

# Avaliação da satisfação de alunos por meio do Modelo de Resposta Gradual da Teoria da Resposta ao Item

Fernando de Jesus Moreira Junior <sup>a</sup>

Andreia Zanella <sup>b</sup>

Luis Felipe Dias Lopes <sup>c</sup>

Enio Júnior Seidel <sup>d</sup>

## Resumo

O objetivo deste estudo foi utilizar a Teoria de Resposta ao Item (TRI) como ferramenta para a avaliação da satisfação dos alunos das disciplinas de Estatística Básica e Estatística Geral da Universidade Federal de Santa Maria (UFSM). Foi aplicado o Modelo de Resposta Gradual (MRG), por meio do qual foi criada uma escala de medida para a satisfação, em que foram posicionados os itens de satisfação e os respondentes, numa mesma unidade de medida. Os resultados mostraram que MRG da TRI foi adequado para a análise dos dados, permitindo a identificação de itens problemáticos, a construção de uma escala para a satisfação dos alunos e a identificação de níveis âncoras. Os achados irão permitir que os responsáveis pelo ensino da Estatística na instituição possam identificar em que é possível melhorar os procedimentos para que tal ensino seja cada vez mais satisfatório.

**Palavras-chave:** Teoria da Resposta ao Item. Modelo de Resposta Gradual. Avaliação da satisfação. Alunos.

## 1 Introdução

Educação pode ser entendida como uma prestação de serviço, gerando satisfação ou insatisfação em relação aos seus aspectos. Nesse sentido, a qualidade dos serviços oferecidos e a satisfação dos discentes são fundamentais para as Instituições de

---

<sup>a</sup> Universidade Federal de Santa Maria – UFSM – Departamento de Estatística. Santa Maria, Rio Grande do Sul, Brasil.

<sup>b</sup> Universidade Federal de Santa Maria – UFSM – Programa de Pós-Graduação em Engenharia de Produção. Santa Maria, Rio Grande do Sul, Brasil.

<sup>c</sup> Universidade Federal de Santa Maria – UFSM – Departamento de Ciências Administrativas. Santa Maria, Rio Grande do Sul, Brasil.

<sup>d</sup> Universidade Federal do Pampa – UNIPAMPA- Campus Itaqui – RS. Itaqui, Rio Grande do Sul, Brasil.

Recebido em: 22 Fev., 2013

Aceito em: 24 Jul., 2014

Ensino Superior (IES) que buscam sobreviver no mercado educacional (VIEIRA; KUNKEL; RIGHI, 2012).

A preocupação com a qualidade na educação tem aumentado nos últimos anos e os governos começam a manifestar, em suas políticas, uma preocupação cada vez maior com essa questão (DAVOK, 2007).

Qualidade em serviços pode ser compreendida como a capacidade que um processo possui de satisfazer uma necessidade, solucionar um problema ou fornecer um benefício a alguém. De acordo com Davok (2007), generalizando, o conceito de qualidade na área educacional abarca as estruturas, os processos e os resultados educacionais.

Para que o processo de ensino esteja inserido na melhoria contínua da qualidade, se faz necessária a busca pelo empenho de todos os envolvidos no processo, procurando desempenhar suas atividades da melhor forma e reconhecendo a importância da autoavaliação no ambiente universitário (ZANELLA; LOPES; SEIDEL, 2009).

Uma metodologia que tem sido utilizada para a avaliação da satisfação é a Teoria da Resposta ao Item (TRI). 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). Também é uma poderosa ferramenta estatística que surgiu para suprir as necessidades decorrentes das limitações da Teoria Clássica da Medida (TCM) ou Teoria Clássica do Teste (TCT), teoria que tradicionalmente era – e ainda é – utilizada nas avaliações. 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).

Este trabalho tem como objetivo utilizar a TRI como ferramenta de avaliação da satisfação dos alunos das disciplinas de Estatística da Universidade Federal de Santa Maria (UFSM), com aplicação do Modelo de Resposta Gradual de natureza unidimensional e cumulativa da TRI, criando uma escala de medida, na qual se localizam os itens de satisfação e os respondentes, numa mesma

unidade de medida. Os dados utilizados são oriundos das pesquisas de Zanella (2008), Zanella, Lopes e Seidel (2009) e Zanella, Seidel e Lopes (2010), em que foi aplicado um questionário para verificar o nível de satisfação quanto aos professores do Departamento de Estatística da UFSM, a infraestrutura oferecida e a contribuição da disciplina de Estatística para a formação profissional dos estudantes, bem como identificar aspectos do perfil desses alunos.

O uso de métodos estatísticos vem crescendo entre acadêmicos e pesquisadores de diversas áreas (BONAFÉ; LOFFREDO; CAMPOS, 2010), e a importância de suas aplicações como ferramenta de auxílio na tomada de decisões, de acordo com Mantovani e Gouvêa (2012), tem crescido e, com isto, surge a necessidade de se compreenderem as dificuldades no ensino desta disciplina.

O estudo justifica-se, então, pelo fato de que é necessário investigar os aspectos de satisfação ou insatisfação dos alunos em relação a tudo o que cerca o ensino da disciplina de Estatística, permitindo entender como está tal processo e o que pode ser melhorado.

Contudo, há de se frisar que o aluno não deve ser entendido como simplesmente um consumidor (um cliente), mas, sim, deve ser encarado como um verdadeiro parceiro no processo de aprendizagem (SOUZA; REINERT, 2010), fazendo com que o processo seja construído em conjunto por gestores, professores, estudantes e demais agentes.

## **2 Avaliação da satisfação**

Segundo Santos e Ferreira (2005), a avaliação pode ser entendida como uma apreciação da qualidade e da eficácia do sistema de ensino como um todo ou parte dele, e deve provocar, no aluno, 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. Esta deve ser compreendida e promovida 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 e, como tal, deve fazer parte da instituição, e realizar-se como cultura (BALZAN; DIAS SOBRINHO, 2005). Neste sentido,

para Souza e Reinert (2010), a satisfação tem grande relevância na avaliação discente de cursos superiores, por resultar de um julgamento formulado a partir da realidade percebida.

Segundo Paulins (2005), a satisfação dos clientes depende muito da qualidade dos serviços que estão consumindo e da qualidade que eles esperam dos serviços. Dessa forma, pode-se entender, no contexto universitário, que a percepção do estudante com relação à qualidade dos serviços educacionais recebidos é proporcional às suas expectativas sobre os mesmos.

De acordo com Vieira, Milach e Huppés (2008) e Vieira, Kunkel e Righi (2012), compreender quais as variáveis que influenciam a satisfação dos acadêmicos é fundamental para as IES que desejam manter a qualidade dos serviços prestados. Neste sentido, segundo Zanella, Lopes e Seidel (2009), a avaliação da satisfação do aluno pode ser entendida como um processo que gera conhecimento e informação para a tomada de decisões.

Monitorar os níveis de satisfação dos acadêmicos, avaliando o seu interesse e o dos professores, as formas de ensino, os currículos e a organização do curso, é uma maneira de garantir a satisfação das necessidades tanto dos alunos como da sociedade (VIEIRA; MILACH; HUPPES, 2008). Além disso, medir a satisfação do aluno é uma questão de sobrevivência para a IES, pois esta deve buscar continuamente a avaliação da satisfação. Para isso, é necessário ter um instrumento que indique se o caminho seguido está ou não em conformidade com aquilo que o consumidor (aluno) espera (MOREIRA JUNIOR, 2010).

Os questionários de satisfação possibilitam a uma instituição a indicação precisa do grau de acerto das diretrizes de seus processos, bem como podem determinar a qualidade dos serviços resultantes desses processos (ZANELLA; SEIDEL; LOPES, 2010). Da mesma forma que outras organizações modernas, as IES devem cumprir exigências de qualidade, competência e produtividade (VIEIRA; KUNKEL; RIGHI, 2012).

Assim, como concluem Zanella, Lopes e Seidel (2009), em seu estudo, a avaliação da satisfação do aluno pode ser entendida como um processo que gera conhecimento e informação para a tomada de decisões sobre os procedimentos de ensino de Estatística.

### 3 Teoria da Resposta ao Item (TRI)

A TRI é uma metodologia que sugere formas de representar a relação entre a probabilidade de um indivíduo apresentar certa resposta a um item e seus traços latentes (habilidades), por meio de um modelo matemático (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.

Existem várias vantagens em se utilizar a Teoria de Resposta ao Item em vez da Teoria Clássica de Medidas (TCM), dentre as quais se destacam:

a) A TRI permite fazer comparações de respondentes de populações diferentes quando são submetidos a testes ou questionário que tenham alguns itens comuns e ainda permite a comparação de indivíduos de mesma população submetidos a testes totalmente diferentes. Na TCM, só é possível comparar indivíduos submetidos ao mesmo teste (testes paralelos) (ANDRADE; TAVARES; VALLE, 2000);

b) Na TRI, os itens e os indivíduos estão na mesma escala; assim, o nível de uma característica que um indivíduo possui pode ser comparado ao nível da característica exigida pelo item, o que facilita a interpretação da escala gerada e permite também conhecer quais itens que estão produzindo informação ao longo da escala. Na TCM, os itens e indivíduos não estão na mesma escala (EMBRETSON; REISE, 2000);

c) A TRI possui o princípio da invariância, isto é, os parâmetros dos itens não dependem do traço latente do respondente e os parâmetros dos indivíduos não dependem dos itens apresentados, sendo este princípio um dos grandes benefícios da TRI. Na TCM, o escore do indivíduo depende do grau de dificuldade dos itens (HAMBLETON; SWAMINATAN; ROGERS, 1991).

