

Avaliação do nível de satisfação de alunos de uma instituição de ensino superior: uma aplicação da Teoria da Resposta ao Item

Consumer satisfaction and item response theory: creating a measurement scale

Silvana Ligia Vincenzi Bortolotti¹
Fernando de Jesus Moreira Junior²
Antonio Cezar Bornia³
Afonso Farias de Sousa Júnior⁴
Dalton Francisco de Andrade⁵

Resumo: O tema deste trabalho é a utilização da Teoria da Resposta ao Item (TRI) como ferramenta de avaliação de aspectos organizacionais específicos. O objetivo é aplicar um modelo cumulativo da TRI para criar uma medida de satisfação de alunos com seus cursos, avaliando também a satisfação no ensino e criando uma escala de medida. Muito utilizada nas áreas educacional e psicológica, a Teoria de Resposta ao Item (TRI) é uma abordagem utilizada para avaliação de traço latente e fornece informações adicionais além daquelas obtidas a partir de técnicas psicométricas clássicas. Neste estudo, foi utilizado o Modelo de Resposta Gradual de Samejima. Um instrumento foi aplicado a 525 alunos e gerou-se uma escala para mensurar a satisfação com seus cursos. Os resultados obtidos demonstram a eficácia da teoria na obtenção da medida de satisfação, uma vez que coloca tanto os itens como os indivíduos numa mesma escala. Também se concluiu que esta teoria pode ser valiosa na avaliação da satisfação do consumidor e de vários outros fenômenos organizacionais. Seus resultados podem auxiliar o tomador de decisão do empreendimento para a retificação de fluxos, processos e procedimentos, assim como para a geração de incremento de eficiência e eficácia nos eventos empresariais. Por fim, as informações obtidas das análises podem atuar na elaboração e/ou avaliação do planejamento institucional, principalmente nas elaborações de objetivos, ações e metas institucionais.

Palavra-chave: Gestão. Medida do nível de satisfação. Teoria de Resposta ao Item. Tomada de decisão.

Abstract: *Today, people have increasingly demanded more from the state and enterprises. Consumer satisfaction is not an organizational option, but rather a matter of survival for any institution. The quest for measurement of consumer satisfaction has been ongoing in many areas of research, and researchers have concentrated efforts to demonstrate the psychometric quality of their measurements. However, the techniques employed by these commitments have not kept pace with the advances in psychometric theory and methods. The Item Response Theory (IRT) is an approach used for assessing latent trait. It is commonly used in educational and psychological tests and provides additional information beyond that obtained from classic psychometric techniques. This article presents a model of cumulative application of item response theory to measure the extent of students' satisfaction with their courses by creating a measurement scale. The Graded Response Model was used. The results demonstrate the effectiveness of this theory in measuring satisfaction since it places both items as individuals on the same scale. This theory may be valuable in the evaluation of customer satisfaction and many other organizational phenomena. The findings may help the decision maker of an enterprise with the correction of flows, processes, and procedures, and, consequently, it may help generate increased efficiency and effectiveness in daily tasks and in event management business. Finally, the information obtained from the analysis can play a role in the development and/or evaluation of institutional planning.*

Keywords: *Management. Measurement of satisfaction. Item response theory. Decision making.*

¹ Universidade Federal do Paraná – UFPR, Av. Brasil, 4232, CP 271, CEP 85884-000, Medianeira, PR, Brasil, e-mail: sligie@globom.com

² Departamento de Estatística, Universidade Federal de Santa Maria – UFSM, Av. Roraima, 1000, Cidade Universitária, Camobi, CEP 97105-900, Santa Maria, RS, Brasil, e-mail: fmjunior@smail.ufsm.br

³ Departamento de Engenharia de Produção, Universidade Federal de Santa Catarina – UFSC, Campus Universitário, CP 476, Trindade, SC, Brasil, e-mail: cezar@deps.ufsc.br

⁴ Universidade da Força Aérea – UNIFA, Av. Marechal Fontenelle, 1200, CEP 21740-000, Rio de Janeiro, RJ, Brasil, e-mail: bvfire@gmail.com

⁵ Departamento de Informática e Estatística, Universidade Federal de Santa Catarina – UFSC, Campus Universitário, CP 476, Trindade, SC, Brasil, e-mail: dandrade@inf.ufsc.br

1 Introdução

Os últimos vinte anos premiaram os Estados, as empresas e a sociedade com inúmeros novos serviços, invenções e inovações para aplicação direta e indireta na vida dos consumidores. As empresas passaram a olhar com mais cuidado para o pós-venda e a continuar valorizando o cliente – mesmo após a compra do produto ou serviço. Fidelizar o consumidor é tarefa árdua e exige competência específica para lidar com este quesito nas dimensões locais, regionais e globais.

