; REISE, 2000). A equalização pode ser feita por meio do método dos Grupos Múltiplos (BOCK; ZIMOWSKI, 1997), se o software permitir, ou por meio de métodos de equalização a posteriori (ANDRADE; TAVARES; VALLE, 2000). O outro grandes avanço da TRI é a possibilidade de criar escalas onde itens e respondentes são posicionados simultaneamente, o que permite a interpretação dessa escala (ANDRADE; TAVARES; VALLE, 2000). Além desses avanços, a TRI possui outras vantagens, tais como, a possibilidade de calcular a probabilidade de dar certa resposta a um determinado item que o indivíduo não tenha respondido (VENDRAMINI; SILVA; CANALE, 2004) e, na avaliação educacional, a possibilidade de verificar o acerto casual (ANDRADE; TAVARES; VALLE, 2000).

A TRI foi utilizada pela primeira vez no Brasil em 1995 no SAEB, para montagem de instrumentos, tratamento de dados e construção de escalas a partir de resultados apresentados por alunos em provas de rendimento (SOUZA, 2005), o que permitiu que alunos de diferentes séries pudessem ser comparados e colocados na mesma escala, que não era possível fazer através da TCT. A maioria das aplicações tem sido na avaliação educacional (MOREIRA JUNIOR, 2010), onde ENEM (Exame Nacional do Ensino Médio) tem sido, atualmente, o maior exemplo que mostra os benefícios da TRI.

A TRI é uma metodologia que propõe uma maneira de representar a relação entre a probabilidade de uma determinada resposta a um item, o traço latente, e as características dos itens, por meio de modelos de probabilidade (ANDRADE; TAVARES; VALLE, 2000; BAKER, 2001; DE AYALA, 2009; EMBRETSON; REISE, 2000; HAMBLETON; SWAMINATHAM, 1985; RECKASE, 2009). Em geral, tem-se: (1) um conjunto de itens, (2) uma amostra de respondentes e (3) um modelo de probabilidade associado, também chamado de Modelo de Resposta ao Item (MRI).

Uma análise de TRI normalmente segue o procedimento descrito a seguir. O conjunto de itens deve ser elaborado da forma tradicional, ou seja, por especialistas sobre o traço latente analisado, consultando a literatura existente sobre o assunto, abrangendo todo o conteúdo necessário (tendo o cuidado para que não fique demasiadamente extenso), fazendo pré-teste a análise de juizes, quando for conveniente. O conjunto de itens resultantes irá compor o questionário, ou seja, o instrumento de avaliação. Esse será submetido a uma amostra suficiente de respondentes. Esses respondentes podem ser indivíduos ou não (empresas, objetos, animais, etc.), dependendo da situação. As respostas obtidas serão analisadas por meio de um MRI, o qual irá fornecer estimativas para os parâmetros dos itens desse modelo. Os itens serão analisados por meio dessas estimativas e por meio da Função de Informação do Item (FII), que mede a qualidade do item para o construto. Se algum item tiver qualidade duvidosa, poderá ser eliminado da análise. Outros itens, com boas qualidades, irão se tornar “itens âncoras”, conceito que será explorado mais adiante. Na sequência, será criada a escala do traço latente, onde os itens serão posicionados. Independentemente da criação da escala, a TRI irá avaliar os

respondentes, atribuindo uma “nota” (um valor) para cada respondente, que é a estimativa para o parâmetro dos respondentes. Após isso, os respondentes poderão ser posicionados na escala criada. Enfim, com itens e respondentes posicionados na mesma escala, pode-se fazer a análise e interpretação da escala.

Tanto para estimação dos parâmetros dos itens (procedimento conhecido como “calibração dos itens”) quanto para a estimação do traço latente, há vários métodos estatísticos sofisticados que podem ser utilizados, tais como o método de Máxima Verossimilhança Marginal (MVM), o método bayesiano da Moda a Posteriori (MAP), o método bayesiano da Média a Posteriori (EAP) e o método da Máxima Verossimilhança Conjunta (MVC). Esses métodos não possuem solução explícita, o que torna necessária a utilização de algum método numérico iterativo, como o Algoritmo Newton-Raphson (ISSAC; KELLER, 1966), o Método Scoring de Fisher (RAO, 1973) e o Algoritmo EM (DEMPSTER; LAIRD; RUBIN, 1977). Essas soluções envolvem cálculos bastante complexos e, conseqüentemente, necessitam de programas computacionais específicos. Existem vários modelos matemáticos utilizados na TRI, diferentes quanto à sua função e à quantidade de parâmetros, e cada um deles é específico para uma (ou mais) situação. Esses modelos podem ser classificados quanto à sua dimensão (unidimensionais ou multidimensionais), quanto ao tipo de traço latente (cumulativo ou não cumulativo), quanto ao tipo de item (dicotômico ou politômico) e quanto ao número de populações envolvidas (MOREIRA JUNIOR, 2011). Nesse trabalho será dada ênfase ao Modelo Logístico Unidimensional de Dois Parâmetros (ML2), que é um modelo unidimensional, dicotômico e para uma única população.

O ML2, desenvolvido por Birnbaum (1968), é representado pela seguinte equação:

$$P(U_{ij} = 1 / \theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}},$$

onde:

a_i é o parâmetro de discriminação do item i , proporcional à inclinação da curva no ponto b_i , onde está localizado o item i ;

b_i é o parâmetro de dificuldade ou de posição do item i , e representa o valor do traço latente θ (no caso, nível de satisfação do aluno) na escala onde há 0,50 de probabilidade do indivíduo j (no caso, o aluno j) responder positivamente ao item;

U é a resposta ao item, que pode ser positiva ($U=1$), no caso, se o aluno considera o item satisfatório, ou negativa ($U=0$), caso contrário;

θ_j representa o valor do traço latente (no caso, nível de satisfação do aluno).

A relação entre a resposta prevista ao item e o traço latente do indivíduo é conhecido através da Curva Característica do Item (CCI). A Figura 1 apresenta um exemplo de uma CCI de um ML2 e a influência do parâmetro a_i (poder de discriminação) na inclinação da curva, e do parâmetro b_i (posição do item na escala). O eixo X indica o valor do traço latente (habilidade), considerando uma escala com média 0 e desvio padrão 1, enquanto que o eixo Y indica a probabilidade do indivíduo j responder positivamente o item dicotômico i , dado que ele tem habilidade θ_j .

