CIÊNCIA e NATURA

UFSM
Ci. e nat., Santa Maria, v. 42
Commemorative Edition: Statistic, e08, 2020
DOI: 10.5902/2179460X39777
ISSN 2179-460X

Statistics
Submissão 26/08/19 aprovação 28/08/19 publicação 27/10/20

Utilização da Teoria da Resposta ao Item na Análise do Instrumento da Autoavaliação Institucional na Pós-graduação da UFSM

Use of Item Response Theory in the Analysis of the Institutional Self-assessment Instrument in Graduate Program at UFSM

Nicásio Gouveia[I], Fernando De Jesus Moreira Junior[II], Angela Pellegrin Ansuj[III]

RESUMO

O objetivo desse trabalho é avaliar a qualidade dos itens do instrumento da Autoavaliação Institucional aplicado aos discentes dos Cursos de Pós-Graduação de todos os Campi da Universidade Federal de Santa Maria (UFSM), no período de 26 de Setembro a 31 de Outubro de 2016. Para a análise do instrumento utilizou-se a Teoria da Resposta ao Item (TRI), por meio do Modelo Logístico Unidimesional de dois Parâmetros (MLU2P). O estudo mostrou que todos os itens desse instrumento são adequados e consistentes, pois apresentam bom poder discriminatório e se posicionam dentro do intervalo esperado de variação. Percebeu-se que os itens que compõem esse instrumento são, em geral, fáceis de serem bem avaliados, segundo os resultados obtidos. Foi possível criar uma escala interpretável de medida com três níveis distintos. Concluiu-se que o instrumento utilizado é adequado para estimar os discentes que avaliam bem ou de forma razoável a instituição, porém o instrumento não consegue estimar com precisão os discentes que avaliam muito bem ou muito mal a instituição.

Palavras-chave: Teoria da Resposta ao Item Autoavaliação Institucional. Avaliação da Qualidade do Instrumento

ABSTRACT

The objective of this work is to evaluate the quality of the items of the Institutional Self-Assessment instrument applied to the students of the Graduate Courses of all Campuses of the Federal University of Santa Maria (UFSM), from September 26 to October 31, 2016. For the analysis of the instrument we used the Item Response Theory (IRT), through the Unidimesional Two Parameter Logistic Model (MLU2P). The study showed that all items of this instrument are adequate and consistent, as they have good discriminatory power and are within the expected range of variation. It was noticed that the items that make up this instrument are generally easy to be well evaluated, according to the results obtained. It was possible to create an interpretable scale of measurement with three distinct levels. It was concluded that the instrument used is adequate to estimate the students who evaluate the institution well or reasonably, but the instrument cannot accurately estimate the students who evaluate the institution very well or very poorly.

Keywords: Item Response Theory. Institutional Self-Assessment. Evaluation of Instrument Quality

[I] Universidade Federal de Santa Maria, Brasil; e-mail: nicasiogouveia@yahoo.com.br;
[II] Universidade Federal de Santa Maria. Brasil. e-mail: fmjunior@smail.ufsm.br;
[III] Universidade Federal de Santa Maria. Brasil. e-mail: angelaansuj@gmail.com

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

1 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. Desse modo, o processo de avaliação institucional tem como objetivo construir uma cultura permanente de avaliação, possibilitando que os resultados obtidos possam contribuir com a gestão das Universidades, permitindo que sejam implementadas ações a fim de alcançar os objetivos estratégicos estabelecidos no Plano de Desenvolvimento Institucional (CPA, 2016). O propósito do processo de Autoavaliação Institucional nas Instituições de Ensino Superior (IES) é proporcionar melhoria contínua da qualidade em todas as esferas, o que nos leva a refletir sobre a importância deste processo de Autoavaliação.

A Autoavaliação Institucional é algo importante às organizações, independentemente do ramo de atividade ou negócio. Consiste numa filosofia de gestão que se fundamenta na perspectiva da construção e da solidificação de cada instituição. Em vista disso, nenhuma instituição é plena de si, por isso, corrigir rumos, diagnosticar tendências, ouvir o público, identificar as limitações, assim como conhecer melhor a sua fortaleza e suas fraquezas, são algumas das razões para efetivar a importância da Autoavaliação Institucional. Em instituições de ensino, especialmente na educação superior, a Autoavaliação Institucional consiste além de uma ferramenta de gestão, numa perspectiva pedagógica, de aprendizado constante, com o corpo discente, docentes, técnico-administrativo, com a comunidade acadêmica e a sociedade em geral.

Existem vários métodos que permitem essa análise. Nesse contexto, um método que pode ser utilizado para analisar instrumentos de Autoavaliação é a Teoria de Resposta ao Item (TRI). A TRI é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da(s) habilidade(s) do respondente (ANDRADE; TAVARES; VALLE, 2000).

A TRI compreende modelos, para avaliar traços latentes (no caso, o nível de avaliação), que apresentam formas de representar a relação entre a probabilidade de um indivíduo dar uma resposta a um item e seu traço latente (ANDRADE; TAVARES; VALLE, 2000). Traços latentes (ou variáveis latentes) são características do indivíduo que não podem ser observadas diretamente, isto é, não existe um aparelho capaz de medi-las diretamente, como, por exemplo, um termômetro que mede diretamente a temperatura. São exemplos de traços latentes: nível de satisfação, conhecimento, habilidade, grau de depressão etc. Portanto, essas características são mensuradas através de variáveis secundárias que sejam relacionadas com o traço latente em estudo.

O traço latente é medido em uma escala criada através da aplicação da TRI. Essa escala pode ser criada com quaisquer valores de média e desvio padrão, entretanto é comum utilizar uma escala (0, 1), isto é, com média igual a zero e desvio padrão igual a um (MOREIRA JUNIOR; ZANELLA; DIAS LOPES; SEIDEL, 2015). Assim, este trabalho tem por objetivo analisar o instrumento da Autoavaliação Institucional aplicado aos discentes dos Cursos de Pós-Graduação da Universidade Federal de Santa Maria (UFSM), por meio da Teoria da Resposta ao Item (TRI). A utilização dessa ferramenta possibilita colocar os itens e os respondentes simultaneamente na mesma escala de medida, o que facilita a sua interpretação.