Embora tenha começado na área de avaliação educacional (SOARES, 2005), a TRI expandiu-se rapidamente para as mais diversas áreas que têm como objetivo a avaliação, como no caso da avaliação da satisfação (BORTOLOTTI, 2003; BORTOLOTTI; ANDRADE, 2007; COSTA, 2001; COSTA; CHAVES NETO, 2002; MENDES, 2006).

Existe uma grande variedade de modelos matemáticos que são utilizados na TRI e a escolha desse modelo depende basicamente do tipo de item (dicotômico, politômico, gradual, etc.), da natureza do traço latente (cumulativo, não cumulativo) e da dimensionalidade do traço latente (unidimensional ou multidimensional). A dimensionalidade está relacionada com a quantidade de traços latentes que estão sendo analisados. A maioria das aplicações da TRI considera apenas um traço latente (AYALA, 2009; EMBRETSON; REISE, 2000; HAMBLETON; SWAMINATHAN; ROGERS, 1991).

Neste estudo, os itens foram construídos em uma escala gradual com cinco categorias de respostas. Dentre os modelos desenvolvidos para itens com resposta gradual, destaca-se o Modelo de Resposta Gradual (MRG) de Samejima (1969, 1972, 1997). Neste modelo, são estimados dois tipos de parâmetros relacionados com o item:

1.º) O parâmetro de dificuldade (“ $b_{i,k}$ ”): este parâmetro é referido como o parâmetro “limiar”. O parâmetro de dificuldade refere-se à probabilidade de um indivíduo comum com determinado nível do traço latente selecionar uma categoria de resposta (por exemplo, insatisfeito) ou uma categoria mais alta ordenada (por exemplo, pouco insatisfeito, satisfeito ou muito satisfeito). Esse parâmetro representa o ponto na escala do traço latente em que há 50% de chance de que uma dada categoria de resposta ou uma categoria mais alta ordenada seja selecionada, isto é, estes parâmetros representam os limiares entre as categorias de resposta;

2.º) O parâmetro de discriminação (“ $a_i$ ”): este parâmetro representa o quanto um item discrimina entre os respondentes de diferentes níveis do traço latente, determinando a “qualidade” do item. Quanto maior o valor deste parâmetro, melhor é o item e melhor é a discriminação entre os indivíduos nos diferentes níveis de traço latente.

Suponha que os escores das categorias de um item  $i$  são arranjados em ordem do menor para o maior e denotados por  $k=0;1;...; m_i$ , em que  $m_i+1$  é o número de categorias do  $i$ -ésimo item. A probabilidade de um indivíduo  $j$  escolher uma particular categoria de resposta ou outra mais alta do item  $i$  pode ser dada por

uma extensão do Modelo Logístico Unidimensional de Dois Parâmetros (MLU2) (ANDRADE; TAVARES; VALLE, 2000), conforme a Equação 1:

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_{i,k})}} \quad (1)$$

em que:

$i = 1, 2, \dots, I$  ( $I$  é a quantidade de itens no teste);

$j = 1, 2, \dots, n$  ( $n$  é o número total de respondentes);

$k = 0, 1, \dots, m_i$  ( $m_i$  é o número de categorias menos 1 do  $i$ -ésimo item);

$b_{i,k}$  é o parâmetro de dificuldade da  $k$ -ésima categoria do item  $i$ ;

$a_i$  é o parâmetro de discriminação do item  $i$ ;

$\theta_j$  representa a habilidade (traço latente) do  $j$ -ésimo indivíduo.

Nesse modelo, a discriminação de uma categoria específica de resposta depende tanto do parâmetro de inclinação  $a_i$ , comum a todas as categorias do item, quanto da distância das categorias de dificuldade adjacentes. Necessariamente, deve-se ter uma ordenação entre o nível de dificuldade das categorias de um dado item, de acordo com a classificação de seus escores, ou seja,

$$b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i}.$$

A probabilidade de um indivíduo  $j$  receber um escore  $k$  no item  $i$  é dada então pela Equação 2:

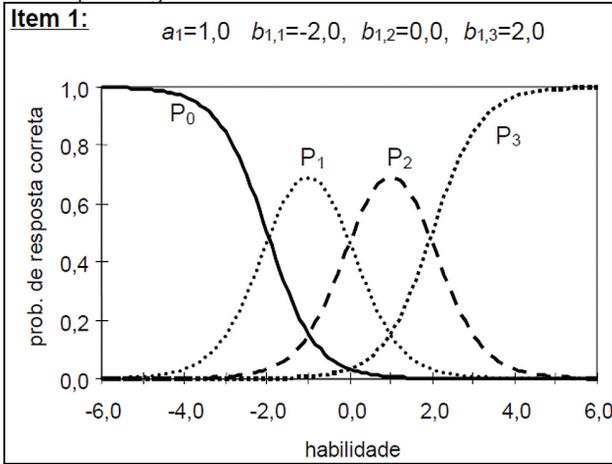
$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-a_i(\theta_j - b_{i,k+1})}} \quad (2)$$

de tal forma que:

$$P_{i,0}^+(\theta_j) = 1;$$

$$P_{i,m_i+1}^+(\theta_j) = 0.$$

Observa-se que em um item com  $(m_i + 1)$  categorias,  $m_i$  valores de dificuldade necessitam ser estimados, além do parâmetro de inclinação do item. Assim, para cada item, o número de parâmetros a ser estimado será dado pelo seu número de categorias de resposta. A Figura 1 apresenta um gráfico de um item com parâmetros estimados pelo MRG, denominado Curva Característica do Item (CCI), para um item hipotético com quatro categorias e com parâmetros  $a_1 = 1$ ;  $b_{1,1} = -2$ ;  $b_{1,2} = 0$ , e  $b_{1,3} = 2$ .

**Figura 1** - Exemplo de gráfico de um MRG

Fonte: Andrade, Tavares e Valle (2000).

Observando-se o gráfico da Figura 1, dá-se a seguinte interpretação: indivíduos que possuem o valor do traço latente inferior a  $-2$  têm maior probabilidade de responder à categoria 1; indivíduos que possuem o valor do traço latente entre  $-2$  e  $0$  têm maior probabilidade de responder à categoria 2; indivíduos que possuem o valor do traço latente entre  $0$  e  $2$  têm maior probabilidade de responder à categoria 3, e indivíduos que possuem o valor do traço latente maior que  $2$  têm maior probabilidade de responder à categoria 4.

Nos modelos politômicos, o parâmetro de discriminação deve ser interpretado cuidadosamente (MATTEUCCI: STRACQUALURSI, 2006). Para avaliar a quantia de discriminação que um item fornece, utiliza-se a função de informação do item (FII), que no MRG é dada pela Equação 3:

$$I_i(\theta) = \sum_{x=1}^{k_i} \frac{P'_{ik}(\theta)^2}{P_{ik}(\theta)} \quad (3)$$

em que  $P'_{ik}(\theta)$  é a primeira derivada da curva de resposta da categoria avaliada em um particular nível do traço latente. Quanto maior for a discriminação de um item, maior será a informação que este fornece ao teste. A função de informação total do teste (FIT) é a soma das informações fornecidas por cada item que compõe o mesmo. A curva de informação total é utilizada para avaliar o desempenho dos itens, ou seja, o quão bem um conjunto de itens está avaliando o traço latente. A função de informação está relacionada com a precisão necessária para estimar o traço latente, de forma que o erro padrão de medida pode ser estimado como o inverso da raiz quadrada do valor da informação total do teste em cada nível do traço latente.

A estimação dos parâmetros no MRG é feita por meio de métodos estatísticos que necessitam de recursos computacionais para serem utilizados. Em geral, utiliza-se o método da Máxima Verossimilhança (MV) ou algum método bayesiano, como, por exemplo, da Esperança a Posteriori (EAP) ou da Moda a Posteriori (MAP) (BAKER, 1992). O processo de estimação dos parâmetros dos itens e dos respondentes está implementado em softwares específicos, tais como MULTILOG (THISSEN, 1991,2003; TOIT, 2003) e PARSCALE (MURAKI; BOCK, 1997; TOIT, 2003). Neste trabalho, os parâmetros do modelo foram estimados por meio do método da Máxima Verossimilhança Marginal, método implementado no software MULTILOG.

Para interpretar os valores dos parâmetros dos itens, é necessário conhecer a escala na qual estes foram determinados. A construção da escala é feita definindo-se níveis âncora e identificando itens âncoras. Considerando-se dois níveis âncora, X e Y com  $X < Y$ , um item é âncora para um nível Y se, e somente se, satisfizer as condições (BEATON; ALLEN, 1992):

- a)  $P(U = 1 | \theta = Y) \geq 0,65$
- b)  $P(U = 1 | \theta = X) < 0,50$ ; e
- c)  $P(U = 1 | \theta = Y) - P(U = 1 | \theta = X) \geq 0,30$ .