Frente às mudanças e transformações técnicas e socioeconômicas, é importante encontrar novos métodos que gerem condições apropriadas para o desenvolvimento das diversas áreas demandadas. A avaliação de procedimentos ou de processos, nesse caso, apresenta-se como ferramenta para compreensão dos problemas existentes e identificação de causas, ao mesmo tempo em que poderá permitir que novas medidas corretivas ou alternativas possam aumentar a eficácia e a eficiência dessas transformações. Avaliação, no sentido de medição, compreende tanto a avaliação de desempenho, como a avaliação da satisfação por determinado produto.

A avaliação da gestão institucional ajuda a perceber os óbices socioadministrativos, uma vez que proporciona a obtenção de informações necessárias para identificar e compreender as causas dos sucessos e fracassos dos empreendimentos, tanto nas questões individuais quanto naquelas relativas às equipes ou divisões (locais, regionais, globais etc.). As informações obtidas por meio da avaliação podem auxiliar na elaboração e revisão do planejamento estratégico e em outras tomadas de decisão. Mattar (2005), relativamente ao processo de administração de *marketing*, salienta que a disponibilidade de informações a respeito do mercado, do meio ambiente, do macroambiente de *marketing* e da própria empresa, em quantidade e com qualidade, é fundamental para a redução do risco na tomada de decisão. Evrard (1995) ratifica essa posição e evidencia que as informações obtidas por meio da avaliação da satisfação são utilizadas para traçar a estratégia de melhoria da qualidade e para implantar ações que efetivamente melhorem a satisfação, assim como a compreensão dos processos de formação da satisfação é empregada como ferramenta para a tomada de decisão.

As decisões gerenciais que são amparadas em fatos e dados, e não meramente na intuição, tendem a ser mais significativas na geração de resultados financeiros positivos para a empresa e ainda é importante verificar a satisfação dos consumidores e confrontar o desempenho da empresa com os concorrentes (PERILLO, 2000).

Empresas que promovem educação buscam a satisfação de seus alunos nos cursos que promovem. Este trabalho tem como objetivo utilizar a Teoria de Resposta ao Item como ferramenta de avaliação de

aspectos organizacionais específicos, com aplicação do Modelo de Resposta Gradual de natureza cumulativa da TRI, na avaliação da satisfação de alunos quanto ao curso que frequentam, desenvolvendo uma escala de medida, em que se localizam os itens de satisfação e os respondentes numa mesma unidade de medida.

A avaliação de medida aplicada ao ensino para averiguar a satisfação de alunos quanto ao curso é essencial para as instituições de ensino, pois proporciona uma reflexão sobre os métodos empregados, bem como fornece subsídios para a melhoria da qualidade do ensino dos cursos oferecidos pela instituição. A pesquisa foi realizada com alunos da Universidade Tecnológica Federal do Paraná, na cidade de Medianeira.

Enfatiza-se que a Teoria de Resposta ao Item compreende modelos para avaliar traços latentes. Esses modelos apresentam formas de representar a relação entre o traço latente de um indivíduo, na área de conhecimento a ser avaliada ou verificada, o qual não pode ser observado diretamente, e a probabilidade dele dar uma certa resposta a um item. (ANDRADE; TAVARES; VALLE, 2000).

O artigo está estruturado da seguinte forma: no início, apresenta-se um breve referencial teórico sobre satisfação e a Teoria da Resposta ao Item, explicitando-se conceitos utilizados neste trabalho. Na sequência, foi abordado o método e apresentam-se os resultados, em que se destacam a coleta de dados, análise da dimensionalidade, estimação dos parâmetros, testes de ajustes, criação da escala de satisfação e posição dos respondentes na escala desenvolvida. As discussões e conclusão encerram o artigo.