Em uma comparação com a TCT, o parâmetro de dificuldade da TRI (b_i) está relacionado ao parâmetro de dificuldade da TCT (proporção de respostas positivas); o parâmetro de discriminação da TRI (a_i) está relacionado ao coeficiente de correlação Ponto Bisserial da TCT; e o parâmetro do traço latente (θ_j) está associado ao escore (contagem de respostas positivas) do indivíduo na TCT.

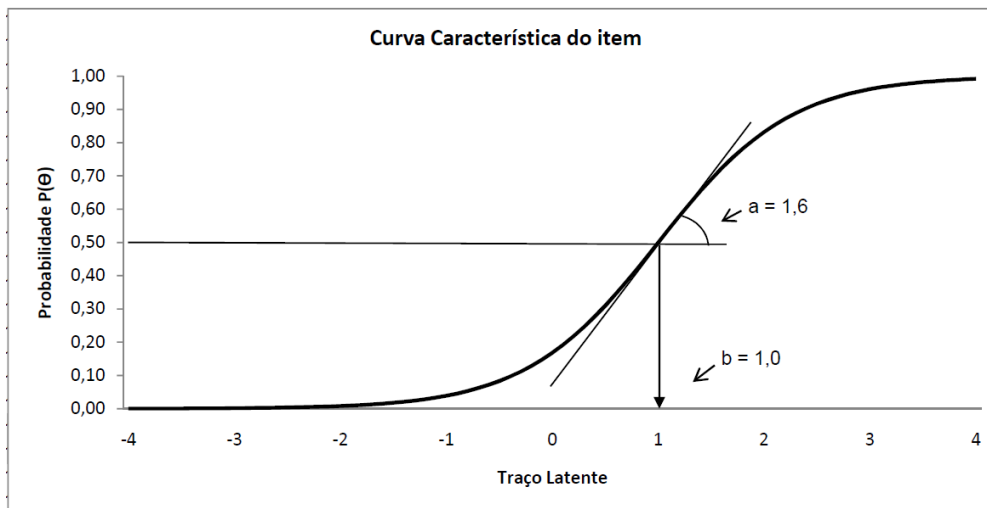


Figura 1: Curva Característica de um item hipotético.

No caso de medição da satisfação do aluno, o eixo x, da Figura 1, representa o nível de satisfação no qual é possível posicionar os itens e os alunos, o que possibilita verificar quais os itens que são satisfatórios para determinado aluno.

A contribuição de um item é a informação que ele fornece para o teste, e está diretamente relacionada com o quadrado da discriminação do item, ou seja, quanto maior a discriminação, mais informação o item fornece. No caso no MLU2P, abordado nesse trabalho, a informação do item é obtida pela expressão:

$$I_i(\theta) = a_i^2 P_i(\theta) Q_i(\theta).$$

A Função de Informação do Teste (FIT) é simplesmente a soma das informações fornecidas por cada item respondido pelo respondente, ou seja:

$$I(\theta) = \sum_{i=1}^I I_i(\theta).$$

Segundo Andrade, Tavares e Valle (2000), os valores da escala do traço latente (ou habilidade) podem assumir teoricamente qualquer valor real entre $-\infty$ e $+\infty$, diferentemente da TCT, onde a escala geralmente varia entre 0 e a quantidade total de questões do teste. No entanto, o importante nessa escala não é a sua magnitude, mas as relações de ordem existentes. Dessa forma, é preciso estabelecer valores para a média e para o desvio padrão que representem a escala de habilidade dos indivíduos na população.

A calibração dos itens, devido à facilidade computacional, geralmente é feita na escala (0,1), ou seja, numa escala com média igual a zero e desvio padrão igual a 1, onde os parâmetros b mais adequados devem variar entre -3 e 3 e os parâmetros a devem ser maior que 0,7. Após essa etapa, é realizada a construção da escala do traço latente, que é colocada na mesma métrica dos itens. Conforme, Fontanive, Elliot e Klein (2007), as escalas de habilidade ordenam o desempenho dos indivíduos do menor para o maior de forma contínua e são cumulativas, isto é, os indivíduos que situam-se em um determinado nível da escala são capazes de demonstrar as habilidades descritas nesse nível e nos níveis anteriores dessa escala.

No entanto, na prática, muitas pessoas não habituadas com a questão da mensuração matemática possuem dificuldades em compreender os valores negativos e decimais que existem na escala (0,1), onde muitas vezes os valores das proficiências são interpretados inadequadamente (VALLE, 2001). Nesse sentido, é bastante comum fazer uma transformação linear em todos os parâmetros envolvidos antes da construção das

escalas, fazendo com que todos os valores fiquem positivos e inteiros, de forma a representar a habilidade em valores numéricos que tornem mais fácil o entendimento.

A construção da escala de habilidade é efetuada após a calibração (e equalização, se necessário) dos itens, com o objetivo de encontrar uma interpretação qualitativa dos valores obtidos pela aplicação do modelo da TRI, possibilitando assim, a interpretação pedagógica dos valores das habilidades. Nesse sentido, surge a ideia dos níveis âncoras e a técnica conhecida como ancoragem (BEATON; ALLEN, 1992).

Andrade, Tavares e Valle (2000) definem níveis âncora como pontos selecionados pelo analista na escala da habilidade para serem interpretados pedagogicamente. Valle (2001) ressalta que esses níveis âncoras não podem ser muito próximos nem muito distantes, podendo-se tomar como base a média e o desvio padrão. Usualmente, os níveis âncoras são estabelecidos em função do desvio padrão em relação à média, obtendo-se, normalmente, cerca de até 7 níveis âncoras ($-3\sigma + \mu$, $-2\sigma + \mu$, $-1\sigma + \mu$, μ , $1\sigma + \mu$, $2\sigma + \mu$, $3\sigma + \mu$). Outras configurações de níveis âncoras podem ser definidas (por exemplo, KLEIN; FONTANIVE; ELLIOT, 2007), porém não são muito utilizadas na prática. No entanto, não se sabe a priori quantos níveis âncoras serão representados, já que isso depende da quantidade de itens âncoras.

Kolen e Brennan (1995) definem item âncora da seguinte forma: considere dois níveis âncora consecutivos Y e Z sendo que $Y < Z$. Um determinado item é âncora para o nível Z se e somente se as 3 condições abaixo forem satisfeitas simultaneamente:

$$P(U = 1 | \theta = Z) \geq 0,65,$$

$$P(U = 1 | \theta = Y) \leq 0,50 \text{ e}$$

$$P(U = 1 | \theta = Z) - P(U = 1 | \theta = Y) \geq 0,30.$$

Isso significa que, para um item ser considerado âncora em um determinado nível âncora, ele precisa ser respondido positivamente por um grande percentual de indivíduos (pelo menos 65%) com este nível de habilidade e por um percentual menor de indivíduos (no máximo 50%) com o nível de habilidade imediatamente anterior. Além disso, a diferença entre essas duas proporções deve ser de pelo menos 0,30. Assim, para um item ser âncora em um nível, ele deve ser um item “típico” desse nível, ou seja, bastante respondido positivamente por indivíduos com aquele nível de habilidade e pouco respondido positivamente por indivíduos com um nível de habilidade imediatamente inferior (ANDRADE; TAVARES; VALLE, 2000).