2 REVISÃO DA LITERATURA

2.1 Histórico do processo de Avaliação Institucional na UFSM

A UFSM sempre manifestou preocupação em avaliar suas atividades acadêmicas por meio de iniciativas pontuais e setorizadas de avaliação porque, até o final da década de 80, não existia no Brasil um processo de avaliação global para as universidades (CPA, 2008). A partir da década de 80, a UFSM vem procurando adotar um Sistema de Avaliação de Desenvolvimento por meio de alguns projetos (CPA, 2008), como:

- Diagnóstico do desempenho da Universidade Federal de Santa Maria (UFSM), 1980;
- Proposta de avaliação da Universidade Federal de Santa Maria (UFSM), 1986;
- Indicadores de Desempenho dos Departamentos da UFSM, 1989;
- Avaliação do Desempenho Docente com vistas a pagamento da Gratificação de Estímulo à Docência na UFSM – GED, 1998, o que proporcionou a universidade uma ampla discussão das atividades docentes, contribuindo para o aumento da qualificação docente, publicações, projetos e uma melhoria na qualidade de ensino;
- E, por último, Avaliação do Desempenho Docente pelo Discente, sendo essa atualmente uma prática permanente na Universidade.

Em 1993, foi instituído o Programa de Avaliação Institucional das Universidades Brasileiras, por meio da Comissão Nacional de Avaliação do Ministério da Educação e do Desporto, vindo ao encontro dos propósitos da administração da UFSM (CPA, 2008). Assim sendo, esses propósitos visam buscar a qualidade dos serviços oferecidos pela instituição à sociedade, refletidas pelo patrimônio ensino-pesquisa-extensão (CPA, 2008).

2.2 Autoavaliação

A avaliação pode ser compreendida como uma apreciação da qualidade e da eficácia do sistema de ensino como um todo ou parte dele, e deve ocasionar, no discente, uma reflexão sobre o que ele vem vivenciando durante as situações de aprendizagem. A avaliação não consiste em um instrumento de medida de atividades de indivíduos isolados e nem em um mecanismo de exposição da fragilidade ou das deficiências de alguns profissionais específicos. Deve ser compreendida como um processo de caráter pedagógico e de construção de valores, não se limitando à elaboração de relatórios e diagnósticos, e ao julgamento de resultados e ações já realizadas. Avaliação é um processo que deve fazer parte da instituição, e realizar-se como cultura (BALZAN; DIAS SOBRINHO, 2005). Para Griboski, Peixoto e Horta (2018),

a principal função da avaliação é auxiliar mudanças e práticas inovadoras no intuito de diagnosticar, apoiar e corrigir os aspectos avaliados.

Segundo Gatti (2006), a Avaliação Institucional nas Universidades consiste em um levantamento de dados na modalidade survey, com um conjunto de questões sobre categorias consideradas importantes no desempenho de uma Universidade. Dessa forma, deve-se salientar a importância da Autovaliação Institucional para o acompanhamento dos processos de gestão nas instituições. A Autovaliação Institucional deve ter a avaliação formativa como pressuposto, porque ela possibilita informações acerca do desenvolvimento do processo de ensino, com o intuito de direcionar a prática pedagógica dos educadores, quanto ao desenvolvimento das ações institucionais, estimular políticas acadêmicas em andamento (GRIBOSKI; PEIXOTO; HORTA, 2018). A Autoavaliação Institucional na Universidade Federal de Santa Maria (UFSM) passou pelas etapas a seguir descritas:

> "O processo de avaliação na UFSM teve início muito antes de ser demandado às universidades como uma exigência legal, o que ocorreu apenas em 2004, com a edição da Lei N. 10.861 de 14 de abril de 2004. O processo de avaliação na UFSM foi oficializado pela emissão da Resolução N. 008 de 23/09/2004, que regulamentou, no âmbito da UFSM, a estrutura e o funcionamento da Comissão Própria de Avaliação (CPA). Esta resolução teve como objetivo normatizar a organização e a condução dos processos internos de avaliação, oferecendo o apoio necessário à avaliação externa, por conta dos processos de autorização, reconhecimento e renovação de reconhecimento de cursos. A referida resolução se manteve válida até maio de 2015, quando foi revogada, pela resolução N. 009/2015 que instituiu o Regimento Interno da Comissão Própria de Avaliação em 26 de maio de 2015. Hoje, a CPA da UFSM é designada pela Portaria nº 78.872 de 23/03/2016, composto por 29 membros, entre eles, docentes, técnico-administrativos, discentes, servidores aposentados e sociedade civil. Atualmente, o processo de autoavaliação é coordenado pela CPA e desenvolvido em colaboração com quinze Comissões Setoriais de Avaliação (CSA), representando cada unidade universitária. Além disso, conta com assessoria da Coordenação de Planejamento e Avaliação Institucional (COPLAI) vinculada a Pró-Reitoria de Planejamento (PROPLAN)". (CPA, 2016, p. 18).

Segundo o Projeto Pedagógico Institucional (CPA, 2016), a Autoavaliação na UFSM tem como objetivo principal detectar seus pontos fortes e fracos para melhorar a qualidade do ensino da Graduação e Pós-Graduação, da pesquisa, da extensão e da gestão em cada setor da UFSM.

2.2 Teoria da Resposta ao Item (TRI)

A abordagem da TRI, tem sido utilizada em muitas avaliações educacionais do ensino. A TRI é um conjunto de modelos matemáticos que define uma maneira de estabelecer uma relação entre variáveis latentes e suas manifestações, possibilitando a criação de medidas padronizadas. A TRI pode ser descrita como uma teoria baseada em inferências estatísticas, na qual características latentes de indivíduos ou sistemas são estimadas, tendo como base as respostas destes a um determinado conjunto de itens (AYALA, 2009).

Moreira Junior e Szinvelski (2013) utilizaram a TRI para avaliar a qualidade dos itens que compõem o instrumento de Autoavaliação Institucional aplicado aos cursos de graduação, que em geral possuíam um bom poder discriminatório. 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.

A TRI é uma ferramenta estatística que surgiu para suprir as necessidades decorrentes das limitações da Teoria Clássica da Medida (TCM) - também chamada de Teoria Clássica do Teste (TCT), - teoria que tradicionalmente era - e ainda é - utilizada nas avaliações. A avaliação tradicional é baseada na TCT, no qual, o traço latente é medido através de um somatório ponderado, ou não, dos pontos atribuídos à cada questão do instrumento. A TCT também utiliza outros indicadores, tais como: o índice de dificuldade, o Coeficiente de Correlação Ponto-Bisserial e o Alfa de Cronbach.

As principais contribuições da TRI são os mecanismos de equalização e a criação de escalas interpretáveis. Os mecanismos de equalização foram um grande progresso para a avaliação educacional, uma vez que permite que indivíduos avaliados por instrumentos de avaliação parcialmente diferentes (com alguns itens em comum) ou totalmente diferentes (em determinados casos) 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 grande 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, por exemplo, 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 a probabilidade do acerto casual (ANDRADE; TAVARES; VALLE, 2000).