2 Satisfação: conceitos e compreensão

A satisfação do consumidor vem sendo abordada por diversas áreas. Atualmente, na era da qualidade total, da globalização da economia e da administração estratégica, este conceito se dissemina em toda a organização.

Um importante critério para avaliar a qualidade das empresas é a satisfação dos consumidores. Este item representa 30% na avaliação geral para o prêmio Malcolm Baldrige nos Estados Unidos, como também é levado em consideração na avaliação do Prêmio Deming no Japão, e ainda refere-se a 17% da pontuação total na avaliação para o Prêmio Nacional de Qualidade - PNQ, no Brasil (Rossi; Slongo, 1997).

Além disso, a importância do estudo da satisfação também pode ser evidenciada pelo fato de que vários países estão realizando a avaliação geral de satisfação de seus Clientes - Cidadãos e utilizando esse índice como uma forma de mensurar o desempenho de suas economias. O primeiro país a medir a satisfação

de seus consumidores foi a Suécia, com o SCSB (*Swedish Customer Satisfaction Barometer*), em 1989. Em seguida, a Alemanha realizou suas análises em 1992, com o *Deutsche Kundenbarometer* (DK), os Estados Unidos em 1994 implementaram o seu ACSI (*American Customer Satisfaction Index*), a Nova Zelândia e Taiwan iniciaram as suas avaliações em 1996 e a União Europeia recomendou aos seus membros que fizessem o mesmo (MEDEIROS; FARIA, 2003; FORNELL et al., 1996).

A satisfação do consumidor é relevante para as empresas, porque seus lucros estão relacionados diretamente com a satisfação do consumidor. Os lucros são maximizados quando o consumidor está satisfeito e permanece leal ao seu produto. Isso é corroborado por Oliver (2009), quando enfatiza que a satisfação do consumidor vem sendo vista como fator crucial para o sucesso dos mais variados tipos de organizações, sobretudo por influenciar diretamente a lealdade à marca, repetição de compras, comunicação boca a boca (positiva), lucratividade e participação de mercado. Seguramente, toda a organização que deseja prosseguir com sucesso no mercado gostaria de saber como influenciar e controlar essas dimensões, especialmente porque o maior escopo das empresas é a maximização do lucro em médio e longo prazos. Assim, quando o assunto é lucratividade, essa relação torna a satisfação do consumidor um aspecto prioritário a ser levado em consideração no planejamento organizacional.

Clientes satisfeitos são mais fiéis à marca, com alto índice de retorno para fazer novos negócios. É dessa forma que a empresa consegue um percentual de fechamento superior a seus concorrentes. Além disso, o cliente fiel não somente volta como indica a companhia e a torna referência, pois suas opiniões são mais consideradas do que toda a propaganda paga e sem nenhum custo adicional.

Os benefícios proporcionados pela pesquisa de satisfação do consumidor, como: a) percepção mais positiva dos consumidores quanto à empresa; b) relações de fidelidade com os consumidores, baseadas em ações corretivas; c) dados precisos e atualizados quanto às necessidades dos consumidores; e d) confiança desenvolvida em função de maior aproximação com o consumidor, fazem com que os recursos destinados a aumentar a satisfação do consumidor sejam considerados como investimentos e não como despesas.

Desse modo, o consumidor passou a ser considerado peça primordial para o sucesso e desenvolvimento da empresa, pois a sua satisfação representa lucro e permanência no mercado (HASTREITER, 1998).

Na literatura, encontram-se vários conceitos de satisfação. Dentre eles, destacam-se os apresentados por Kotler (1994, p. 50), “[...] satisfação é o nível de sentimento de uma pessoa resultante da comparação

do desempenho (ou resultado) de um produto em relação a suas expectativas.”, e Mowen (1995, p. 511), “[...] é a atitude geral referente a um produto ou serviço depois de sua aquisição e uso. É a avaliação pós-compra resultante de uma seleção de compra específica.”.