Isto significa que, para um item ser considerado âncora em um determinado nível âncora Y, é necessário que seja respondido positivamente por pelo menos 65% dos respondentes com este nível do traço latente e por uma proporção menor do que 50% dos respondentes de um nível de traço latente imediatamente inferior X. E ainda que a diferença entre a proporção de respondentes destes dois níveis consecutivos deve ser pelo menos de 30% (BEATON; ALLEN, 1992). Portanto, um item é considerado âncora quando for típico daquele nível âncora Y e atender às três condições descritas. Essas condições são válidas para os Modelos Logísticos (dicotômicos) da TRI, porém, conforme Mafra (2010), podem ser adaptadas à Modelos Politômicos, em que não é verificado se um item é âncora, mas se uma categoria do item é âncora. Para os Modelos Politômicos, esse é um processo de dicotomização das k categorias dos itens, de tal forma que é possível verificar se k-1 categorias são âncoras ou não. No entanto, não é possível verificar se a categoria mais baixa da escala é âncora, conforme a aplicação de Mafra (2010).

A partir das respostas obtidas de respondentes a um conjunto de itens, a TRI possibilita a estimação dos parâmetros dos itens e dos indivíduos em uma escala de medida. Por exemplo, neste estudo, em que está sendo avaliado nível de satisfação, a análise com a TRI irá estimar o nível de satisfação do respondente (traço latente) e também os parâmetros dos itens de modo a criar uma escala de medida do nível de satisfação.

## 4 Metodologia

O desenvolvimento do presente trabalho constitui-se de pesquisa bibliográfica e de campo, com abordagem quantitativa, desenvolvida para avaliar a satisfação dos alunos das disciplinas de Estatística em relação aos professores do Departamento de Estatística da Universidade Federal de Santa Maria (UFSM), à infraestrutura das salas de aula e à disciplina. A amostra foi constituída de 500 alunos dos cursos de graduação e pós-graduação que tiveram aula da disciplina de Estatística no primeiro semestre do ano de 2007, conforme especificado por Zanella (2008), Zanella, Lopes e Seidel (2009) e Zanella, Seidel e Lopes (2010). Estudos anteriores sobre o modelo de resposta gradual indicam que é necessária uma amostra de pelo menos 500 respondentes, distribuídos nos diferentes níveis do traço latente, para obter uma calibração adequada dos itens (REISE; YU, 1990). A coleta dos dados foi realizada por meio de um questionário fechado, composto por 30 itens. Os sete primeiros eram qualitativos e buscavam caracterizar o perfil dos alunos. Os demais, num total de 23 itens, eram quantitativos e buscavam investigar a satisfação do aluno, sendo que 13 itens avaliavam o professor, três itens avaliavam a infraestrutura e sete itens avaliavam a disciplina. Esses itens foram construídos baseados na escala de Likert, que permite respostas com níveis variados de classificação, e passaram por uma validação do construto, sendo validados por Zanella, Seidel e Lopes (2010). O embasamento teórico utilizado na construção dos itens pode ser encontrado em Zanella (2008). As opções de resposta variavam em “muito insatisfeito” a “muito satisfeito”, em uma escala de 1 a 5. A descrição dos itens é apresentada no Quadro 1.

As análises utilizadas foram estatísticas descritivas, tabelas de frequências, análise multivariada (análise fatorial) e a Teoria da Resposta ao Item (TRI) com o Modelo de Resposta Gradual (MRG). Foi utilizado o software FACTOR, versão 9.2 (LORENZO-SEVA; FERRNADO, 2013), para a análise fatorial de dados qualitativos, e o software MULTILOG (TOIT, 2003), para a análise da TRI. A TRI foi utilizada para a avaliação dos 23 itens, para a estimação do valor do traço latente e para a criação da escala de medida.

**Quadro 1** - Itens avaliados na medida de satisfação

Item	Descrição
Professor	
1	Disponibilidade do professor em responder às solicitações dos alunos
2	Confiança inspirada pela postura do professor da disciplina
3	O domínio do assunto apresentado pelo professor da disciplina
4	Capacidade de síntese do conteúdo pelo professor da disciplina
5	Capacidade de estímulo/motivação apresentada pelo professor
6	Capacidade do professor em manter o aluno atento e interessado na aula
7	Clareza do professor ao apresentar os conteúdos
8	Apresentação da teoria e aplicação prática dos temas abordados
9	Possibilidade de contatar o professor fora da sala de aula
10	Forma de avaliação proposta pelo professor da disciplina
11	Pontualidade e assiduidade apresentadas pelo professor
12	Informações/esclarecimento quanto ao planejamento da disciplina
13	Estruturação com que as aulas são apresentadas
Infraestrutura	
14	As instalações utilizadas durante as aulas da disciplina
15	Qualidade dos recursos didáticos e demais materiais relacionados à disciplina
16	Disponibilidade das referências bibliográficas e demais materiais didáticos
Disciplina	
17	Profundidade/desenvolvimento dos temas em relação aos objetivos da disciplina
18	Aplicabilidade/aprendizagem dos temas abordados
19	Interação curricular do curso com a disciplina
20	Afinidade entre sua ocupação profissional e os conhecimentos recebidos
21	Entendimento, participação e acompanhamento dos assuntos abordados
22	Melhoria no seu desempenho profissional, gerando resultados positivos
23	Atendimento da disciplina às suas expectativas

Fonte: Elaborado pelos autores (2013).

## 5 Resultados

Primeiramente, foi feita uma análise descritiva dos dados para a caracterização do perfil da amostra. Observou-se que 92,2% dos alunos eram de cursos de graduação, 49,3% tinham entre 20 e 24 anos, e 36,5% tinham vínculo empregatício ou estavam realizando algum estágio. Além disso, 38,9% nunca estudavam Estatística fora da sala de aula, enquanto que 55,1% estudavam fora da sala de aula por até duas horas semanais. Quanto à bibliografia indicada para a disciplina de estatística, 27,1% nunca consultavam, 65,9% consultavam somente quando necessário e 7,0% consultavam com frequência. Em relação às tarefas solicitadas pelo professor, 5% nunca realizavam, 45,3% sempre realizavam e 49,7% só realizavam quando tinham tempo disponível. Cerca de 5,6% dos alunos eram repetentes na disciplina de Estatística.

A Tabela 1 apresenta o percentual das respostas válidas aos 23 itens utilizados na avaliação da satisfação, descritos no Quadro 1. Também são apresentados os valores médios de satisfação, os quais variam de 1 a 5, sendo dessa forma classificados: quanto mais próximo de 1, menor é o grau de satisfação, e quanto mais próximo de 5, maior é o grau de satisfação.

**Tabela 1 - Distribuição de Frequências Relativas dos Itens e Avaliação Média**

CATEGORIAS	1	2	3	4	5	Média
ITEM 1	1,8%	2,4%	17,8%	28,2%	49,8%	4,2
ITEM 2	1,4%	5,8%	19,7%	34,1%	39,0%	4,0
ITEM 3	0,6%	3,2%	12,7%	28,9%	54,6%	4,3
ITEM 4	1,8%	3,6%	18,4%	35,0%	41,1%	4,1
ITEM 5	5,2%	13,6%	26,8%	28,6%	25,8%	3,6
ITEM 6	5,8%	13,1%	29,3%	28,9%	22,9%	3,5
ITEM 7	3,0%	8,2%	21,1%	33,2%	34,4%	3,9
ITEM 8	2,4%	9,2%	24,3%	30,5%	33,5%	3,8
ITEM 9	5,3%	8,9%	28,5%	27,9%	29,4%	3,7
ITEM 10	4,0%	5,8%	19,1%	27,0%	44,1%	4,0
ITEM 11	0,4%	3,0%	8,4%	18,6%	69,5%	4,5
ITEM 12	1,2%	3,8%	19,9%	37,6%	37,6%	4,1
ITEM 13	1,0%	8,1%	21,6%	36,3%	33,1%	3,9
ITEM 14	8,8%	14,0%	29,2%	24,4%	23,6%	3,4
ITEM 15	5,8%	12,2%	31,5%	29,3%	21,1%	3,5
ITEM 16	9,5%	14,7%	34,4%	24,9%	16,5%	3,2
ITEM 17	3,0%	5,5%	32,9%	35,8%	22,8%	3,7
ITEM 18	5,4%	11,4%	24,6%	34,9%	23,6%	3,6
ITEM 19	11,6%	13,5%	31,1%	28,7%	15,1%	3,2
ITEM 20	9,3%	11,5%	33,4%	31,6%	14,2%	3,3
ITEM 21	3,4%	9,6%	28,9%	34,9%	23,1%	3,6
ITEM 22	7,2%	11,9%	28,4%	33,2%	19,3%	3,5
ITEM 23	8,2%	12,0%	27,1%	29,7%	23,0%	3,5

Fonte: Elaborado pelos autores (2013).

Observando-se a Tabela 1, podem-se verificar diversos itens que trazem satisfação ou não aos alunos. Entretanto, essa é uma análise basicamente descritiva e não possui a capacidade de fornecer informação sobre a qualidade dos itens e dos seus resultados. Por exemplo, pode-se verificar que o item 11 (pontualidade e assiduidade do professor) é o que possui maior satisfação média (4,5), em que 88,1% dos alunos afirmaram que estavam “satisfeitos” ou “muito satisfeitos”; entretanto, não há indicadores que avaliam se esse item é bom para avaliar a satisfação do aluno. Os itens 16 (disponibilidade de material bibliográfico) e

19 (interação do curso com a disciplina) tiveram a menor satisfação média, porém também não é possível saber se estes são bons indicadores para avaliar a satisfação. A análise feita por meio da TRI pode verificar se esses itens são bons indicadores.