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

Atualmente, a TRI vem sendo bastante difundida no mundo todo, principalmente na área de educação e testes psicológicos. Uma relação de trabalhos sobre TRI publicados no Brasil até o ano de 2009 encontra-se disponível em Moreira Junior (2010).

O uso da TRI vem se destacando em várias pesquisas de satisfação, assim como na Autoavaliação Institucional. Nascimento et al (2016) usaram a TRI para avaliação Discente em Ciências Contábeis e concluiuram ser um imprescindível subsídio para o constante aprimoramento de todo o processo ensino-aprendizagem em contabilidade.

Para a estimação dos parâmetros dos itens (procedimento conhecido como “calibração dos itens”) e para a estimação do traço latente, há vários métodos estatísticos que podem ser utilizados. Os mais conhecidos são: o método de Máxima

Verossimilhança Marginal (MVM), o método bayesiano da Moda a Posteriori (MAP), o método bayesiano da Esperança a Posteriori (EAP) e o método da Máxima Verossimilhança Conjunta (MVC). No entanto, 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, consequentemente, necessitam de programas computacionais específicos.

Existem vários modelos matemáticos utilizados na TRI, diferentes quanto à sua natureza, 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).

3 MATERIAL E MÉTODO

Para a realização deste trabalho, foi utilizado o banco de dados disponibilizado pelo Centro de Processamento de Dados (CPD) da Universidade Federal de Santa Maria (UFSM), No total, participaram da Autoavalição 1.290 discentes dos Cursos de Pós-Graduação dos diversos Campi da UFSM, no período de 26 de setembro a 31 de outubro de 2016 (CPA, 2016). O instrumento foi disponibilizado em duas partes (Questões Gerais e Discente de Pós-Graduação) no site da UFSM, composto ao todo por 44 questões, sendo que 34 são de caráter objetivo e 10 de caráter subjetivo, conforme instrumento em anexo.

Inicialmente, para a análise dos dados foi utilizada a Teoria Clássica dos Testes (TCT), onde foi calculada a proporção de avaliação positiva e não positiva, o Coeficientes de Correlação Ponto Bisserial com e sem item e o Alfa de Crombach sem item. Após utilizou-se a Teoria de Resposta ao Item (TRI), para avaliar os itens do instrumento da Avaliação Institucional. A TRI necessita de uma amostra grande para fornecer resultados consistentes, assim como só é possível analisar questões objetivas, portanto as questões subjetivas foram ignoradas. Todas as 34 questões

objetivas solicitavam a avaliação do discente em relação a diversos itens e possuíam as mesmas categorias de respostas, com variações de gênero e número: (1) péssimo(a)(os)(as), (2) ruim/ruins, (3) regular(es), (4) bom/bons/boa(s) e (5) excelente(s). A tabela 1 apresenta as 34 questões que compõe o instrumento.

Quadro 1 Questões do Instrumento da Autovaliação Institucional da Pós-Graduação da Universidade Federal de Santa Maria

ITEM	DESCRIÇÃO
1	Como você avalia a divulgação dos resultados da Pesquisa de Autoavaliação Institucional na sua unidade/subunidade?
2	Como você avalia a utilização dos resultados da Pesquisa de Autoavaliação Institucional como subsídio à revisão, proposição e implementação de ações na sua unidade/subunidade?
3	Como você avalia as ações voltadas para a inclusão social promovidas pela Instituição?
4	Como você avalia a acessibilidade da unidade/subunidade em que você atua?
5	Como você avalia a gestão ambiental da Instituição?
6	A missão da UFSM é construir e difundir conhecimento, comprometida com a formação de pessoas capazes de inovar e contribuir com o desenvolvimento da sociedade, de modo sustentável. Como você avalia a atuação da Instituição no cumprimento da sua missão?
7	Como você avalia os canais de comunicação interna entre os diversos setores da Instituição?
8	Como você avalia os canais de comunicação da Instituição com a comunidade externa?
9	Como você avalia o Ambiente Virtual de Ensino e Aprendizagem (AVEA) oferecido pela Instituição?
10	Como você avalia a capacitação e a qualificação dos servidores docentes e técnico-administrativos em educação nos diversos setores administrativos da Instituição?
11	Como você avalia o acesso e a funcionalidade dos portais institucionais (Portal do RH, Portal do Aluno, Portal do Professor, entre outros)?
12	Como você avalia a disponibilidade e a atualização do acervo das bibliotecas que você utiliza?
13	Como você avalia os serviços terceirizados disponibilizados na Instituição (recepção, limpeza, manutenção e segurança)?
14	Como você avalia a prestação de serviços privados disponíveis na Instituição (restaurantes/lancherias e reprografias)?
15	Como você avalia os horários de atendimento dos serviços privados disponíveis à comunidade (transporte coletivo, restaurantes/lancherias e reprografias)?
16	Como você avalia os horários de atendimento dos serviços internos disponíveis à comunidade universitária (restaurante universitário, bibliotecas e setores administrativos)?
17	Como você avalia as condições de infraestrutura da Instituição?
18	Como você avalia as condições de infraestrutura de sua unidade/subunidade?
19	Como você avalia a disponibilidade, conservação e acesso aos equipamentos na Instituição?
20	Como você avalia a disponibilidade e o acesso à internet e à intranet na Instituição?
21	Como você avalia as condições das instalações sanitárias que você utiliza?
22	Como você avalia a disponibilidade e a conservação dos espaços de convivência que você utiliza?
23	Como você avalia a contribuição dos planos de ensino para a sua formação acadêmica, no que diz respeito à atualização, conteúdos e bibliografias das disciplinas?
24	Como você avalia o acompanhamento, cumprimento e divulgação dos trabalhos de conclusão de curso?

25	Como você avalia a divulgação e o apoio da coordenação do curso para a sua participação em eventos científicos, técnicos ou culturais?
26	Como você avalia a orientação e o acompanhamento das atividades práticas desenvolvidas nos laboratórios?
27	Como você avalia o acesso e o treinamento oferecidos pela Instituição para busca bibliográfica em portais de periódicos?
28	Como você avalia os critérios adotados para a concessão de bolsas?
29	Como você avalia o corpo docente quanto à sua experiência, conhecimento, dedicação e comprometimento, em relação à proposta do curso?
30	Como você avalia a atuação do diretor do seu centro de ensino/campus, considerando a melhoria contínua e o atendimento às necessidades dos estudantes.
31	Como você avalia a atuação do coordenador do seu curso, considerando a melhoria contínua e o atendimento às necessidades dos estudantes?
32	Como você avalia o funcionamento e o atendimento da secretaria do seu curso, considerando a melhoria contínua e o atendimento às necessidades dos estudantes?
33	Como você avalia a atuação do seu orientador em relação ao acompanhamento e orientação do seu projeto de pós-graduação?
34	Como você avalia as condições oferecidas para a realização das pesquisas necessárias ao seu embasamento acadêmico-profissional?