O termo “avaliação” pressupõe a existência de um parâmetro. Para a maioria dos autores, o grau de satisfação deriva de uma comparação realizada pelo consumidor, comparação essa que confronta o resultado da transação com uma referência anteriormente existente (SPRENG; MACKENZIE; OLSHAVSKY, 1996). O resultado desse processo comparativo pode ser positivo, negativo ou neutro, gerando satisfação ou insatisfação.

A satisfação é definida como um estado psicológico (EVRARD, 1995) e não pode ser observada diretamente. Ela é considerada uma variável latente e referida como traço latente ou construto, devendo ser inferida a partir da observação de variáveis secundárias que estejam relacionadas a ela. Dessa forma, tem-se procurado desenvolver escalas para medi-la. Frequentemente, essa medida é feita por levantamentos periódicos, empregando-se questionários de pesquisa que abordam questões qualitativas e quantitativas.

A análise para obter a expectativa, a percepção e o grau de satisfação dos consumidores, assim como a percepção da administração, é realizada por meio de técnicas estatísticas, como a análise fatorial (utilizando, principalmente, o método de componentes principais) e a análise de regressão.

Assim, medir a satisfação de um consumidor é uma questão de sobrevivência para qualquer organização. A empresa necessita liderar o processo para a busca contínua da satisfação, pois é um dos resultados mais importantes de um atendimento de boa qualidade. É necessário, então, haver um instrumento que indique se o caminho trilhado está ou não em conformidade com a expectativa do consumidor.

3 Teoria da Resposta ao Item e escala de medida

A TRI foi desenvolvida principalmente para suprir limitações da Teoria Clássica de Medidas (TCM). Uma das limitações da TCM é que o instrumento de medida depende das características dos respondentes que se submetem ao teste ou ao questionário (Hambleton; Swaminatan; ROGERS, 1991).

Inicialmente as aplicações da TRI foram desenvolvidas no âmbito educacional e, posteriormente, ela foi encontrando espaço em outras áreas, especialmente na área de gestão e produção com os trabalhos de Alexandre et al. (2002) e Vargas et al. (2008).

Dentre as vantagens da Teoria de Resposta ao Item sobre a Teoria Clássica de Medidas, destaca-se que a TRI:

a) Permite fazer comparações entre traço latente de indivíduos 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 (ANDRADE; TAVARES; VALLE, 2000);

b) Possibilita uma melhor análise de cada item que forma o instrumento de medida, pois considera suas características específicas de construção de escalas;

c) 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, facilitando a interpretação da escala gerada e permitindo conhecer quais itens estão produzindo informação ao longo da escala (EMBRETSON; REISE, 2000);

d) 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. Este princípio é um dos grandes benefícios da TRI (Hambleton; Swaminatan; ROGERS, 1991).

A TRI surgiu como uma forma de considerar cada item particularmente, sem relevar os escores totais. Portanto, as conclusões não dependem do teste ou questionário como um todo, como é feito na TCM, mas sim de cada item que o compõe (ARAÚJO; ANDRADE; BORTOLOTTI, 2009).

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, considere-se o construto nível de satisfação: uma análise feita por meio da TRI pode estimar o nível de satisfação do respondente (isto é, um parâmetro do indivíduo) e também os parâmetros dos itens, de modo a criar uma escala de medida do nível de satisfação.

Os modelos utilizados na TRI requerem dois pressupostos relevantes (EMBRETSON; REISE, 2000), a Curva Característica do Item – CCI e a independência local ou dimensionalidade.

A forma de uma curva característica do item descreve como a mudança do traço latente relaciona-se com a mudança na probabilidade de uma resposta específica (EMBRETSON; REISE, 2000).

A independência local é obtida quando, controlados pelo nível do traço latente, os itens do teste são independentes. Assim, a probabilidade de responder a um item é precisamente determinada pelo nível do traço latente do respondente e não por suas respostas a outros itens do conjunto (EMBRETSON; REISE, 2000; ANDRADE; TAVARES; VALLE, 2000). A independência local é vista como consequência da correta determinação da dimensionalidade dos dados (LORD, 1980). A dimensionalidade pode ser verificada por meio de uma Análise Fatorial apropriada para

dados categorizados (EMBRETSON; REISE, 2000; ANDRADE; TAVARES; VALLE, 2000).