Antes da análise por meio da TRI, foi feita uma Análise Fatorial baseada na matriz das correlações tetracóricas, para verificar a suposição de unidimensionalidade dos dados. O procedimento utilizado para determinar o número de fatores foi a implementação ótima da Análise Paralela (TIMMERMAN; LORENZO-SEVA, 2011). As variáveis do presente estudo são de natureza qualitativa e utilizam uma escala Likert equilibrada com cinco categorias. A Figura 2 apresenta o resultado da análise fatorial obtida pelo software FACTOR.

Conforme se observa na Figura 2, o número aconselhado de dimensões, quando o percentil 95 é considerado, é de uma dimensão, em que o componente principal dominante, que se destaca fortemente dos demais, explica 53,5% da variância, o que sugere que existe um fator dominante. Isso não significa que o construto seja propriamente unidimensional, mas a existência de um fator dominante permite que uma análise com modelos unidimensionais da TRI possa ser realizada (ANDRADE; TAVARES; VALLE, 2000). Esse fator dominante é o traço latente

**Figura 2 - Resultado da Análise Fatorial**

Variable	Real-data % of variance	Mean of random % of variance	95 percentile of random % of variance
1	53.5**	8.9	9.9
2	8.3*	8.2	8.9
3	6.7	7.7	8.3
4	4.1	7.3	7.8
5	3.3	6.9	7.4
6	3.0	6.5	7.0
7	2.7	6.1	6.5
8	2.3	5.8	6.2
9	2.0	5.4	5.8
10	1.8	5.0	5.4
11	1.8	4.7	5.1
12	1.7	4.3	4.7
13	1.5	4.0	4.4
14	1.4	3.6	4.1
15	1.3	3.3	3.7
16	1.1	2.9	3.3
17	1.0	2.5	3.0
18	0.8	2.1	2.6
19	0.8	1.8	2.2
20	0.6	1.4	1.9
21	0.2	1.0	1.5
22	0.1	0.6	1.0
23	0.0	0.0	0.0

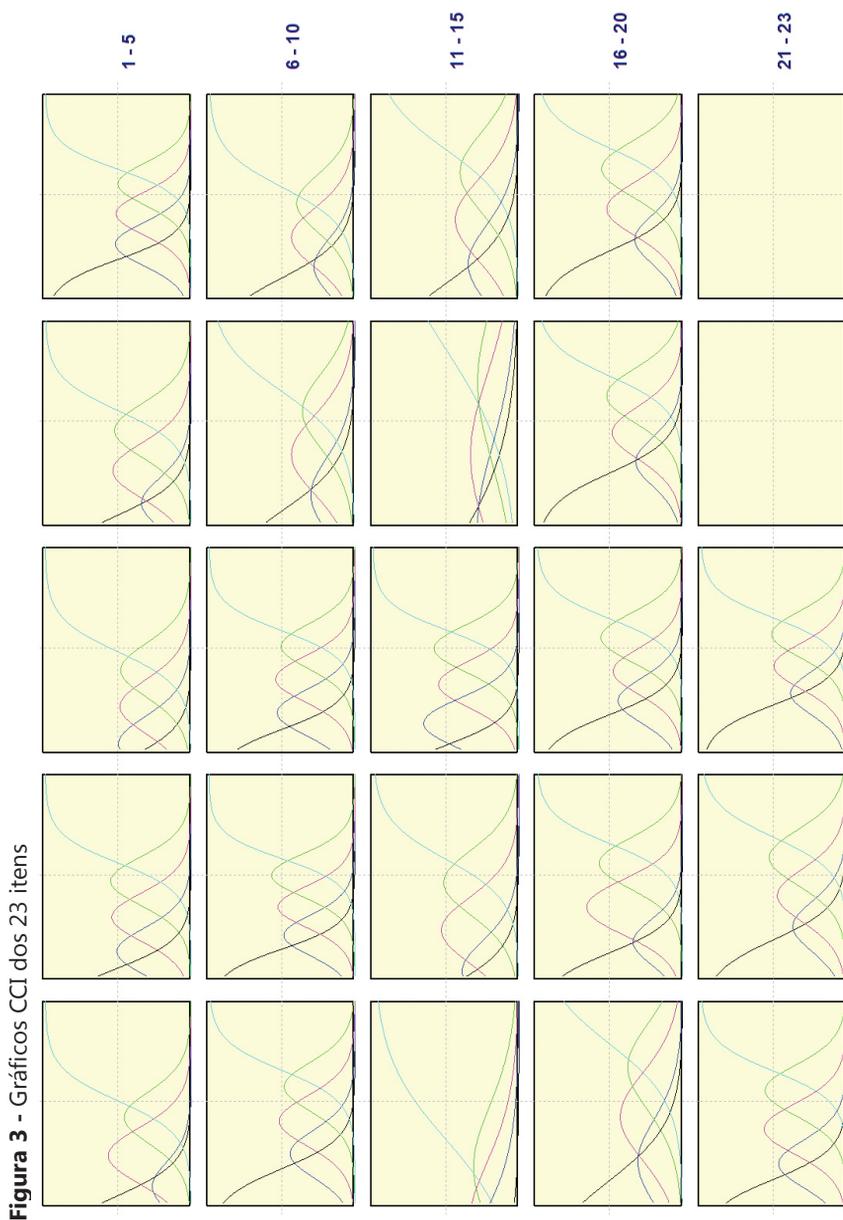
\*\* Advised number of dimensions when 95 percentile is considered: 1  
\* Advised number of dimensions when mean is considered: 2

Fonte: Elaborado pelos autores (2013).

que está sendo medido, que foi denominado como “a satisfação dos alunos das disciplinas de Estatística da UFSM”.

Com a suposição de unidimensionalidade verificada, pode-se utilizar o MRG da TRI. O tratamento dos dados foi feito utilizando-se um software específico para a estimação dos parâmetros dos itens e do traço latente, chamado MULTILOG (TOIT, 2003). No processo de calibração, são estimados os parâmetros de discriminação “ $a_i$ ” de cada item, os da localização “ $b_i, k$ ” de cada categoria e de cada item, e os traços latentes dos respondentes (nível de satisfação de cada respondente). A Figura 3 apresenta os gráficos das CCI dos 23 itens avaliados. Nessa Figura, as categorias estão apresentadas nas curvas dos gráficos, da esquerda (muito insatisfeito) para a direita (muito satisfeito). Observa-se que, quando as categorias do item não se sobressaem em alguma região do gráfico, isso pode ser um indicativo de que o item não possui uma boa qualidade para avaliar o traço latente ou um problema da amostra, que pode não conter sujeitos em todos os níveis do traço latente. Nota-se que os itens 11 e 14 não apresentaram um bom desempenho. O item 11 possui baixa discriminação (0,79) e está posicionado muito abaixo do esperado (-3,92) para uma escala com média zero e desvio padrão 1 (0, 1). Observa-se que a sua curva da categoria 5 (muito satisfeito) se sobressai em relação às demais categorias em praticamente todo o intervalo do traço latente no gráfico entre -3 e 3. Isso indica que, nesse item, todos os indivíduos se consideram satisfeitos e, por isso, não há discriminação entre satisfeitos e insatisfeitos. Já o item 14 está bem localizado (-1,00), mas possui uma discriminação baixa (0,59). Observa-se que apenas duas de suas curvas se sobressaem em relação às demais: as categorias 3 e 5. A princípio, esses dois itens não se mostram adequados para compor o questionário. Uma solução possível seria recategorizar (diminuindo a quantidade de categorias) o item, de tal forma que as categorias resultantes se sobressaíssem em alguma região no gráfico. Outra alternativa seria eliminar esses itens da análise. Os demais itens parecem adequados, embora alguns deles apresentem alguma categoria que não se sobressai em nenhuma região do gráfico.

Pelo fato de haver dois itens (11 e 14) que não foram ajustados adequadamente, a análise foi refeita, desconsiderando-se esses itens. Para desconsiderar itens no Software MULTILOG, codifica-se com o valor 0 (zero) todas as categorias dos itens que não devem ser considerados na análise. Embora esses itens sejam desconsiderados, o Software MULTILOG gera um resultado para esses itens e até o gráfico, que deverão ser ignorados na análise; no entanto, o erro padrão é estimado como sendo igual a 0 (zero). Na Tabela 2, são visualizadas as estimativas



**Figura 3** - Gráficos CCI dos 23 itens

Fonte: Elaborado pelos autores (2013).