Na prática, às vezes um item se caracteriza por ser “quase âncora”, ao violar “levemente” uma das três condições necessárias. Nessas situações, muitos autores consideram esse item como sendo âncora, se ele for importante ou se existirem poucos itens no instrumento de pesquisa. Valle (2001) salienta que alguns níveis âncoras extremos podem ser mal caracterizados por serem definidos por itens muito fáceis ou muito difíceis, os quais geralmente são poucos.

Depois de identificados os itens âncoras de cada nível âncora, os especialistas no traço latente estudado devem caracterizar os níveis âncoras, segundo o conteúdo abordado no conjunto de itens que compõem cada nível.

Após essa etapa, a escala está pronta para ser utilizada, por exemplo, para o posicionamento das populações ou dos indivíduos a fim de verificar os conteúdos dominados ou para identificar o percentual de indivíduos em cada nível de habilidade (VALLE, 2001).

Assim, um sujeito que se situa acima de um determinado nível dominará os conteúdos caracterizados por esse nível e pelos níveis inferiores à esse, porém não dominará os assuntos dos níveis superiores a esse.

Resultados e Discussões

Após a formatação dos dados (dicotomização e exclusão dos casos com “não resposta”), foi feita uma análise tradicional baseada na TCT com recursos fornecidos pelo pacote *ltm* do R. A Tabela 1 apresenta a proporção de respostas “insatisfatório”, a proporção de respostas “satisfatório”, o Coeficiente do Correlação Ponto-Bisserial incluindo o item (CCPBc/I), o Coeficiente do Correlação Ponto-Bisserial desconsiderando o item (CCPBs/I) e o Alfa de Cronbach desconsiderando o item.

Observa-se que todos os itens foram considerados mais satisfatórios do que insatisfatórios, onde o Item 12 foi o menos satisfatórios (0,5148) e o item 21 foi o mais satisfatório (0,9166). Os Coeficientes da Correlação Ponto-Bisserial (incluindo ou não o item), que calcula a correlação entre a resposta dada aos itens pelo indivíduo e os escores dos indivíduos (soma de respostas positivas), medindo a coerência do item, mostraram que os itens não eram incoerentes com relação ao questionário. O Alfa de Cronbach incluindo todos os itens foi de 0,9420, praticamente o mesmo valor quando se excluía algum item (Tabela 1), mostrando que o questionário é consistente. Os resultados da TCT mostraram que não havia nenhum item inadequado.

Quanto aos alunos, embora não seja um dado com o qual a CPA trabalhe em seu relatório, obteve-se a frequência dos escores de satisfação dos alunos, ou seja, a contagem dos itens considerados satisfatórios para cada indivíduo. A informação é visualizada na Figura 2.

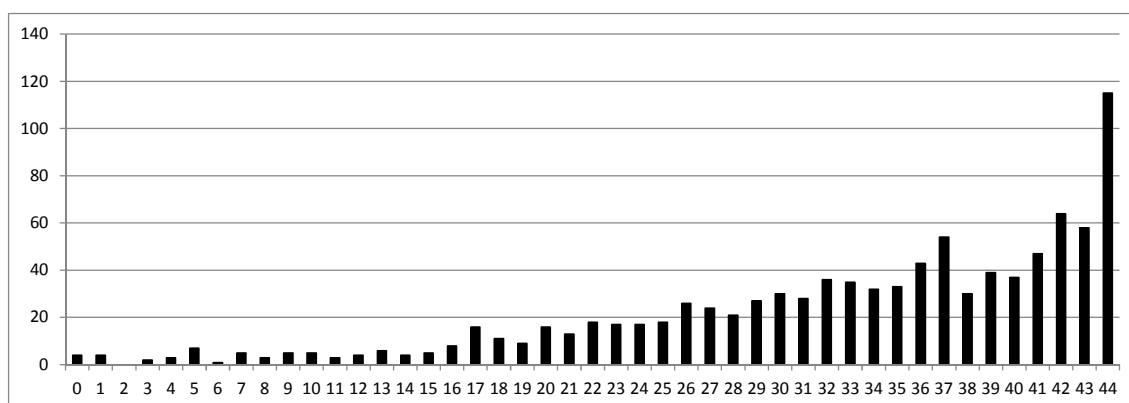


Figura 2 – Escore dos alunos

O gráfico da Figura 2 mostra que realmente são poucos os alunos que apresentam insatisfação com muitos itens. Cerca de 50% dos alunos estão satisfeitos com mais de 90% dos itens, sendo que 29% estão satisfeitos com mais de 90% dos itens. Apenas 15% estão insatisfeitos com 50% dos itens ou mais.

Finalizada a análise tradicional, passou-se então à análise dos dados por meio da TRI. A Tabela 2 apresenta a descrição dos itens e os valores estimados dos parâmetros de discriminação (*a*) e de dificuldade (*b*) dos itens, por meio do ML2 em uma escala com média 0 (zero) e desvio padrão 1 (um).

Tabela 1: Análise clássica dos itens.