Os vários modelos existentes da TRI distinguem-se na forma matemática da função característica do item e/ou no número de parâmetros especificados no modelo. Todos os modelos podem conter um ou mais parâmetros relacionados aos itens e ao indivíduo (ANDRADE; BORTOLOTTI, 2007).

Foram desenvolvidos diversos modelos da Teoria da resposta ao item e a diferença principal entre estes modelos refere-se à suposição quanto ao relacionamento entre a seleção de opções de uma resposta e o nível do traço latente. Dentre os modelos desenvolvidos para itens com resposta gradual, destaca-se o Modelo de Resposta Gradual (MRG) de Samejima (1969, 1972, 1997). No MRG de Samejima, são estimados dois parâmetros relacionados com o item (EMBRETSON; REISE, 2000):

1º) $b_{i,k}$, parâmetro de dificuldade da k -ésima categoria do item i : este parâmetro é conhecido como o parâmetro “limiar”. O parâmetro de dificuldade refere-se ao nível do traço latente no qual a probabilidade de um indivíduo selecionar uma categoria de resposta (por exemplo, insatisfeito) ou uma categoria mais alta ordenada (por exemplo, pouco insatisfeito, satisfeito ou muito satisfeito) é 0,5. Assim, ele representa o ponto na escala do traço latente onde há 50% de probabilidade 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;

2º) a_i , parâmetro de discriminação do item 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 a discriminação entre os indivíduos nos diferentes níveis de traço latente.

Suponha que os escores das categorias de um item i sejam arranjados em ordem do menor para o maior e denotados por $k = 0; 1; \dots; m_i$, 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 de dois parâmetros (ML2) (ANDRADE; TAVARES; VALLE, 2000), conforme a Equação 1:

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

com $i = 1, 2, \dots, I, j = 1, 2, \dots, n$, e $k = 0, 1, \dots, m_i$, em que:

- $b_{i,k}$ e a_i já foram definidos anteriormente,
- θ_j representa o traço latente do respondente j , neste caso, o nível de satisfação do respondente em relação ao curso que frequenta, e

- D é uma constante de escala igual a 1; utiliza-se o valor 1,7 quando se quer que os valores da função logística aproximem-se da função ogiva normal.

No MRG de Samejima, a discriminação de uma categoria específica de resposta depende tanto do parâmetro de discriminação a , comum a todas as categorias do item, quanto da distância das categorias de dificuldade adjacentes. Necessariamente, deve haver uma ordenação entre o nível de dificuldade das categorias de um dado item, de acordo com a classificação de seus escores, ou seja, $b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i}$.

Desse modo, a probabilidade de um indivíduo j receber um escore k no item i é dada pela expressão da Equação 2, que define o MRG de Samejima :

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

de tal forma que:

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

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. Neste trabalho, na medida de satisfação, com quatro categorias de respostas, são estimados três parâmetros de dificuldade e um de discriminação.

A Figura 1 apresenta a curva característica do item 22 (“medidas adotadas para aprimorar a metodologia das aulas nas disciplinas do curso”) que compõe o instrumento de medida avaliado neste trabalho. Cada item contém quatro categorias de respostas: insatisfeito, pouco insatisfeito, satisfeito e muito satisfeito.

Na Figura 1, o traço latente (θ - nível de satisfação) está representado no eixo x e a probabilidade de selecionar uma determinada categoria de resposta em um dado nível do traço latente está representada no eixo y . A curva característica do MRG tem a seguinte interpretação: respondentes que possuem nível de satisfação inferior a $-1,48$ têm maior probabilidade de responder a categoria 1 (insatisfeito), respondentes que estão situados entre $-1,48$ e $0,80$ têm maior probabilidade de responder a categoria 2 (pouco insatisfeito), respondentes que possuem nível de satisfação entre $0,80$ e $2,89$ têm maior probabilidade de responder a categoria 3 (satisfeito) e respondentes que apresentam o nível de satisfação maior que $2,89$ têm maior probabilidade de responder a categoria 4 (muito satisfeito).