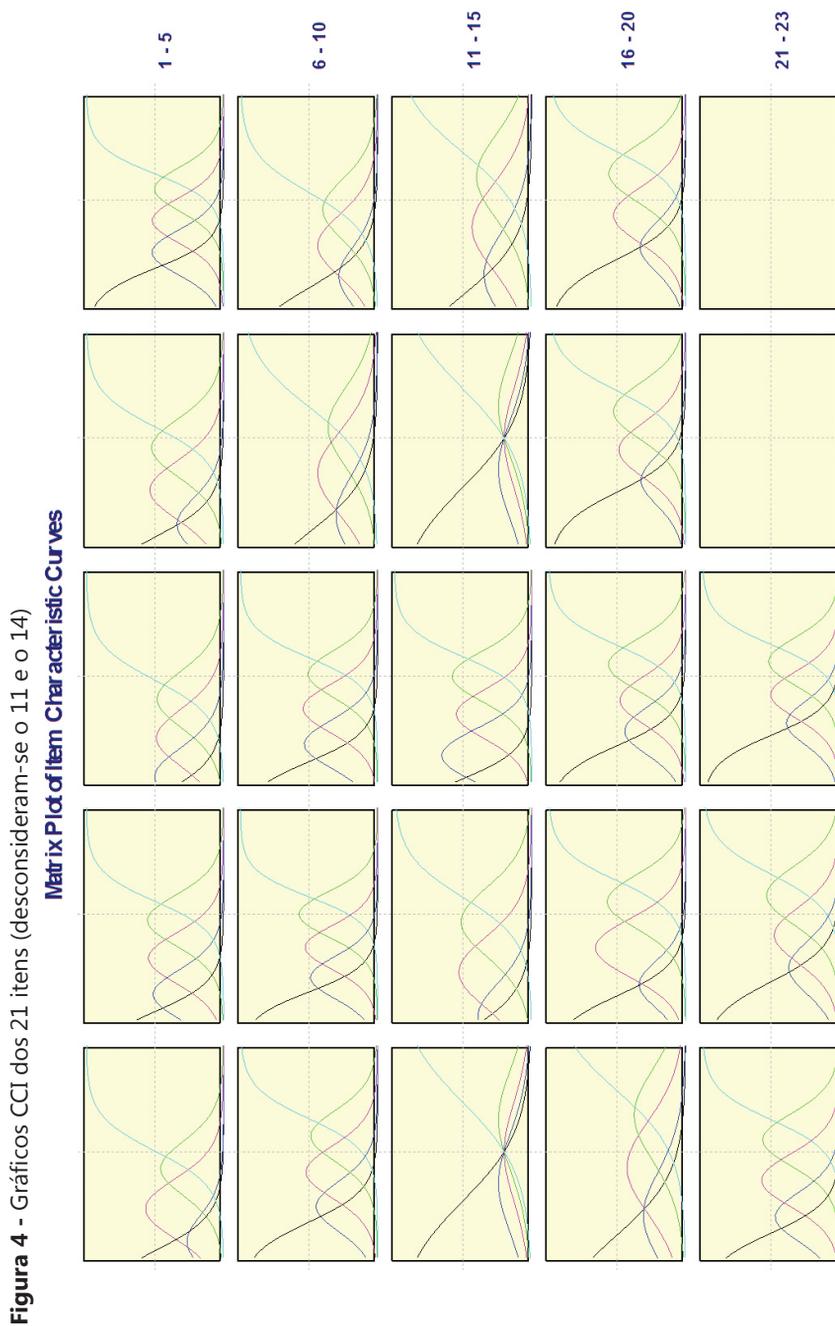
dos parâmetros de discriminação dos itens e de localização das categorias e os respectivos erros padrão (entre parênteses), além do valor médio do parâmetro  $b$  para cada item, para a nova calibração que desconsidera os itens 11 e 14. A Figura 4 apresenta os gráficos das CCI desses 21 itens avaliados.

O primeiro  $b_{i,k}$ , no caso o  $b_{i,1}$ , indica o ponto de inflexão da curva da primeira categoria. O último  $b_{i,k}$ , no caso o  $b_{i,4}$ , indica o ponto de inflexão da curva da última categoria. Os  $b_{i,k}$  intermediários referem-se aos pontos médios dos picos entre duas categorias adjacentes (EMBRETSON; REISE, 2000). Nesse estudo, existem dois  $b_{i,k}$  intermediários: no caso, o  $b_{i,2}$ , cujo valor é a média entre os pontos de picos entre as categorias 2 e 3, e o  $b_{i,3}$ , cujo valor é a média entre os pontos de picos entre as categorias 3 e 4. Já a média dos valores dos  $b_{i,k}$  indica qual a posição do item na escala criada e, neste estudo, indica também o nível médio de satisfação do item. Por exemplo, o item 19 (interação do curso com a disciplina) possui o maior valor de  $b$  médio ( $-0,22$ ) enquanto que o item 3 (domínio do assunto apresentado pelo professor) possui o menor valor de  $b$  médio ( $-1,76$ ), sendo que a média geral foi  $-0,88$ . Isso indica que os itens estão

**Tabela 2 - Estimativas dos parâmetros dos itens e seus respectivos erros padrão**

ITEM	a	b1	b2	b3	b4	b médio
1	2,00 (0,20)	-2,81 (0,30)	-2,26 (0,21)	-0,97 (0,10)	0,02 (0,08)	-1,50
2	2,31 (0,20)	-2,77 (0,26)	-1,79 (0,14)	-0,72 (0,08)	0,36 (0,07)	-1,23
3	1,94 (0,20)	-3,43 (0,47)	-2,30 (0,22)	-1,20 (0,11)	-0,12 (0,08)	-1,76
4	1,96 (0,18)	-2,80 (0,30)	-2,08 (0,18)	-0,86 (0,10)	0,32 (0,08)	-1,36
5	2,54 (0,19)	-1,93 (0,15)	-1,02 (0,08)	-0,12 (0,06)	0,75 (0,08)	-0,58
6	2,17 (0,18)	-1,99 (0,17)	-1,11 (0,10)	-0,04 (0,08)	0,93 (0,09)	-0,55
7	2,63 (0,21)	-2,21 (0,17)	-1,40 (0,10)	-0,51 (0,07)	0,48 (0,07)	-0,91
8	2,39 (0,20)	-2,43 (0,22)	-1,43 (0,11)	-0,41 (0,07)	0,52 (0,07)	-0,94
9	1,30 (0,14)	-2,67 (0,30)	-1,71 (0,19)	-0,28 (0,11)	0,89 (0,14)	-0,94
10	1,75 (0,18)	-2,48 (0,25)	-1,82 (0,17)	-0,75 (0,10)	0,22 (0,09)	-1,21
12	1,62 (0,16)	-3,39 (0,42)	-2,38 (0,23)	-0,93 (0,11)	0,47 (0,10)	-1,56
13	2,38 (0,20)	-2,91 (0,30)	-1,60 (0,12)	-0,57 (0,08)	0,54 (0,08)	-1,13
15	1,24 (0,14)	-2,68 (0,32)	-1,51 (0,18)	-0,02 (0,11)	1,35 (0,18)	-0,72
16	1,15 (0,13)	-2,36 (0,30)	-1,25 (0,18)	0,35 (0,13)	1,72 (0,22)	-0,38
17	2,20 (0,18)	-2,32 (0,20)	-1,68 (0,13)	-0,25 (0,07)	0,93 (0,10)	-0,83
18	2,30 (0,18)	-1,97 (0,16)	-1,14 (0,10)	-0,23 (0,07)	0,87 (0,10)	-0,62
19	2,03 (0,18)	-1,53 (0,14)	-0,86 (0,09)	0,19 (0,08)	1,33 (0,12)	-0,22
20	2,05 (0,18)	-1,67 (0,15)	-1,00 (0,10)	0,14 (0,07)	1,37 (0,13)	-0,29
21	2,23 (0,18)	-2,30 (0,20)	-1,39 (0,11)	-0,23 (0,07)	0,90 (0,09)	-0,75
22	1,98 (0,18)	-1,92 (0,17)	-1,13 (0,11)	-0,08 (0,08)	1,12 (0,11)	-0,51
23	2,39 (0,20)	-1,68 (0,14)	-0,99 (0,09)	-0,07 (0,07)	0,89 (0,09)	-0,46

Fonte: Elaborado pelos autores (2013).



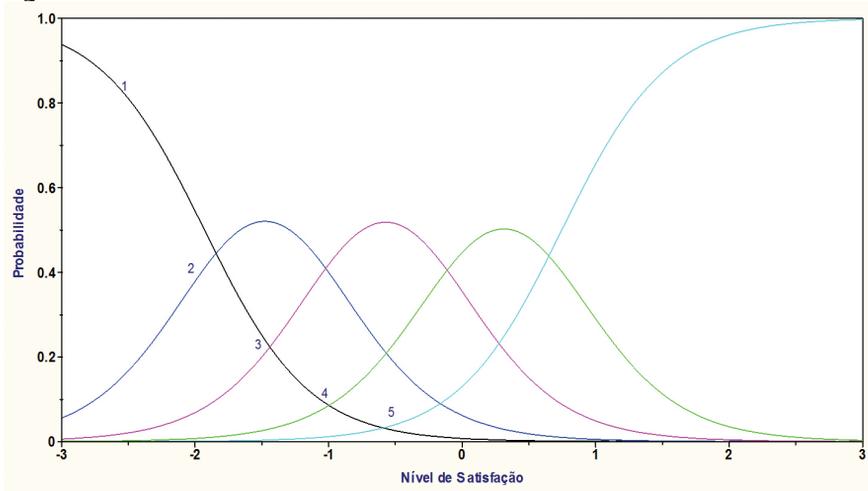
posicionados entre os valores  $-1,76$  e  $-0,22$  na escala, ou seja, como todos os valores médios dos  $b_{i,k}$  dos itens são negativos, conclui-se que todos os itens foram avaliados mais como satisfatórios do que insatisfatórios.