Fonte: Elaborado pelo autor

Para realizar a análise da TRI é necessário ter: (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 da TRI usualmente segue o procedimento descrito a seguir.

a) 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 juízes, quando for conveniente.
b) O conjunto de itens resultantes irá compor 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.
c) 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.

d) 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.

Nesse trabalho será dada ênfase ao Modelo Logístico Unidimensional de Dois Parâmetros (MLU2P), que é um modelo unidimensional, dicotômico e para uma única população.

O MLU2P, desenvolvido por Birnbaum (1968), é representado pela seguinte equação:

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

onde,

$P(U_{ij} = 1/\theta_j)$ é a probabilidade do respondente j, condicionado no seu traço latente θj, responder ou avaliar positivamente ao item i;

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 na escala onde há 0,50 de probabilidade do indivíduo *j* (no caso, o discente *j*) avaliar positivamente ao item i;

U_{ij} é a resposta ao item, que pode ser positiva (U_{ij}=1), no caso, o discente avaliar positivamente o item i, ou negativa (U_{ij}=0), caso contrário;

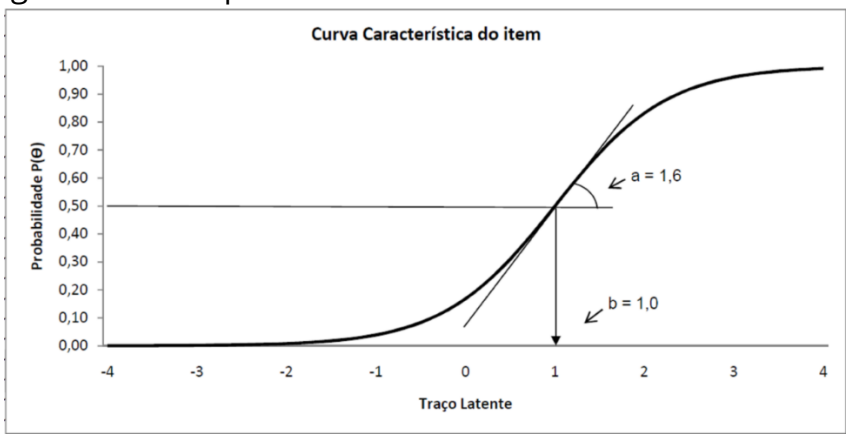
Θ_j representa o valor do traço latente (no caso, avaliação da UFSM pelo discente de Pós-Graduação da UFSM).

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 MLU2P 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 – Exemplo de uma Curva Característica de um Item



Fonte: Moreira Junior, Tezza e Bornia (2011).

No caso de medição do grau de avaliação da qualidade do instrumento da Pós-Graduação da UFSM pelo discente, o eixo x, da Figura 1, representa o grau de avaliação da qualidade do instrumento no qual é possível posicionar os itens e os discentes, o que possibilita verificar quais os itens que são bem avaliados para determinado discente. 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:

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

onde:

$I_i(\theta)$ é a "informação" fornecida pelo item *i* no nível de habilidade θ;

a_i^2 é o quadrado do parâmetro de discriminação do item *i*,

$P_i(\theta)$ é a proporção de discentes que avaliaram positivamente o item *i* no nível de habilidade θ;

$Q_i(\theta)$ é a proporção de discentes que avaliaram negativamente o item *i* no nível de habilidade θ.

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

$$I(\theta) = \sum_{i=1}^{I} I_i(\theta) \qquad (3)$$

Segundo Andrade, Tavares e Valle (2000), os valores da escala do traço latente (ou habilidade) podem assumir teoricamente qualquer valor real entre -∞ e +∞, 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 maiores que 0,65 (BAKER, 2001). Após essa etapa, é realizada a construção da escala do traço latente, que é colocada na mesma métrica dos itens. Conforme, Klein, Fontanive e Elliot (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 se situam em um determinado nível da escala são capazes de demonstrar as habilidades descritas nesse nível e nos níveis anteriores dessa escala.

Todavia, na realidade, 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 matemáticas/estatísticas abaixo forem satisfeitas simultaneamente (U representa a resposta ao item):

1. $P(U=1|\theta=Z)\geq 0{,}65$, (4)
2. $P(U=1|\theta=Y)<0{,}50$ e (5)
3. $P(U=1|\theta=Z)-P(U=1|\theta=Y)\geq 0{,}30$. (6)

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 da identificação dos 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). Dessa forma, um sujeito que se situa acima de um determinado nível dominará os conteúdos caracterizados por esse nível e pelos níveis inferiores a esse, porém não dominará os assuntos dos níveis superiores a esse. Para análise dos dados, foi utilizado o *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. Porém, foi escolhido o pacote *irtoys* (PARTCHEV, 2017) pelo fato de trabalhar com Modelos dicotômicos e apresentar boas opções de estimação e de análises gráficas. A análise por meio da TCT foi feita com o pacote *ltm* (RIZOPOULOS, 2018). No pacote *irtoys*, a estimação dos parâmetros dos itens é feita pelo pacote *ltm,* o qual utiliza o método da Máxima Verossimilhança Marginal. Por outro lado, para estimar os parâmetros dos indivíduos, o pacote *irtoys* oferece quatro métodos. Dentre esses, foi escolhido o método bayesiano da Esperança a Posteriori (EAP) por ser um método que obtém o menor erro padrão e a maior fidelidade

(HONTANGAS; PONSODA; OLEA, 1999), além do mais, 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 recomendado.

Para utilizar o Modelo Logístico Unidimensional de Dois Parâmetros (MLU2P), as respostas referentes às questões foram dicotomizadas. O processo da dicotomização foi realizada da seguinte forma: as alternativas (1) péssimo(a)(os)(as), (2) ruim/ruins e (3) regular(es) foram consideradas como uma única categoria denominada (0) avaliação não positiva; enquanto que as alternativas (4) bom/bons/boa(s) e (5) excelente(s) foram consideradas como uma única categoria denominada (1) avaliação positiva. Várias questões ainda possuíam a alternativa "Não sei responder/Não se aplica", que foi considerada como "não resposta" na presente análise. Além disso, nem todos os discentes que responderam o instrumento das questões gerais responderam também as questões específicas do instrumento da Pós-Graduação. Para essa análise foram considerados somente os discentes que responderam ambos os instrumentos. Diante disso, a análise foi realizada com uma amostra de 892 discentes, que é considerado uma amostra suficientemente grande para uma análise da TRI como o modelo proposto (NUNES; PRIMI; 2005; EMBRETSON; REISE, 2000; HAMBLETON; SWAMINATHAM, 1985).