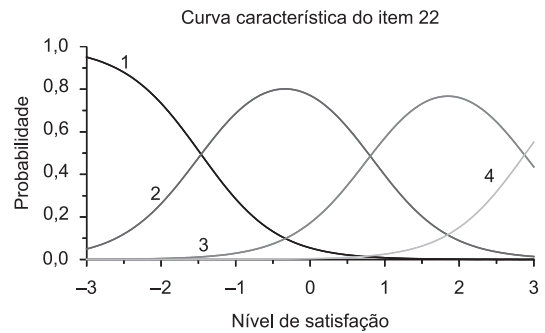


Figura 1. CCI do MRG do item 22 ($a = 1,94$, $b_1 = -1,48$, $b_2 = 0,80$, e $b_3 = 2,89$).

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 curva de informação do item (CCI). A função de informação do item é dada pela expressão da Equação 3:

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

em que $P'_{ik}(\theta)$ é a primeira derivada da curva de resposta da categoria avaliada em um particular nível do traço latente. A curva de informação total do teste é a soma das informações fornecidas por cada item que o compõe. A curva de informação do 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 do MRG pode ser feita por Máxima Verossimilhança ou métodos bayesianos (BAKER, 1992). O processo de estimação dos parâmetros dos itens e dos respondentes está implementado em *softwares* tais como MULTILOG (THISSEN, 1991; TOIT, 2003) e PARSCALE (MURAKI; BOCK, 1997; TOIT, 2003). Neste trabalho, os parâmetros do modelo foram estimados no *software* MULTILOG (THISSEN, 2003), por meio do método da Máxima Verossimilhança Marginal.

O Modelo de Resposta Gradual avalia somente um traço latente, ou seja, é unidimensional. A verificação da unidimensionalidade é feita por meio da Análise Fatorial. Neste trabalho, considera-se o critério de Reckase (1979), o qual sugere que os resultados podem indicar um fator dominante quando o primeiro fator corresponde pelo menos a 20% da variância total.

Para interpretar os valores dos parâmetros dos itens (a, b), é necessário conhecer a escala na qual foram determinados (ANDRADE; TAVARES; VALLE, 2000). A construção da escala é feita definindo-se níveis âncora e identificando-se itens âncora. Na identificação de itens âncora, utilizam-se três condições. Considerando-se dois níveis no traço latente, X e Y, com $X < Y$, um item é âncora para um nível Y se, e somente se, satisfizer as condições das Equações 4,5,6 (BEATON; ALLEN, 1992):

$$a) \quad P(U = 1\theta = Y) \geq 0,65 \quad (4)$$

$$b) \quad P(U = 1\theta = X) \geq 0,50 \quad (5)$$

$$c) \quad P(U = 1\theta = Y) - P(U = 1\theta = X) \geq 0,30 \quad (6)$$

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 das Equações 4,5,6). Entretanto, neste trabalho, considerou-se também que um item que atende pelo menos a duas das três condições é como “quase âncora” e deve ser incluído no seu respectivo nível âncora. Essas condições são válidas para os Modelos Logísticos (dicotômicos) da TRI, porém podem ser adaptadas ao MRG (MAFRA, 2010), no qual não é verificado se um item é âncora, mas se uma categoria do item é âncora .

4 Método e resultados

4.1 Coleta de dados

O instrumento utilizado faz parte de um questionário de avaliação, referente ao ensino, desenvolvido pela UNIOESTE, Cascavel, PR, e encontra-se no Apêndice A. O instrumento continha 35 itens, e cada item apresentava seis categorias de resposta: 1) totalmente insatisfeito, 2) insatisfeito, 3) pouco insatisfeito, 4) pouco satisfeito, 5) satisfeito, 6) totalmente satisfeito. Foi aplicado a 525 discentes dos cursos de Tecnologia daquela Instituição .