O grau de discriminação de um item (o valor de “ai”) determina a “qualidade” do item, isto é, quanto maior o valor de “ai”, maior o grau de discriminação do item. Observa-se, na Tabela 2, que os parâmetros de discriminação variaram de 1,15 a 2,63, indicando que todos os itens possuem um poder de discriminação satisfatório, já que os itens com discriminação baixa foram removidos. Por exemplo, o item 7 (clareza do professor) possui o maior valor de ai enquanto que o item 16 (disponibilidade das referências bibliográficas) possui o menor valor de ai.

Para uma interpretação mais detalhada da CCI, utilizou-se o item 5 ( $a_5 = 2,54$ ;  $b_{5,1} = -1,93$ ;  $b_{5,2} = -1,02$ ;  $b_{5,3} = -0,12$ ;  $b_{5,4} = 0,75$ ), cuja CCI é apresentada na Figura 5.

A Figura 5 apresenta o gráfico do MRG para o item 5, considerando-se o nível de satisfação entre  $-3$  e  $3$ , que é o intervalo em que praticamente todos os indivíduos estarão situados na escala (0,1). Este item possui a seguinte descrição: “Capacidade de estímulo/motivação utilizada pelo professor”. Indivíduos com satisfação nesta escala, entre  $-3,0$  e  $-1,8$ , têm maior probabilidade de responder a categoria 1 (muito insatisfeito); indivíduos com grau de satisfação entre  $-1,8$  e

**Figura 5** - CCI do Item 5



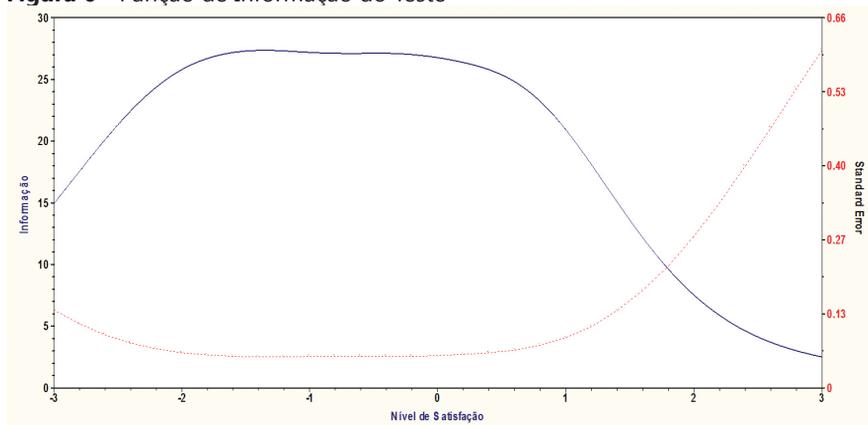
Fonte: Elaborado pelos autores (2013).

-1,0 têm maior probabilidade de responder a categoria 2 (insatisfeito); indivíduos com grau de satisfação entre -1,0 e 0,0 têm maior probabilidade de responder a categoria 3 (indiferente); indivíduos com grau de satisfação entre 0,0 e 0,7 têm maior probabilidade de responder a categoria 4 (satisfeito), e indivíduos com grau de satisfação maior que 0,7 têm maior probabilidade de responder a categoria 5 (muito satisfeito). Esse é um exemplo de item que funciona adequadamente para avaliar a satisfação dos alunos, pois consegue discriminar os alunos entre as categorias de satisfação.

A Figura 6 apresenta a Função de Informação Total (FIT) do instrumento de medida. Observa-se que o instrumento de medida tem maior informação no intervalo entre -3 e 1, o que corrobora com a posição dos itens na escala, segundo esse modelo, os quais situam-se entre -1,76 e -0,22. Isso significa que esse questionário de avaliação é mais adequado para medir a satisfação de indivíduos que possuem nível de satisfação entre -3 e 1. Além disso, os itens são, por estarem posicionados na região negativa da escala (0,1), “fáceis”, ou seja, a maioria dos indivíduos estará satisfeito com eles, como foi observado na Tabela 2.

A Tabela 3 apresenta a distribuição de frequência do traço latente, ou seja, do nível de satisfação. Observa-se que os dados encontram-se pouco dispersos e que nenhum respondente teve nível de satisfação estimado abaixo de -3 ou acima de 3 na escala (0, 1). O nível de satisfação médio foi de 0,003, com desvio padrão igual a 0,94, sendo que a maioria dos valores está localizada entre -2 e 2. O erro

**Figura 6** - Função de Informação do Teste



Fonte: Elaborado pelos autores (2013).

padrão das estimativas do traço latente variou entre 0,17 e 0,48, indicando uma boa precisão das estimativas. Observa-se que 50,4% dos respondentes possuem satisfação acima do valor da média da escala (zero).

Após a estimação dos parâmetros dos itens e da satisfação dos respondentes, estabeleceram-se as categorias âncoras dos itens e os níveis âncoras, com base nas condições mencionadas na Seção 3. O posicionamento dos itens foi realizado com os resultados da calibração dos itens. Os níveis âncoras foram estabelecidos

**Tabela 3** - Distribuição de Frequência dos Valores do Traço Latente

Intervalo	Quantidade	Percentual
-3,00 a -2,01	10	2,0%
-2,00 a -1,01	64	12,8%
-1,00 a -0,01	174	34,8%
0,00 a 0,99	187	37,4%
1,00 a 1,99	54	10,8%
2,00 a 3,00	11	2,2%

Fonte: Elaborado pelos autores (2013).

em função do desvio padrão da escala utilizada na calibração dos itens, que é igual a 1 (um), no intervalo entre -3 e 3. A Tabela 4 apresenta as categorias âncoras de cada nível âncora. Observa-se que, conforme os resultados de Mafra (2010), não é possível caracterizar a categoria mais baixa, no caso, “Muito Insatisfeito”, como categoria âncora, segundo o procedimento adotado. No entanto, pode-se afirmar que aqueles que estão posicionados abaixo do nível âncora mais baixo, estão muito insatisfeitos com todos os itens. Observa-se, ainda, que, dentre os 21 itens avaliados, apenas o item 16 não possui nenhuma categoria âncora.

**Tabela 4** – Categorias âncoras dos Itens e Níveis âncoras

Categoria\Nível	-3	-2	-1	0	1	2	3
2 – Insatisfeito	3, 12	1, 2, 4, 8, 9, 10, 13, 17, 21	5, 6, 18, 19, 20, 22, 23	-	-	-	-
3 – Parcialmente Satisfeito	-	12	2, 7, 8, 9, 10, 13, 15, 17, 21	19, 20, 23	-	-	-
4 – Satisfeito	-	-	-	1, 2, 4, 7, 8, 10, 12, 13	19, 20	-	-
5 – Muito Satisfeito	-	-	-	-	1, 2, 4, 5, 7, 8, 10, 12, 13	19, 20, 22	-

Fonte: Elaborado pelos autores (2013).

A interpretação de um indivíduo posicionado na escala por meio do valor estimado do seu traço latente é feita em relação ao nível em que ele se encontra. Com base na descrição dos itens, os níveis da escala de satisfação podem ser definidos assim:

- Nível –3: Alunos posicionados nesse nível estão insatisfeitos em relação ao domínio do assunto apresentado pelo professor da disciplina e às informações ou esclarecimentos quanto ao planejamento da disciplina. No estudo realizado, 2,0% dos alunos da amostra estão posicionados nesse nível.
- Nível –2: Alunos posicionados nesse nível estão insatisfeitos em relação aos seguintes aspectos: disponibilidade do professor em responder às solicitações dos alunos; confiança inspirada pela postura do professor da disciplina; capacidade de síntese do conteúdo pelo professor da disciplina; apresentação da teoria e aplicação prática dos temas abordados; possibilidade de contatar o professor fora da sala de aula; forma de avaliação proposta pelo professor da disciplina; estruturação com que as aulas são apresentadas; profundidade/desenvolvimento dos temas em relação aos objetivos da disciplina, e entendimento, participação e acompanhamento dos assuntos abordados. Mas esses alunos também estão parcialmente satisfeitos em relação às informações ou aos esclarecimentos quanto ao planejamento da disciplina. No estudo realizado, 12,8% dos alunos da amostra estão posicionados nesse nível.
- Nível –1: Alunos posicionados nesse nível estão insatisfeitos em relação aos seguintes aspectos: capacidade de estímulo/motivação utilizada pelo professor; capacidade do professor em manter o aluno atento e interessado na aula; aplicabilidade/aprendizagem dos temas abordados; interação curricular do curso com a disciplina; afinidade entre sua ocupação profissional e os conhecimentos recebidos; melhoria no seu desempenho profissional, gerando resultados positivos, e atendimento da disciplina às suas expectativas. Esses alunos também estão parcialmente satisfeitos em relação a confiança inspirada pela postura do professor da disciplina; clareza apresentada pelo professor ao apresentar os conteúdos; apresentação da teoria e aplicação prática dos temas abordados; possibilidade de contatar o professor fora da sala de aula; forma de avaliação proposta pelo professor da disciplina; estruturação com que as aulas são apresentadas; qualidade dos recursos didáticos e demais materiais relacionados à disciplina; profundidade/desenvolvimento dos temas em relação aos objetivos da disciplina, e entendimento, participação e acompanhamento dos assuntos

abordados. No estudo realizado, 34,8% dos alunos da amostra estão posicionados nesse nível.