4 RESULTADOS

A Tabela 1 apresenta a quantidade e o percentual dos discentes analisados com relação à modalidade do curso de Pós-Graduação.

Tabela 1 Caracterização da Amostra Segundo a Modalidade do Curso

Modalidade do Curso	Quantidade	Percentual
Doutorado	353	39,57%
Especialização	44	4,93%
Mestrado	478	53,59%
Pós-Doutorado	17	1,91%
Total	892	100,00%

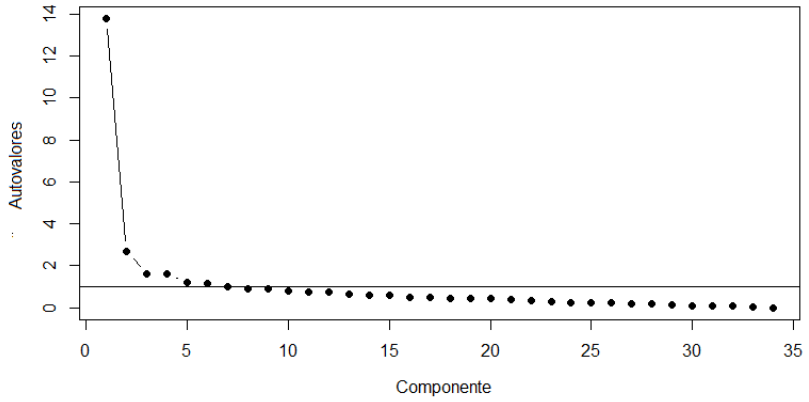
Fonte: Elaborado pelos autores

Logo após a formatação dos dados (dicotomização e codificação), foi realizada uma análise tradicional baseada na Teoria Clássica dos Testes (TCT) com recursos

concedidos pelo pacote *ltm* do R. A Tabela 1 apresenta a proporção de respostas "de avaliação não positiva", a proporção de "avaliação positiva", o Coeficiente do Correlação Ponto-Bisserial que inclui o item (CCPBc/I), o Coeficiente do Correlação Ponto-Bisserial que desconsidera o item (CCPBs/I) e o Alfa de Cronbach que desconsidera o item. Percebe-se que o Item 20 foi o que obteve a menor proporção de avaliação positiva (0,2875), por ser o mais difícil de ser avaliado positivamente pelo discente, enquanto que o item 11 foi o que obteve a maior proporção de avaliação positiva (0,8939). Os Coeficientes da Correlação Ponto-Bisserial (incluindo ou não o item) que calculam 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, em geral, não eram incoerentes em relação ao instrumento. O Alfa de Cronbach incluindo todos os itens foi de 0,9180, praticamente o mesmo valor quando se excluía algum item (Tabela 2), o que mostra claramente que o instrumento é consistente.

Para verificar a suposição da unidimensionalidade, foi realizada a Análise Fatorial para dados qualitativos baseada na matriz de correlações tetracóricas. Na TRI, para satisfazer a suposição da unidimensionalidade, é considerável reconhecer que haja uma habilidade dominante (um fator dominante) responsável pelo conjunto de itens. (ANDRADE; TAVARES; VALLE, 2000). O gráfico da Figura 2 mostra o *Scree Plot,* resultado gráfico da Análise Fatorial, com base nos autovalores.

Figura 2 – Scree Plot.



Fonte: Elaborado pelos autores

Pode-se observar graficamente que o primeiro fator se destacou em relação aos demais, evidenciando que existe um fator dominante, que é suficiente para verificar a suposição da unidimensionalidade para fins de análise com modelos unidimensionais da TRI.

Tabela 2: Análise Clássica dos Itens

ITEM	Proporção de avaliação não positiva	Proporção de avaliação positiva	CCPBc/I	CCPBs/I	Alfa de Cronbach s/I
Item 1	0,4798	0,5202	0,5530	0,5090	0,9154
Item 2	0,4897	0,5103	0,5020	0,4546	0,9168
Item 3	0,3379	0,6621	0,4963	0,4515	0,9184
Item 4	0,4964	0,5036	0,4315	0,3805	0,9171
Item 5	0,5032	0,4968	0,5447	0,5002	0,9150
Item 6	0,3576	0,6424	0,5160	0,4724	0,9160
Item 7	0,4443	0,5557	0,5340	0,4895	0,9157
Item 8	0,5310	0,4690	0,5524	0,5081	0,9143
Item 9	0,2649	0,7351	0,5123	0,4702	0,9165
Item 10	0,2890	0,7110	0,5792	0,5412	0,9148
Item 11	0,1061	0,8939	0,3858	0,3538	0,9172
Item 12	0,3173	0,6827	0,5315	0,4904	0,9158
Item 13	0,3984	0,6016	0,5551	0,5119	0,9153
Item 14	0,6568	0,3432	0,5846	0,5438	0,9145
Item 15	0,6186	0,3814	0,5450	0,5012	0,9153
Item 16	0,2916	0,7084	0,5215	0,4801	0,9158
Item 17	0,3131	0,6869	0,5815	0,5427	0,9149
Item 18	0,3833	0,6167	0,5998	0,5601	0,9146
Item 19	0,4920	0,5080	0,6093	0,5691	0,9145
Item 20	0,7125	0,2875	0,5159	0,4720	0,9155
Item 21	0,6171	0,3829	0,4774	0,4296	0,9163
Item 22	0,4479	0,5521	0,5786	0,5366	0,9146
Item 23	0,2597	0,7403	0,4910	0,4506	0,9165
Item 24	0,3983	0,6017	0,5769	0,5363	0,9151
Item 25	0,4707	0,5293	0,6143	0,5745	0,9138
Item 26	0,3324	0,6676	0,5304	0,4879	0,9137
Item 27	0,4156	0,5844	0,5524	0,5090	0,9148
Item 28	0,5181	0,4819	0,4002	0,3473	0,9173
Item 29	0,1917	0,8083	0,5124	0,4771	0,9160
Item 30	0,3017	0,6983	0,5431	0,5017	0,9152
Item 31	0,2016	0,7984	0,4909	0,4546	0,9163
Item 32	0,2269	0,7731	0,4285	0,3868	0,9170
Item 33	0,1146	0,8854	0,3257	0,2951	0,9181
Item 34	0,3185	0,6815	0,5410	0,4994	0,9155

Fonte: Elaborado pelos autores

Diante disso, passou-se à análise dos dados por meio da TRI. A Tabela 3 apresenta a descrição dos itens e os valores estimados dos parâmetros de discriminação (*a*) e de dificuldade (*b*) dos itens com respectivos erros padrão, por meio do MLU2P em uma escala com média zero (0) e desvio padrão um (1).