ITEM	Proporção de insatisfatório	Proporção de satisfatório	CCPbc/I	CCPBs/I	Alfa de Cronbach s/I
Item 1	0,0966	0,9034	0,4685	0,4448	0,9412
Item 2	0,2370	0,7630	0,5177	0,4854	0,9409
Item 3	0,2075	0,7925	0,5426	0,5128	0,9407
Item 4	0,1322	0,8678	0,5042	0,4781	0,9409
Item 5	0,1465	0,8535	0,4837	0,4557	0,9411
Item 6	0,1506	0,8494	0,4985	0,4707	0,9410
Item 7	0,1882	0,8118	0,4209	0,3878	0,9415
Item 8	0,2055	0,7945	0,5185	0,4878	0,9408
Item 9	0,4059	0,5941	0,4877	0,4486	0,9412
Item 10	0,3693	0,6307	0,5826	0,5491	0,9404
Item 11	0,3459	0,6541	0,6024	0,5706	0,9402
Item 12	0,4852	0,5148	0,4117	0,3686	0,9419
Item 13	0,2024	0,7976	0,5775	0,5496	0,9404
Item 14	0,1424	0,8576	0,5217	0,4953	0,9408
Item 15	0,2248	0,7752	0,5143	0,4824	0,9409
Item 16	0,2045	0,7955	0,5875	0,5600	0,9404
Item 17	0,2136	0,7864	0,5782	0,5498	0,9404
Item 18	0,1892	0,8108	0,4742	0,4428	0,9411
Item 19	0,1689	0,8311	0,5669	0,5405	0,9405
Item 20	0,2330	0,7670	0,5655	0,5355	0,9405
Item 21	0,0834	0,9166	0,3868	0,3628	0,9416
Item 22	0,2085	0,7915	0,6152	0,5888	0,9401
Item 23	0,1099	0,8901	0,5099	0,4860	0,9409
Item 24	0,3540	0,6460	0,6220	0,5911	0,9401
Item 25	0,1017	0,8983	0,5161	0,4933	0,9409
Item 26	0,1089	0,8911	0,5074	0,4835	0,9410
Item 27	0,2574	0,7426	0,5664	0,5353	0,9405
Item 28	0,3103	0,6897	0,6165	0,5864	0,9401
Item 29	0,2635	0,7365	0,5873	0,5572	0,9403
Item 30	0,1740	0,8260	0,5072	0,4780	0,9409
Item 31	0,2889	0,7111	0,5188	0,4842	0,9409
Item 32	0,3988	0,6012	0,5335	0,4969	0,9408
Item 33	0,3744	0,6256	0,5703	0,5361	0,9405
Item 34	0,1943	0,8057	0,4323	0,3991	0,9414
Item 35	0,4283	0,5717	0,6313	0,5999	0,9400
Item 36	0,3703	0,6297	0,5912	0,5582	0,9403
Item 37	0,4201	0,5799	0,6166	0,5844	0,9401
Item 38	0,2157	0,7843	0,6188	0,5923	0,9401
Item 39	0,2970	0,7030	0,5844	0,5529	0,9404
Item 40	0,3530	0,6470	0,6254	0,5948	0,9400
Item 41	0,3245	0,6755	0,4583	0,4198	0,9414
Item 42	0,2970	0,7030	0,6140	0,5841	0,9401
Item 43	0,2594	0,7406	0,3976	0,3595	0,9418
Item 44	0,3143	0,6857	0,5375	0,5030	0,9408

Tabela 2: estimativas dos parâmetros de discriminação (a) e de dificuldade (b) dos itens.

ITEM	DESCRIÇÃO	a	b
Item 1	As disciplinas obrigatórias que compõem o currículo do seu curso para a sua formação técnica, profissional e cidadã, são:	1,7188	-1,8043
Item 2	As disciplinas complementares (DCG) que compõem o currículo do seu curso para a sua formação técnica, profissional e cidadã, são:	1,4454	-1,0497
Item 3	As atividades complementares (ACG) que compõem o currículo do seu curso para a sua formação técnica, profissional e cidadã, são:	1,6949	-1,1027
Item 4	A atuação do coordenador para buscar a constante melhoria do curso é:	1,6614	-1,5563
Item 5	A atuação do coordenador para atender às necessidades dos alunos é:	1,5539	-1,5158
Item 6	A carga horária de dedicação do coordenador à administração e à condução do curso no exercício de sua função é:	1,6430	-1,4432
Item 7	A carga horária das disciplinas do curso para atender os conteúdos programáticos é:	1,1284	-1,5423
Item 8	A atualização das disciplinas do curso no que diz respeito a ementas, conteúdos e bibliografias, você considera:	1,6015	-1,1472
Item 9	Nas condições de infra-estrutura, são considerados os aspectos ambientais (acústica, iluminação, ventilação, temperatura, etc.) bem como a disposição e adequação de instalações e equipamentos. Num contexto geral, como você avalia a sua unidade/subunidade:	1,2369	-0,3438
Item 10	As instalações acadêmicas (salas de aulas, laboratórios, anfiteatros, auditórios, diretório acadêmico) que existem em sua unidade universitária são:	1,7792	-0,4128
Item 11	Os recursos, equipamentos, informações, entre outros, existentes para a realização das atividades acadêmicas em sua unidade universitária são:	1,8909	-0,4856
Item 12	As condições de acesso para pessoas com necessidades especiais na unidade em que você atua são:	0,9399	-0,0112
Item 13	O acompanhamento das atividades de prática profissional e/ou estágio pela coordenação do curso é:	1,9486	-1,0537
Item 14	Quanto às formas de apresentação dos resultados da prática profissional e/ou estágio, você as considera:	1,7966	-1,4275
Item 15	A relação do número de alunos por orientador nas atividades de prática profissional e/ou estágio do curso é:	1,5219	-1,0799
Item 16	O funcionamento das atividades de prática profissional e/ou estágio internas à Instituição é:	1,9950	-1,0331
Item 17	O funcionamento das atividades de prática profissional e/ou estágio externas à Instituição é:	1,9868	-0,9932
Item 18	A carga horária de prática profissional e/ou estágio prevista para a sua formação é:	1,3890	-1,3404
Item 19	Quanto aos mecanismos de acompanhamento e de cumprimento do trabalho de conclusão de curso, você os considera:	2,0319	-1,1985
Item 20	A divulgação de trabalhos de conclusão de curso é:	1,8201	-0,9462
Item 21	De maneira geral, o relacionamento dos professores com os alunos do curso é:	1,3154	-2,2592
Item 22	O acompanhamento e a orientação, pela coordenação do curso, das Atividades Complementares previstas no curso são:	2,1868	-0,9753
Item 23	A oferta de Atividades Complementares de Graduação pelo curso é:	1,9794	-1,5777
Item 24	Quanto aos incentivos para a participação dos alunos em atividades complementares de graduação fora da UFSM, você os considera:	2,0651	-0,4403
Item 25	As experiências e os conhecimentos do corpo docente, em relação à proposta do curso, são:	2,0238	-1,6228