- **Nível 0:** Alunos posicionados nesse nível estão parcialmente satisfeitos em relação aos seguintes aspectos: interação curricular do curso com a disciplina; afinidade entre sua ocupação profissional e os conhecimentos recebidos, e atendimento da disciplina às suas expectativas. Esses alunos também estão parcialmente satisfeitos com a disponibilidade do professor em responder às solicitações dos alunos; a confiança inspirada pela postura do professor da disciplina; a capacidade de síntese do conteúdo pelo professor da disciplina; a clareza apresentada pelo professor ao apresentar os conteúdos; a apresentação da teoria e aplicação prática dos temas abordados; a forma de avaliação proposta pelo professor da disciplina; as informações/esclarecimento quanto ao planejamento da disciplina, e a estruturação com que as aulas são apresentadas. No estudo realizado, 37,4% dos alunos da amostra estão posicionados nesse nível.
- **Nível 1:** Alunos posicionados nesse nível estão satisfeitos em relação à interação curricular do curso com a disciplina e à afinidade entre sua ocupação profissional e os conhecimentos recebidos. Esses alunos também estão muito satisfeitos em relação aos seguintes aspectos: disponibilidade do professor em responder às solicitações dos alunos; confiança inspirada pela postura do professor da disciplina; capacidade de síntese do conteúdo pelo professor da disciplina; capacidade de estímulo/motivação utilizada pelo professor; clareza apresentada pelo professor ao apresentar os conteúdos; apresentação da teoria e aplicação prática dos temas abordados; forma de avaliação proposta pelo professor da disciplina; informações/esclarecimento quanto ao planejamento da disciplina, e estruturação com que as aulas são apresentadas. No estudo realizado, 10,8% dos alunos da amostra estão posicionados nesse nível.
- **Nível 2:** Alunos posicionados nesse nível estão muito satisfeitos com a interação curricular do curso com a disciplina; a afinidade entre sua ocupação profissional e os conhecimentos recebidos, e a melhoria no seu desempenho profissional, gerando resultados positivos. No estudo realizado, 2,2% dos alunos da amostra estão posicionados nesse nível.
- **Nível 3:** Não foram encontradas categorias âncoras nesse nível. Na amostra utilizada neste estudo, não há nenhum aluno posicionado nesse nível.

## 6 Conclusões

Neste trabalho, foi utilizada a Teoria de Resposta ao Item como ferramenta para a avaliação da satisfação dos alunos das disciplinas de Estatística Básica e Estatística Geral da Universidade Federal de Santa Maria (UFSM). Foi aplicado o Modelo de Resposta Gradual (MRG) da TRI, de natureza unidimensional e cumulativa, por meio do qual foi criada uma escala de medida para a satisfação, na qual foram posicionados os itens de satisfação e os respondentes, numa mesma unidade de medida. Os níveis da escala criada foram identificados e caracterizados, sendo então feita uma avaliação da satisfação dos alunos das disciplinas de Estatística Básica e Estatística Geral da UFSM.

O resultado da análise fatorial mostrou que existe um fator dominante (satisfação do aluno) que caracteriza a unidimensionalidade do construto. Quanto aos parâmetros dos itens, na primeira análise, apenas dois itens mostraram-se inadequados. Posteriormente, esses itens foram removidos e a análise foi refeita, em que se observou que todos os itens apresentaram resultados adequados. Em relação ao parâmetro de posição “bk” das categorias dos itens, verificou-se que todos os valores médios dos “bk” foram negativos, concluindo-se, assim, que todos os itens foram avaliados mais como satisfatórios do que insatisfatórios. Dessa forma, se conclui que não há um item com o qual a maioria está insatisfeita. Já os parâmetros de discriminação “a” dos itens indicaram que todos os itens possuíam um poder de discriminação satisfatório. Por meio da FIT, foi possível verificar que os itens que compunham o questionário eram “fáceis”, ou seja, são itens nos quais a maioria dos indivíduos irá declarar que está satisfeita com estes. Também foram identificadas várias categorias âncoras de itens, num total de 53 categorias âncoras distribuídas em seis níveis âncoras, que foram devidamente caracterizados. Apenas um item não apresentou nenhuma categoria âncora. Foi possível identificar em que nível da escala de satisfação os alunos estão situados.

Pode-se concluir que o MRG da TRI foi adequado para a análise dos dados, a identificação de itens problemáticos, a construção de uma escala para a satisfação dos alunos e a identificação de níveis âncoras. Além disso, o uso do MRG permite obter mais informação das respostas dos indivíduos do que um modelo dicotômico permitiria, uma vez que leva em consideração a probabilidade de ele responder cada categoria de resposta e não somente se está ou não satisfeito, no caso de um modelo dicotômico.

Segundo a propriedade da invariância dos parâmetros dos itens, todos os itens continuarão com a mesma probabilidade de serem considerados satisfatórios ou

não, até que ocorra uma mudança prática que altere o valor dos seus parâmetros. Por exemplo, o item com maior probabilidade de ser considerado insatisfatório é o item 19, descrito por “Interação curricular do curso com a disciplina”. Poderia ser feita uma reformulação curricular nos cursos que têm disciplina de Estatística, a fim de mostrar a integração das ferramentas estatísticas com as demais disciplinas da grade curricular do curso, indicando quais as outras disciplinas que precisam utilizar ferramentas estatísticas, mostrando a importância da disciplina de Estatística no contexto curricular do curso. Na TRI, a escala deve ser construída para posicionar os parâmetros dos itens e ser utilizada ao longo do tempo. Isso funciona bem na avaliação educacional, em que um item que é difícil continua a ser difícil ao longo do tempo, caso ele não se torne conhecido do público. No entanto, quando há mudança nas propriedades psicométricas de algum item, este precisa ter seus parâmetros estimados novamente. Isso pode ocorrer na situação supracitada, em que houve uma suposta melhoria no processo e aquele item pode passar a ser mais satisfatório do que os outros, que não tiveram seus parâmetros alterados. É claro que isso só pode ser verificado por meio da aplicação dos itens a uma nova amostra de indivíduos. Nesse caso hipotético, apenas aquele item teria seus parâmetros modificados. Isso não significa que seria construída uma nova escala de tempos em tempos, mas que alguns itens podem sofrer mudanças nas suas propriedades psicométricas ao longo do tempo, devendo ter as suas estimativas atualizadas. Eventualmente, isso pode ocorrer também na avaliação educacional, quando um item, por exemplo, de Geografia, pode ter seus parâmetros alterados, devido à emancipação de um novo país ou ao surgimento de um novo bloco econômico.

O segundo item com maior probabilidade de ser considerado insatisfatório é o 20, descrito como “Afinidade entre sua ocupação profissional e os conhecimentos recebidos”, indicando que falta ao professor fazer uma conexão entre os conteúdos e a prática profissional do aluno.

Outros dois itens com alta probabilidade de serem considerados insatisfatórios tratam da relação entre a disciplina e o atendimento às expectativas do aluno e à sua melhoria profissional. Isto indica que há mais chance de insatisfação dos alunos em questões que relacionam a disciplina de Estatística e a sua área de atuação profissional.

Já o item 3, descrito como “O domínio do assunto apresentado pelo professor da disciplina”, foi o item com maior probabilidade de ser considerado satisfatório, mostrando que, na visão do aluno, o professor tem bom domínio dos conteúdos ensinados em sala de aula. Além disso, os cinco itens com maior probabilidade

de serem considerados satisfatórios pelos alunos tratam do conhecimento, da capacidade de planejamento e da disponibilidade que o professor apresenta ao ministrar a disciplina de Estatística.

Os resultados obtidos permitem que os responsáveis pelo ensino da Estatística possam observar os aspectos da satisfação ou da insatisfação dos alunos, para detectar em que é possível melhorar os procedimentos para que o ensino seja cada vez mais satisfatório.

Para estudos futuros, sugere-se:

- Aumentar a quantidade de itens do tipo “difícil”, ou seja, mais difíceis de serem considerados satisfatórios pelo aluno, para melhorar a estimativa do traço latente dos alunos satisfeitos;
- Aumentar o tamanho da amostra para melhorar a estimativa dos parâmetros dos itens;
- Diminuir o erro padrão;
- Verificar o desempenho por meio de estudos de simulação.

## 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.

AYALA, R. J. *The theory and practice of Item Response Theory*. New York: The Guilford Press, 2009.

BAKER, F. B. *Item response theory parameter estimation techniques*. New York: Marcel Dekker, 1992.

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*, Washington, v. 17, p. 191-204, 1992.

BONAFÉ, F. S. S.; LOFFREDO, L. C. M.; CAMPOS, J. A. D. B. Atitudes em relação à Bioestatística de discentes e docentes da Faculdade de Ciências Farmacêuticas de Araraquara-UNESP. *Revista de Ciências Farmacêuticas Básica e Aplicada*, Araraquara, v. 31, n. 2, p. 143-147, 2010.