Tabela 3: Estimativas dos Parâmetros de Discriminação (a) e de Dificuldade (b) dos itens com respectivo erro padrão (EP)

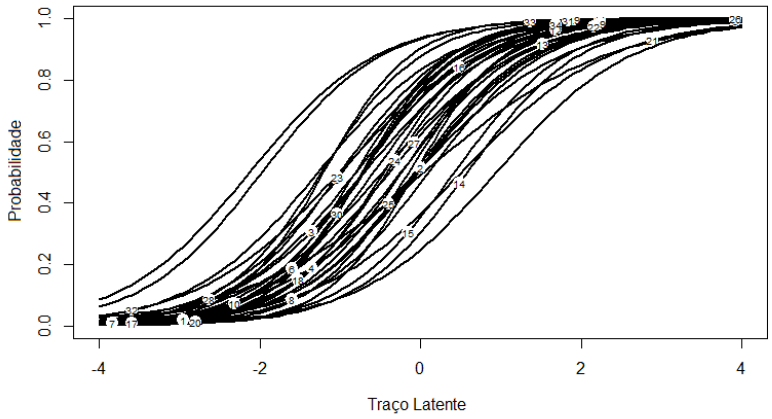
ITEM	*a*	EP(*a*)	*b*	EP(*b*)
1	1,40	0,13	-0,08	0,07
2	1,38	0,13	-0,05	0,07
3	1,24	0,13	-0,70	0,09
4	1,10	0,11	-0,05	0,08
5	1,35	0,12	-0,02	0,07
6	1,45	0,13	-0,59	0,07
7	1,35	0,12	-0,24	0,07
8	1,40	0,13	0,10	0,07
9	1,32	0,15	-0,98	0,11
10	1,68	0,15	-0,81	0,07
11	1,27	0,16	-2,13	0,20
12	1,30	0,12	-0,81	0,08
13	1,19	0,11	-0,48	0,08
14	1,43	0,13	0,58	0,08
15	1,43	0,13	0,42	0,07
16	1,13	0,12	-1,01	0,10
17	1,66	0,15	-0,73	0,07
18	1,64	0,14	-0,46	0,06
19	1,81	0,15	-0,06	0,06
20	1,18	0,11	0,94	0,10
21	1,08	0,10	0,51	0,09
22	1,51	0,13	-0,24	0,06
23	1,46	0,14	-1,00	0,09
24	1,54	0,14	-0,42	0,06
25	1,66	0,14	-0,15	0,06
26	1,47	0,14	-0,73	0,08
27	1,28	0,12	-0,38	0,07
28	0,89	0,10	0,04	0,09
29	1,84	0,18	-1,21	0,08
30	1,79	0,16	-0,73	0,07
31	1,60	0,16	-1,23	0,09
32	1,26	0,13	-1,28	0,11
33	1,35	0,16	-1,97	0,18
34	1,64	0,15	-0,71	0,07

Fonte: Elaborado pelos autores.

Percebe-se que os itens que compõem o instrumento apresentaram boa discriminação, com parâmetro estimado acima do valor um (a > 1), com exceção do item 28, que apesar de que tenha sido estimado abaixo o valor, ficou bem próximo (0,89). A estimação dos valores para o parâmetro de dificuldade mostra que a maioria dos itens foram avaliados positivamente, pois todos os outros itens apresentaram valores negativos (b < 0), com exceção dos itens (8, 14, 15, 20, 21 e 28), que apresentaram valores positivos (b > 0), isto é, eles são mais difíceis de serem avaliados positivamente.

A Figura 3 apresenta as CCIs de cada item, onde é possível visualizar a posição e a discriminação de cada um, além da probabilidade do item ser avaliado positivamente. Percebe-se que praticamente todos os itens se apresentam dentro de intervalo compreendido entre -2 e 1, com um poder similar/parecido a discriminação. Nesse seguimento, dois itens se destacaram na Figura 3 aos demais: o item 11 (aquele posicionado mais à esquerda) e o item 20 (aquele posicionado mais à direita e com menor discriminação). O item 11 (b = -2,13) é o item com maior facilidade de ser avaliado positivamente pelos discentes, ou seja, o acesso e a funcionalidade dos portais institucionais (Portal do RH, Portal do Aluno, Portal do Professor, entre outros) é o item mais bem avaliado do instrumento. Em contrapartida, o item 20 (b = 0,94) tem maior dificuldade de ser avaliado positivamente pelos discentes, ou seja, os discentes avaliaram mais negativamente a disponibilidade e o acesso à internet e à intranet na instituição.

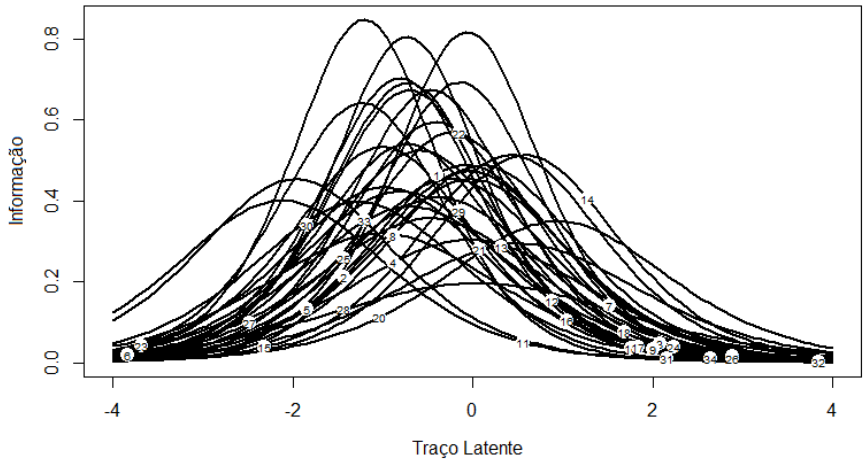
Figura 3 – Curva Característica de todos os Itens.



Fonte: Elaborado pelos autores.