Item 26	A dedicação e o comprometimento do corpo docente ao curso para que se cumpram as atividades acadêmicas previstas são:	1,9398	-1,6005
Item 27	Os programas voltados para a assistência estudantil são:	1,7686	-0,8519
Item 28	O apoio para participar de eventos científicos, técnicos ou culturais é:	2,0258	-0,5990
Item 29	O acesso aos sistemas de rede, registros e arquivos na obtenção de informações acadêmicas é:	1,8541	-0,8074
Item 30	Os serviços prestados pelos técnico-administrativos vinculados ao seu curso são:	1,5780	-1,3312
Item 31	A quantidade de servidores técnico-administrativos do seu curso é:	1,3963	-0,8192
Item 32	O acervo de livros, quanto à quantidade e atualização, pertinência e relevância acadêmico-científica para as disciplinas do curso, é:	1,4386	-0,3417
Item 33	O acervo de periódicos, bases de dados específicas, jornais e revistas, quanto à quantidade e atualização, pertinência e relevância acadêmico-científica para as disciplinas do curso, é:	1,6487	-0,4092
Item 34	O sistema de acesso aos recursos bibliográficos (consulta e empréstimo) das Bibliotecas é:	1,1953	-1,4427
Item 35	A quantidade de ambientes/laboratórios para atender as necessidades de atividades práticas dos alunos matriculados no curso é:	2,2945	-0,1857
Item 36	A dimensão espacial, acústica, iluminação, ventilação, mobiliário e limpeza dos ambientes/laboratórios são:	1,9204	-0,3959
Item 37	A disponibilidade de equipamentos nos ambientes/laboratórios para atender às necessidades do curso é:	2,1576	-0,2150
Item 38	A orientação das atividades práticas desenvolvidas nos ambientes/laboratórios é:	2,2959	-0,9266
Item 39	A conservação das instalações hidráulicas, elétricas, eletrônicas e de telecomunicações dos ambientes/laboratórios é:	1,9465	-0,6588
Item 40	Os equipamentos de proteção individual e coletiva (EPI e EPC) nos ambientes/laboratórios são:	2,1560	-0,4365
Item 41	A organização e participação do Diretório Acadêmico nas atividades do curso são:	1,1249	-0,7580
Item 42	O apoio e incentivo da sua unidade de ensino à organização dos estudantes são:	1,9032	-0,6653
Item 43	Como você avalia a sua participação nas discussões e atividades inerentes ao seu curso de formação:	1,0042	-1,1965
Item 44	A divulgação dos resultados para a comunidade acadêmica sobre as avaliações anteriores é:	1,5754	-0,6577

Observa-se que os itens que compõem o questionário possuem boa discriminação, com parâmetro de discriminação estimado acima do valor um ($a > 1$), com exceção de um item que, embora tenha sido estimado abaixo do valor, ficou bem próximo dele (0,9399). Os valores estimados para o parâmetro de dificuldade mostram que todos os itens são considerados mais satisfatórios do que insatisfatórios, pois todos os itens apresentaram valores negativos ($b < 0$). Esse resultado corrobora com o resultado da TCT (Tabela 2), onde o parâmetro de dificuldade (proporção de satisfação do item) aponta os mesmos resultados. A Figura 3 apresenta as CCIs de cada item, onde é possível verificar a posição e a discriminação de cada um. Observa-se que quase todos os itens se apresentam dentro de um intervalo compreendido entre 0 e -2, com um poder semelhante de discriminação. Nessa Figura, dois itens se destacam dos demais: o item 21 (aquele posicionado mais à esquerda) e o item 12 (aquele posicionado mais à direita e com menor discriminação). O item 21 ($b = -2,2592$) é o item com maior facilidade de ser considerado satisfatório pelos alunos, ou seja, o relacionamento dos professores com os alunos é o item mais satisfatório do questionário. Por outro lado, o item 12 ($b = -0,0112$) tem maior dificuldade de ser considerado satisfatório pelos alunos, ou seja, os alunos estão mais divididos em considerar satisfatórias as condições de acesso para pessoas com necessidades especiais, esse é o item menos satisfatório.

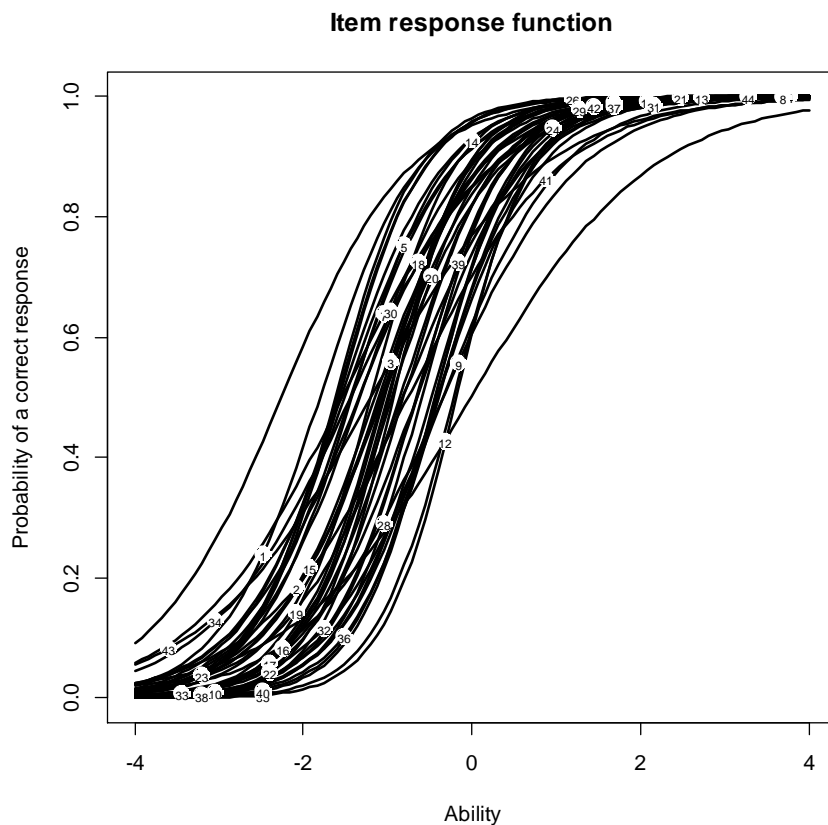


Figura 3 – Curva Característica de todos os Itens

A Figura 4 apresenta a FII, ou seja, a contribuição de cada item em termos de agregação de informação para o questionário. Pode-se observar claramente o quanto de informação que cada itens fornece e em que região do traço latente. Por exemplo, observa-se que o item 12 ($a = 0,9399$) oferece pouca informação onde o item está situado, em comparação com os demais itens, no entanto, esse é o item que fornece mais informação na região onde não há itens posicionados ($2 < b < 4$). A FII é uma contribuição da TRI que não possui equivalência direta na TCT.