BORTOLOTTI, S. L. V. *Aplicação de um modelo de desdobramento graduado generalizado da teoria da resposta ao item – TRI*. 2003. 107 f. Dissertação (Mestrado em Engenharia de Produção)-Programa de Pós-Graduação em Engenharia de Produção, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2003.

BORTOLOTTI, S. L. V.; ANDRADE, D. F. Aplicação de um modelo de desdobramento graduado generalizado – GGUM da teoria da resposta ao item. *Estudos em Avaliação Educacional*, São Paulo, v. 18, n. 37, p. 157-188, 2007.

COSTA, M. B. F. *Técnica derivada da teoria da resposta ao item aplicada ao setor de serviços*. 2001. Dissertação (Mestrado em Ciências)-Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, 2001.

COSTA, M. B. F.; CHAVES NETO, A. Aplicação da Teoria da Resposta ao Item (TRI) na avaliação do grau de satisfação do consumidor em um serviço específico. In: REUNIÃO REGIONAL DA ASSOCIAÇÃO BRASILEIRA DE ESTATÍSTICA 34., 2002, Fortaleza - CE. *Resumos...* São Paulo: ABE - Associação Brasileira de Estatística, 2002. p. 31-31.

DAVOK, D. F. Qualidade em educação. *Avaliação: Revista da Avaliação da Educação Superior*, Sorocaba, v. 12, n. 3, p. 505-513, 2007.

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

HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. *Fundamentals of item response theory*. Newbury Park: Sage, 1991.

LORENZO-SEVA, U.; FERRNADO, P. J. *Manual of the Program Factor*. Tarragona: Departament de Psicologia, Universitat Rovira i Virgili, 2013. Disponível em: <<http://psico.fcep.urv.cat/utilitats/factor/documentation/Manual-of-the-Factor-Program-v92.pdf>>. Acesso em: 30 maio 2014.

MAFRA, P. M. R. *Proposta de uma sistemática para a modelagem de risco de crédito sob a perspectiva da teoria da criação do conhecimento: uma abordagem da teoria da resposta ao item*. 2010. Tese (Doutorado em Engenharia e Gestão do Conhecimento)-Universidade Federal de Santa Catarina, Florianópolis, 2010.

MANTOVANI, D. M. N.; GOUVÊA, M. A. Estatística aplicada à Administração: um estudo de atitudes versus desempenho do aluno. *Revista Iberoamericana de Educacion*, Madrid, n. 58/2, p. 1-12, 2012.

MATTEUCCI, M.; STRACQUALURSI, L. Student assessment via Graded Response Model. *Statistica*, Bologna, anno LXVI, n. 4, p. 435-447, 2006.

MENDES, E. L. *Uma metodologia para avaliação da satisfação do consumidor com os serviços prestados pelas distribuidoras de energia elétrica*. 2006. 148 f. Tese (Doutorado em Engenharia Elétrica)-Programa de Pós-Graduação em Engenharia Elétrica, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006.

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.

MURAKI, E.; BOCK, R. D. *Parscale: IRT Based Test Scoring and Item Analysis for Graded Open-Ended Exercises and Performance Tasks*. Chicago: Scientific Software Inc., 1997.

PAULINS, V. A. An analysis of customer service quality to college students as influenced by customer appearance through dress during the in-store shopping process. *Journal of Retailing and Consumer Services*, v. 12, n. 5, p. 345-355, 2005. <http://dx.doi.org/10.1016/j.jretconser.2004.11.003>.

REISE, S. P.; YU, J. Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement*, Washington, v. 27, n. 2, p. 133-144, 1990. <http://dx.doi.org/10.1111/j.1745-3984.1990.tb00738.x>.

SAMEJIMA, F. A general model for free response data. *Psychometrika Monograph Supplement*, Stacks, n. 18, 1972.

\_\_\_\_\_. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, Stacks, n. 17, 1969.

\_\_\_\_\_. Graded response model. In: VAN DER LINDEN, W. J.; HAMBLETON, R. K. (Ed.). *Handbook of modern item response theory*. New York: Springer, 1997. p. 85-100.

SANTOS, C. R.; FERREIRA, M. C. L. *Avaliação Educacional: um olhar reflexivo sobre sua prática*. São Paulo: Avercamp, 2005.

SOARES, T. M. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. *Pesquisa Operacional*, Rio de Janeiro, v. 25, n. 1, 2005.

SOUZA, S. A.; REINERT, J. N. Avaliação de um curso de ensino superior através da satisfação/insatisfação discente. *Avaliação: Revista da Avaliação da Educação Superior*, Sorocaba, v. 15, n. 1, p. 159-176, 2010.

THISSEN, D. *Multilog 7.0. Multiple, categorical item analysis and test scoring using item response theory*. Lincolnwood: Scientific Software International, 2003.

\_\_\_\_\_. *Multiloguser's guide: multiple, categorical item analysis and testing score using Item Response Theory*. Chicago: Scientific Software, 1991.

TIMMERMAN, M. E.; LORENZO-SEVA, U. Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, Washington, v. 16, n. 2, p. 209-220, 2011. <http://dx.doi.org/10.1037/a0023353>. PMID:21500916

TOIT, M. *IRT from SSI: Bilog-Mg, Multilog, Parscale, Testfact*. Chicago: Scientific Software, 2003.

VIEIRA, K. M.; KUNKEL, F. I. R.; RIGHI, M. B. Desenvolvimento e validação do modelo PLS-PM aplicado à satisfação dos alunos de uma instituição de ensino superior. *Perspectiva*, Erechim, v. 36, n. 134, p. 109-124, 2012.

VIEIRA, K. M.; MILACH, F. T.; HUPPES, D. Equações estruturais aplicadas à satisfação dos alunos: um estudo no curso de ciências contábeis. *Revista Contabilidade & Finanças*, São Paulo, v. 19, n. 48, p. 65-76, 2008.

ZANELLA, A. *Diagnóstico da qualidade do ensino-aprendizagem e satisfação dos alunos nas disciplinas de estatística da UFSM*. 2008. 119 f. Dissertação (Mestrado em Engenharia de Produção)-Programa de Pós-Graduação em

Engenharia de Produção, Centro de Tecnologia, Universidade Federal de Santa Maria, Santa Maria, 2008.

ZANELLA, A.; LOPES, L. F. D.; SEIDEL, E. J. Diagnóstico do ensino-aprendizagem e satisfação dos alunos nas disciplinas de estatística da UFSM. *Revista GEPROS: Gestão da Produção, Operações e Sistemas*, Bauru, v. 4, n. 3, p. 123-140, 2009.

ZANELLA, A.; SEIDEL, E. J.; LOPES, L. F. D. Validação de questionário de satisfação usando análise fatorial. *Revista INGEPRO: Inovação, Gestão e Produção*, Santa Maria, v. 2, n. 12, p. 102-112, 2010.

## **Student satisfaction evaluation through the Gradual Response Model of Item Response Theory**

### **Abstract**

*The aim of this study was to use the Item Response Theory (IRT) as a tool to assess the satisfaction of students of the Basic and General Statistics course at the Federal University of Santa Maria (UFSM). Through the application of the Graded Response Model (GRM), a measurement scale for satisfaction was created, where the items of satisfaction and the respondents were positioned in the same unit of measure. The results showed that the GRM of the IRT was suitable for the analysis of data, allowing the identification of problematic items, the construction of a scale for student satisfaction, and the identification of anchor levels. The findings will allow the staff responsible for the teaching of Statistics at the institution to identify how procedures can be improved so that teaching becomes increasingly satisfactory.*

**Keywords:** *Item Response Theory. Gradual Response Model. Satisfaction evaluation. Students.*

## **Evaluación de satisfacción de alumnos a través del Modelo de Respuesta Gradual de la Teoría de Respuesta al Ítem**

### **Resumen**

*El objetivo de este estudio fue utilizar la Teoría de Respuesta al Ítem (TRI) como una herramienta para la evaluación de satisfacción de los alumnos de las asignaturas de Estadística básica y Estadística general de la Universidad*

*Federal de Santa Maria (UFSM). Se aplicó el Modelo de Respuesta Gradual (MRG) a través del cual se creó una escala de medida de satisfacción, donde los ítems de satisfacción junto con los encuestados se colocaron en la misma unidad de medida. Los resultados mostraron que el MRG del TRI eran adecuados para el análisis de los datos, permitiendo la identificación de ítems problemáticos, la construcción de una escala de satisfacción de alumnos y la identificación de niveles ancorados. Los resultados permitirán que los responsables de la enseñanza de estadística en la institución puedan identificar los procedimientos más adecuados para que la enseñanza sea cada vez más satisfactoria.*

**Palabras-clave:** Teoría de Respuesta al Ítem. Modelo de Respuesta Gradual. Evaluación de satisfacción. Alumnos.

## Informações dos autores

Fernando de Jesus Moreira Junior: Doutor em Engenharia de Produção. Contato: fmjunior@smail.ufsm.br

Andreia Zanella: Mestre em Engenharia de Produção. Contato: andreia\_zanella@yahoo.com.br

Luis Felipe Dias Lopes: Doutor em Engenharia de Produção. Contato: lflopes67@yahoo.com.br

Enio Júnior Seidel: Doutor em Estatística e Experimentação Agropecuária. Contato: ejrseidel@hotmail.com