A Figura 4 apresenta a Função de Informação dos Itens (FII), ou seja, a contribuição de cada item em termos de agregação de informação para o instrumento. Pode-se observar graficamente o quanto de informação que cada item fornece e em que região do traço latente. Por exemplo, pode-se observar que o item 28 (a = 0,89) oferece pouca informação em relação aos demais itens. O item 29 (a = 1,84) é o item que fornece mais informação na região onde não há itens posicionados ($2 < b < 4$).

Figura 4 – Função de Informação de todos os Itens.

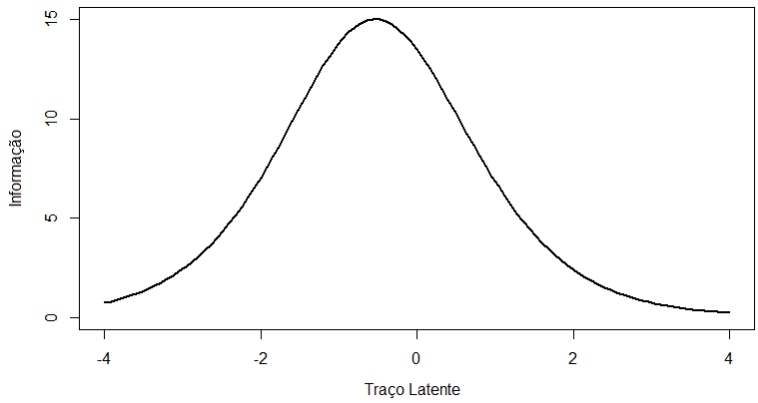


Fonte: Elaborado pelos autores.

A Figura 5 apresenta a Função de Informação do Teste (FIT), ou seja, a contribuição que o conjunto de itens agrega e fornece de informação para o instrumento de avaliação. Pode-se observar graficamente que essa curva está posicionada ligeiramente para esquerda em relação à média da escala, o que significa que, em geral, os itens que compõem esse instrumento são fáceis de serem bem avaliados, segundo os resultados obtidos, o que, de fato, é um bom resultado quando se trata de Autoavaliação Institucional. A FIT mostrou ainda, que o instrumento conseguiu estimar com positivamente os itens do instrumento da autoavaliação dos discentes da Pós-Graduação da UFSM posicionando-os entre -2 e 1 na escala do traço latente, região onde há mais informação. Porém, a estimativa da avaliação dos discentes situados fora desse intervalo fica prejudicada, visto que o instrumento não

consegue estimar com precisão os discentes que avaliam muito bem ou muito mal a qualidade do instrumento, pela falta de itens posicionados nas regiões extremas.

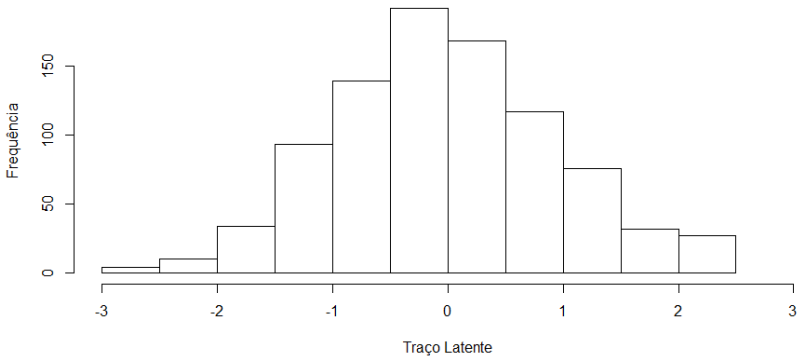
Figura 5 – Função de Informação do teste.



Fonte: Elaborado pelos autores.

A Figura 6 apresenta a distribuição do valor de traço latente estimado (grau de avaliação dos discentes da Pós-Graduação da UFSM). Pode-se observar que a distribuição não é simétrica nas caudas e que a maioria dos discentes encontra-se entre -1,5 e 1,5 na escala.

Figura 6 – Histograma do nível de avaliação dos discentes.



Fonte: Elaborado pelos autores.

Foi criada uma escala (0,1), isto é, com média igual a zero e desvio padrão igual a um. No que se refere à criação da escala, dos 34 itens avaliados, 16 foram considerados âncoras e 1 item “quase âncora”, no entanto, apenas três níveis âncoras foram caracterizados (Níveis -1, 0 e 1), visto que os itens estão concentrados basicamente nessa região, conforme Tabela 4. Observa-se que nenhum item está

situado nos níveis -3, -2 e 2. Isso significa que o instrumento não possui itens que avaliados positiva ou negativamente pela maioria dos respondentes.

Tabela 4: Itens Característicos Segundo Nível Âncora

Níveis Âncoras	Itens Âncoras	Percentual de Respondentes
- 3	-	1,57%
- 2	-	14,24%
- 1	33	37,11%
0	3,6,9,10,12,17,18,23,24,26,30 e 34	31,95%
1	8,14,15 e 28	12,11%
2	-	0,03%

Fonte: Elaborado pelos autores

Os níveis âncoras e as interpretações são apresentados a seguir:

Nível -1 (Possui um Item): discentes situados nesse nível avaliam positivamente:

- Item 33: A atuação do seu orientador em relação ao acompanhamento e orientação do seu projeto de Pós-Graduação.

Nível 0 (Possui doze Itens): discentes situados nesse nível avaliam positivamente o item do nível -1e os demais itens:

- Item 3: As ações voltadas para a inclusão social promovidas pela Instituição;
- Item 6: A atuação da Instituição no cumprimento da sua missão;
- Item 9: O Ambiente Virtual de Ensino e Aprendizagem (AVEA) oferecido pela Instituição;
- Item 10: A capacitação e a qualificação dos servidores docentes e técnico-administrativos em educação nos diversos setores administrativos da Instituição;
- Item 12: A disponibilidade e a atualização do acervo das bibliotecas que utiliza;
- Item 17: As condições de infraestrutura da Instituição;
- Item 18: As condições de infraestrutura de sua unidade/subunidade;
- Item 23: A contribuição dos planos de ensino para a sua formação acadêmica, no que diz respeito à atualização, conteúdos e bibliografias das disciplinas;
- Item 24: O acompanhamento, cumprimento e divulgação dos trabalhos de conclusão de curso;

- Item 26: A orientação e o acompanhamento das atividades práticas desenvolvidas nos laboratórios;
- Item 30: A atuação do diretor do seu centro de ensino/campus, considerando a melhoria contínua e o atendimento às necessidades dos estudantes;
- Item 34: As condições oferecidas para a realização das pesquisas necessárias ao seu embasamento acadêmico-profissional.