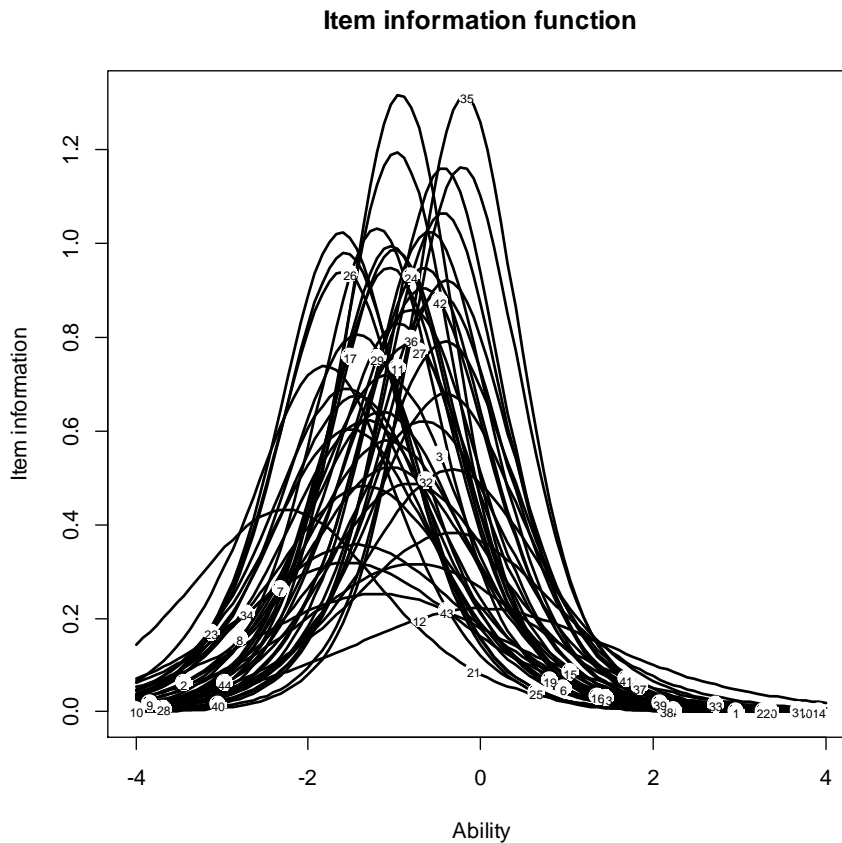


Figura 4 – Função de Informação de todos os Itens

A Figura 5 apresenta a FIT, ou seja, a contribuição que o conjunto de itens agrega fornece de informação para o questionário. Pode-se observar claramente que essa curva está posicionada para esquerda em relação à média da escala. Pode-se concluir que esse questionário consegue estimar com eficiência a satisfação de indivíduos que estão posicionados entre -3 e 1 na escala do traço latente, no entanto, a estimativa da satisfação dos alunos situados fora desse intervalo (principalmente entre 1 e 3) fica ineficiente com o uso desse questionário, ou seja, não há itens que meçam com precisão indivíduos muito satisfeitos. Por exemplo, ao observarmos a Figura 3, vemos que um indivíduo com traço latente igual a 2, tem alta probabilidade de estar satisfeito com todos os itens desse questionário. Dessa forma, não existe no questionário itens insatisfatórios, o que é um bom resultado em se tratando desse tipo de avaliação.

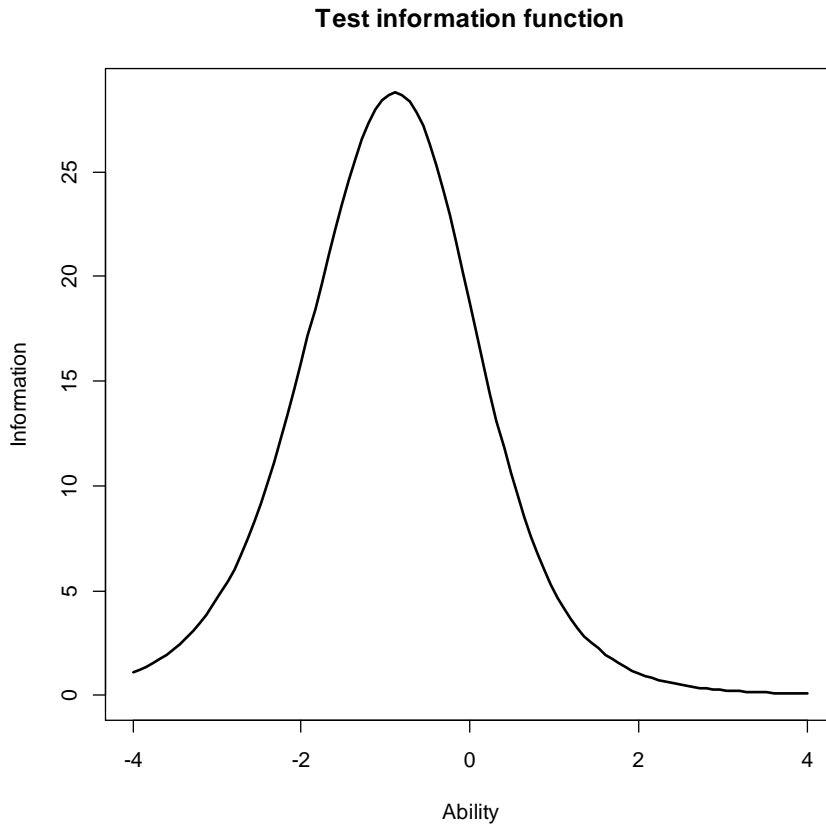


Figura 5 – Função de Informação do Teste

A Figura 6 apresenta a distribuição do valor traço latente estimado (nível de satisfação do aluno), por meio de um gráfico de Histograma. Observa-se que a maioria dos alunos situa-se entre -1,0 e 0,5 na escala, mas há uma grande quantidade deles nos intervalos entre -1,5 e -1,0, e 0,5 e 2. A distribuição assimétrica dos dados assemelha-se em parte à distribuição dos escores dos alunos na TCT (Figura 2). O intervalo entre 1,5 e 2,0 apresenta uma grande quantidade de alunos, no entanto, após esse intervalo, não há nenhum aluno situado. Isso não significa que não haja aluno situado após o valor 2,0, mas que o conjunto de itens não consegue dar uma estimativa maior que 2,0 para um aluno, mesmo que ele tenha respondido que está satisfeito com todos os itens do questionário, pois não há itens posicionados nessa região do traço latente. Isso corrobora com o que á foi dito anteriormente, ou seja, que o questionário não consegue medir com precisão indivíduos posicionados nessa região do traço latente.

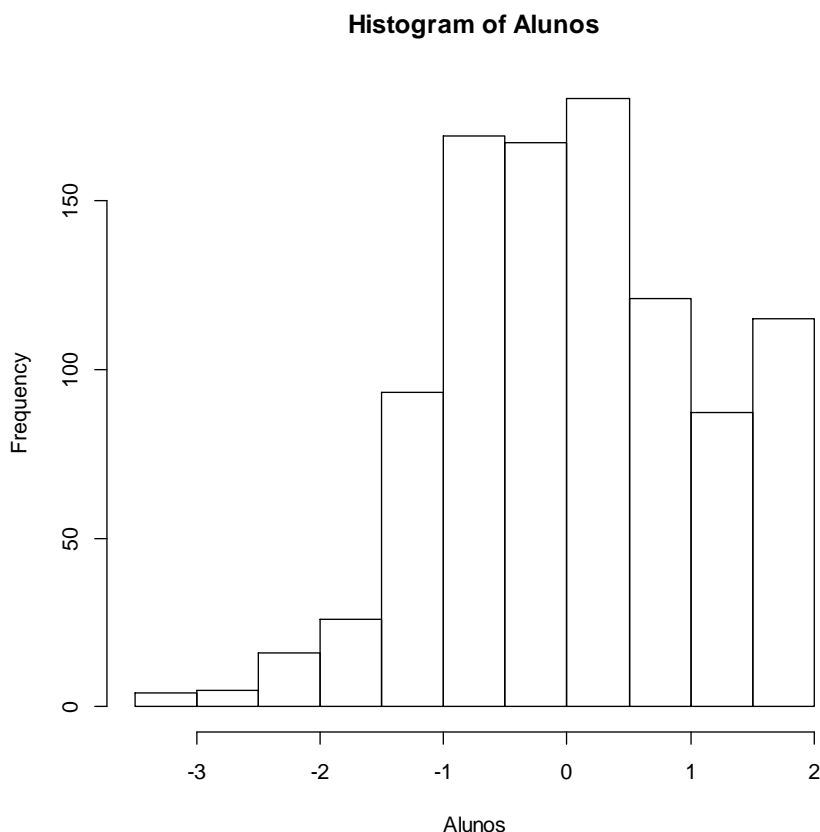


Figura 6 – Histograma do nível de satisfação dos alunos

Quanto à criação da escala, muitos itens foram considerados âncoras (25) e “quase âncoras” (14), no entanto, apenas três níveis âncoras foram caracterizados (Níveis -2, -1 e 0), já que os itens estão concentrados basicamente nessa região. Os níveis âncoras e as interpretações são dadas a seguir:

Nível -2 (1 Item): Indivíduos situados nesse nível ou acima, consideram satisfatório:

- O relacionamento dos professores com os alunos do curso, de maneira geral.

Nível -1 (17 Itens): Indivíduos situados nesse nível ou acima, consideram satisfatório:

- As disciplinas obrigatórias que compõem o currículo do seu curso para a sua formação técnica, profissional e cidadã;
- As experiências e os conhecimentos do corpo docente, em relação à proposta do curso;
- A dedicação e o comprometimento do corpo docente ao curso para que se cumpram as atividades acadêmicas previstas;
- A oferta de Atividades Complementares de Graduação pelo curso;
- A atuação do coordenador para buscar a constante melhoria do curso;
- A atuação do coordenador para atender às necessidades dos alunos;
- A carga horária de dedicação do coordenador à administração e à condução do curso no exercício de sua função;
- As formas de apresentação dos resultados da prática profissional e/ou estágio;
- A carga horária de prática profissional e/ou estágio prevista para a sua formação;
- Os serviços prestados pelos técnico-administrativos vinculados ao seu curso;

- Os mecanismos de acompanhamento e de cumprimento do trabalho de conclusão de curso;
- A atualização das disciplinas do curso no que diz respeito a ementas, conteúdos e bibliografias;
- As atividades complementares (ACG) que compõem o currículo do seu curso para a sua formação técnica, profissional e cidadã;
- A relação do número de alunos por orientador nas atividades de prática profissional e/ou estágio do curso;
- O acompanhamento das atividades de prática profissional e/ou estágio pela coordenação do curso;
- As disciplinas complementares (DCG) que compõem o currículo do seu curso para a sua formação técnica, profissional e cidadã;
- O funcionamento das atividades de prática profissional e/ou estágio internas à Instituição.

Nível 0 (21 Itens): Indivíduos situados nesse nível ou acima, consideram satisfatório:

- O funcionamento das atividades de prática profissional e/ou estágio externas à Instituição;
- O acompanhamento e a orientação, pela coordenação do curso, das Atividades Complementares previstas no curso;
- A divulgação de trabalhos de conclusão de curso;
- A orientação das atividades práticas desenvolvidas nos ambientes/laboratórios;
- Os programas voltados para a assistência estudantil;
- A quantidade de servidores técnico-administrativos do seu curso;
- O acesso aos sistemas de rede, registros e arquivos na obtenção de informações acadêmicas;
- A organização e participação do Diretório Acadêmico nas atividades do curso;
- O apoio e incentivo da sua unidade de ensino à organização dos estudantes;
- A conservação das instalações hidráulicas, elétricas, eletrônicas e de telecomunicações dos ambientes/laboratórios;
- A divulgação dos resultados para a comunidade acadêmica sobre as avaliações anteriores;
- O apoio para participar de eventos científicos, técnicos ou culturais;
- Os recursos, equipamentos, informações, entre outros, existentes para a realização das atividades acadêmicas em sua unidade universitária;
- Os incentivos para a participação dos alunos em atividades complementares de graduação fora da UFSM;
- Os equipamentos de proteção individual e coletiva (EPI e EPC) nos ambientes/laboratórios;
- As instalações acadêmicas (salas de aulas, laboratórios, anfiteatros, auditórios, diretório acadêmico) que existem em sua unidade universitária;
- O acervo de periódicos, bases de dados específicas, jornais e revistas, quanto à quantidade e atualização, pertinência e relevância acadêmico-científica para as disciplinas do curso;
- A dimensão espacial, acústica, iluminação, ventilação, mobiliário e limpeza dos ambientes/laboratórios;
- O acervo de livros, quanto à quantidade e atualização, pertinência e relevância acadêmico-científica para as disciplinas do curso;

- A disponibilidade de equipamentos nos ambientes/laboratórios para atender às necessidades do curso;
- A quantidade de ambientes/laboratórios para atender as necessidades de atividades práticas dos alunos matriculados no curso.

Considerações Finais

Esse trabalho teve como objetivo apresentar as vantagens e as oportunidades que a utilização da Teoria da Resposta ao Item (TRI) pode fornecer para complementar a análise que tradicionalmente é realizada autoavaliação institucional, por meio de uma análise inicial da TRI.