Nível 1 (Possui quatro Itens): Discentes situados nesse nível avaliam positivamente os itens dos níveis anteriores e mais os seguintes:

- Item 8: Os canais de comunicação da Instituição com a comunidade externa;
- Item 14: A prestação de serviços privados disponíveis na Instituição (restaurantes/lancherias e reprografias);
- Item 15: Os horários de atendimento dos serviços privados disponíveis à comunidade (transporte coletivo, restaurantes/lancherias e reprografias);
- Item 28: Os critérios adotados para a concessão de bolsas.

5 CONCLUSÃO

Todos os objetivos propostos nesta pesquisa foram alcançados. De acordo com análise realizada por meio do MLU2P, pode-se concluir que todos os itens são adequados, visto que possuem bom poder discriminatório e se posicionaram dentro do intervalo esperado de variação. Os itens do instrumento da autoavaliação respondido pelos discentes dos cursos de Pós-Graduação da UFSM foram ajustados adequadamente pelo Modelo utilizado, ficando posicionados entre -2 e 1 na escala do traço latente.

Porém, a estimativa da avaliação dos discentes situados fora desse intervalo fica prejudicada, visto que o instrumento não consegue estimar com precisão os discentes que avaliam muito bem ou muito mal a qualidade do instrumento, pela falta de itens posicionados nas regiões extremas. Quanto ao instrumento de Autoavaliação, percebeu-se que os itens que compõem esse instrumento são, em geral, fáceis de serem bem avaliado, segundo os resultados obtidos, o que, de fato, é um bom resultado quando se trata de Autoavaliação Institucional.

O traço latente "grau de avaliação dos discentes da Pós-Graduação da UFSM" foi estimado e os resultados mostram que a maior parte dos discentes se situa entre -1 e 1 na escala. No que se refere à criação de escala, muitos itens foram considerados âncoras (16) e "quase âncora" (1), porém, apenas três níveis âncoras foram caracterizados: Nível -1 (Um Item), Nível 0 (Doze Itens) e Nível 1 (Quatro Itens). Isso significa que os itens possuem uma avaliação positiva, isto é, o instrumento não possuem itens que são avaliados negativamente. Nesse sentido, sugere-se que as perguntas do instrumento de avaliação institucional sejam desmembradas para que os itens sejam melhor avaliados, pois no instrumento atual cada questão avalia vários aspectos simultaneamente, o que ocasiona um equilíbrio na resposta final do item, sendo que um aspecto pode estar muito bom e outro não, mas na média é avaliado como sendo bom.

REFERÊNCIAS

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da resposta ao item: conceitos e aplicações.** São Paulo: Associação Brasileira de Estatística, 2000.

BALZAN, N. C.; DIAS SOBRINHO, J. **Avaliação institucional: teoria e experiências**. São Paulo: Cortez, 2005.

BEATON, A. E.; ALLEN, N. L. Interpreting Scales through Scale Anchoring. **Journal of Educational Statistics**, n. 17, p. 191-204, 1992.

BAKER, F. B. **The Basics of Item Response Theory**. 2 ed. USA: ERIC Clearinghouseon Assessment and Evaluation, 2001.

BOCK, R. D.; ZIMOWSKI, M. F. Multiple Group IRT. In: VAN DER LINDER, W. J.; HAMBLETON, R. K. **Handbook of Modern Item Response Theory**. New York: SpringVerlag, 1997.

BIRNBAUM, A. Some Latent Trait Models and Their Use in Infering an Examinee's Ability. In: LORD, F. M.; NOVICK, M. R. **Statistical Theories of Mental Test Scores**. Reading, MA: Addison-Wesley, 1968.

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 2016**. Ministério da Educação. Universidade Federal de Santa Maria. Comissão Própria de Avaliação da UFSM, 2016. Disponível em <http://coral.ufsm.br/avaliacaoufsm/images/documentos/relatorios/2016-2017.pdf>. Acesso em 13/06/2017.

CPA. **Relatório de Avaliação Interna: Autoavaliação da UFSM 2008**. Ministério da Educação. Universidade Federal de Santa Maria. Comissão Própria de Avaliação da UFSM, 2008. Disponível em <http://w3.ufsm.br/proplan/images/coplai/UFSM_Relatrio_AutoAvaliao_2008.pdf >. Acesso em 12/10/2018.

DE AYALA, R. J. **The Theory and Practice of Item Response Theory**. New York, USA: The Guilford Press, 2009.

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.

GATTI, B. A. Avaliação institucional: processo descritivo, analítico ou reflexivo?. **Estudos em avaliação educacional**, v. 17, n. 34, p. 7-14, 2006.

GRIBOSKI, C. M.; PIXOTO, M. C. L.; HORTA, P. M. **Avaliação Externa, Autoavaliação e o PDI**. Avaliação, Campinas, Sorocaba, SP, v. 23, n. 1, p. 178-197, mar. 2018.

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.

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

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

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 Pratices**. New York, USA: Springer, 1995.

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

MOREIRA JUNIOR, F. J.; SZINVELSKI, C. R. P. . Utilização da Teoria da Resposta ao Item na Análise dos Dados da Autoavaliação Institucional da Universidade Federal de Santa Maria: Vantagens e Oportunidades. In: **Seminários Regionais sobre Autoavaliação Institucional e Comissões Próprias de Avaliação (CPA)**, 2013, Porto Alegre. Anais dos Seminários Regionais sobre Autoavaliação Institucional e Comissões Próprias de Avaliação (CPA) 2013. Brasília: Inep, 2013. p. 189-190.

MOREIRA JUNIOR, F. J.; ZANELLA, A.; LOPES, L. F. D.; SEIDEL, E. J. Avaliação da satisfação de alunos por meio do Modelo de Resposta Gradual da Teoria da Resposta ao Item, **Ensaio: aval. pol. públ. Educ**., Rio de Janeiro, v. 23, n. 86, p. 129-158, jan./mar. 2015.

NASCIMENTO, J. C. H. B. et al. Avaliação Institucional: Aplicação da Teoria da Resposta ao Item Para Avaliação Discente em Ciências Contábeis. **Future Studies Research Journal: Trends and Strategies**, v. 8, n. 2, p. 117-143, 2016.

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**, 2017. CRAN.R project, Disponível em <http://cran.rproject.org/web/packages/irtoys/irtoys.pdf> . Acesso em 05/04/2017.

RIZOPOULOS, D. **Package ltm: Latent Trait Models under IRT**, 2018. CRAN.R project, Disponível em <http://cran.r-project.org/web/packages/ltm/ltm.pdf> . Acesso em 16/04/2018.

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

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.