Todos os objetivos propostos foram alcançados. Segundo a análise realizada por meio do ML2, pode-se concluir que todos os itens são adequados, uma vez que possuem bom poder discriminatório e se posicionam dentro do intervalo adequado de variação. O questionário consegue estimar com eficiência o nível de satisfação de indivíduos que estão posicionados entre -3 e 1 na escala do traço latente, no entanto, a estimativa da satisfação dos alunos situados fora desse intervalo fica prejudicada, pois não há itens que meçam com precisão indivíduos muito satisfeitos, ou seja, o questionário carece de itens insatisfatórios, o que é um bom resultado em se tratando de avaliação de satisfação. O traço latente “nível de satisfação do aluno de graduação” foi estimado e os resultados mostram que a maior parte dos alunos situa-se entre -1,5 e 2 na escala. Quanto à criação da escala, muitos itens foram considerados âncoras (25) e “quase âncoras” (14), no entanto, apenas três níveis âncoras foram caracterizados: Nível -2 (1 Item), Nível -1 (17 Itens) e Nível 0 (21), já que os itens estão concentrados basicamente nessa região.

Como futuras análises, sugerem-se:

- A utilização de pacotes do *Software R* que reconheçam os dados de “não resposta”, a fim de aumentar o tamanho da amostra e verificar se há diferença nos resultados obtidos;
- A utilização de pacotes do *Software R* que utilizem Modelos Politômicos da TRI, com o propósito de modelar todas as categorias de respostas dos itens, sem a necessidade de dicotomizá-los para a análise.
- Utilizar métodos de equalização para criar uma escala única com todos os grupos de respondentes, por meio de itens comuns entre os diferentes questionários, com o objetivo de comparar a satisfação de indivíduos de diferentes grupos.

Referências

- 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.
- BAKER, F. B. **The Basics of Item Response Theory**. 2 ed. USA: ERIC Clearinghouse on Assessment and Evaluation, 2001.
- BEATON, A. E.; ALLEN, N. L. Interpreting Scales through Scale Anchoring. **Journal of Educational Statistics**, n. 17, p. 191-204, 1992.
- BIRNBAUM, A. **Some Latent Trait Models and Their Use in Inferring an Examinee's Ability**. In: LORD, F. M.; NOVICK, M. R. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley, 1968.
- BOCK, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. **Psychometrika**, v. 37, p. 29-51, 1972.
- COSTA, D. R. **Métodos Estatísticos em Testes Adaptativos Informatizados**. Dissertação. 2009. 120 f. (Mestrado em Estatística) – Departamento de Métodos

Estatísticos, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

CPA. **Relatório de Avaliação Interna: Autoavaliação da UFSM 2012**. Ministério da Educação. Universidade Federal de Santa Maria. Comissão Própria de Avaliação da UFSM, 2013. Disponível em <http://w3.ufsm.br/proplan/images/stories/file/COPLAI/UFSM_Relatorio_Autoavaliacao_2012.pdf>. Acesso em 27/06/2013.

DE AYALA, R. J. **The Theory and Practice of Item Response Theory**. New York, USA: The Guilford Press, 2008.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm (with discussion). **Journal of the Royal Statistical Society, Series B**, 39, 1-38, 1977.

EMBRETSON, S. E.; REISE, S. P. **Item Response Theory for Psychologists**. New Jersey, USA: Lawrence Erlbaum Associates, 2000.

FONTANIVE, N. S.; ELLIOT, L. G.; KLEIN, R. Os desafios da apresentação dos resultados da avaliação de sistemas escolares a diferentes públicos. **REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 5, n. 2e, 2007.

HAMBLETON, R. K.; SWAMINATHAN, H. **Item Response Theory: Principles and Applications**. Boston: Kluwer Nijoff; 1985.

HONTANGAS, P.; PONSODA, V.; OLEA, J. Procedimientos de integración numérica y estimación bayesiana en tests adaptativos informatizados. **VI Congreso de Metodología de las Ciencias Sociales y de la Salud**, Oviedo, Espanha, 1999.

KLEIN, R.; FONTANIVE, N. S.; ELLIOT, L. G. O Exame Nacional do Ensino Médio – Tecnologia e Principais Resultados Em 2005. **REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 5, n. 2e, p. 116-131, 2007.

KOLEN, M. J.; BRENNAN, R. L. **Test Equating - Methods and Practices**. New York, USA: Springer, 1995.

ISSAC, E; KELLER, H. B. **Analysis of Numerical Methods**. New York: Wiley & Sons, 1966.

MOREIRA JUNIOR, F. J. Aplicações da Teoria da Resposta ao Item (TRI) no Brasil. **Revista Brasileira de Biometria**, São Paulo, v.28, n.4 , p. 137-170, out.-dez. 2010.

MOREIRA JUNIOR, F. J. **Sistemática para a Implantação de Testes Adaptativos Informatizados baseados na Teoria da Resposta ao Item**. 2011. 334 f. Tese (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2011.

NUNES, C. H. S. S.; PRIMI, R. Impacto do Tamanho da Amostra na Calibração de Itens e Estimativa de Escores por Teoria de Resposta do Item. **Avaliação Psicológica**, v. 4, n. 2, p. 141-153, 2005.

PARTCHEV, I. **Package irtoys**: Simple interface to the estimation and plotting of IRT models, 2013. CRAN.R project, Disponível em <<http://cran.r-project.org/web/packages/irtoys/irtoys.pdf>> . Acesso em 16/04/2013.

RAO, C. R. **Linear Statistical Inference and Its Applications**. New York: Wiley & Sons, 1973.

RECKASE, M. D. **Multidimensional Item Response Theory. Statistical for Social and Behavioral Sciences**. New York: Springer Science+Business Media, LLC, 2009.

RIZOPOULOS, D. **Package ltm**: Latent Trait Models under IRT, 2013. CRAN.R project, Disponível em <<http://cran.r-project.org/web/packages/ltm/ltm.pdf>> . Acesso em 16/04/2013.

SOUZA, S. Z. 40 Anos de Contribuição à Avaliação Educacional. **Estudos em Avaliação Educacional**, v. 16, n. 31, jan./jun. 2005.

VALLE, R. C. A Construção e a Interpretação de Escalas de Conhecimento – Considerações Gerais e uma Visão do que vem sendo feito no SARESP. **Estudos em Avaliação Educacional**, n. 23, p. 71-92, 2001.

VENDRAMINI, C. M. M.; SILVA, M. C.; CANEL, M. Análise de Itens de uma Prova de Raciocínio Estatístico. **Psicologia em Estudo**, Maringá, v. 9, n. 3, p. 487-498, set./dez, 2004.