Estudos anteriores sobre o modelo de resposta gradual indicam que é necessária uma amostra de pelo menos 500 respondentes para obter uma calibração adequada dos itens (REISE; YU, 1990). Entretanto, itens com 6 categorias necessitam de uma amostra maior, já que muitas categorias ficam com poucas respostas, o que prejudica a estimação

dos parâmetros. Por isso, os itens foram agrupados em 4 categorias: 1) insatisfeito, compreendendo as categorias 1 e 2; 2) pouco insatisfeito, compreendendo as categorias 3 e 4; 3) satisfeito, compreendendo a categoria 5; e 4) muito satisfeito, compreendendo a categoria 6.

4.2 Análise de dimensionalidade

A suposição de unidimensionalidade foi verificada por meio da Análise Fatorial seguindo o critério de Reckase (1979). Segundo o autor, os resultados podem indicar um fator dominante quando o primeiro fator corresponde pelo menos a 20% da variância total. A Tabela 1 mostra o resultado da extração dos autovalores da análise fatorial pelo método dos componentes principais.

Os resultados da análise foram obtidos por meio do *software* SPSS, versão 17. Observa-se que o primeiro fator explica 33% da variância, atendendo ao critério de Reckase (1979). Portanto, pode-se concluir que há um fator dominante nos dados, satisfazendo a condição de unidimensionalidade para o uso do Modelo de Resposta Gradual .

4.3 Análises dos itens por meio do Modelo de Resposta Gradual

O processo de estimação dos parâmetros dos itens é denominado de calibração. Na Tabela 2, são visualizadas as estimativas dos parâmetros dos itens (a_i e $b_{i,k}$) para cada categoria ($b_{i,1}$, $b_{i,2}$ e $b_{i,3}$), obtidas por meio do *software* MULTILOG .

O parâmetro $b_{i,1}$ indica o ponto de inflexão da curva da primeira categoria (insatisfeito). O $b_{i,3}$ (muito satisfeito), 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 (pouco insatisfeito e satisfeito) (EMBRETSON; REISE, 2000). Neste trabalho, só existe um b intermediário, no caso o $b_{i,2}$, cujo valor é a média entre os pontos de picos entre as categorias 2 e 3. Já a média dos valores dos “ $b_{i,k}$ ” indica qual a posição do item na escala. Os três itens com maior valor de $b_{i,k}$ médio foram, em ordem crescente, os itens 20, 21 e 13. Já os itens com menor valor de $b_{i,k}$ médio, também em ordem crescente, foram os itens 34, 35, 10 e 31.

Tabela 1. Extração dos autovalores por meio dos componentes principais.

Fator	Autovalor	Variância total	Autovalor cumulativo (AC)	% AC
1	11,295	33,222	11,295	33,217
2	2,060	6,058	13,355	39,279

Tabela 2. Estimativas dos parâmetros dos itens para cada categoria de resposta.

Item	a	$b_{1,1}$	$b_{1,2}$	$b_{1,3}$	$b_{\text{médio}}$
1	1,30	-2,95	-0,40	2,52	-0,28
2	1,10	-2,74	0,15	3,47	0,29
3	1,45	-2,42	0,29	2,85	0,24
4	1,26	-3,16	-0,03	2,87	-0,11
5	1,04	-3,38	0,23	3,36	0,07
6	1,01	-2,73	-0,05	3,14	0,12
7	0,92	-2,65	0,44	3,41	0,40
8	1,37	-2,23	0,47	2,81	0,35
9	0,85	-3,63	0,02	4,02	0,14
10	1,31	-2,80	-0,46	1,91	-0,45
11	1,37	-2,54	-0,22	2,33	-0,14
12	0,95	-2,82	0,15	3,09	0,14
13	1,10	-1,75	1,09	4,08	1,14
14	1,24	-1,91	0,94	3,56	0,86
15	1,43	-1,70	0,83	3,56	0,90
16	1,16	-2,20	0,50	2,99	0,43
17	1,36	-1,64	0,75	2,73	0,61
18	1,18	-1,91	0,20	2,29	0,19
19	0,79	-2,73	0,71	4,03	0,67
20	1,53	-1,20	1,04	3,05	0,96
21	1,65	-1,22	1,08	3,02	0,96
22	1,94	-1,48	0,80	2,89	0,74
23	1,71	-1,69	0,74	3,47	0,84
24	2,18	-1,94	0,44	2,53	0,35
25	1,27	-2,29	0,50	2,94	0,38
26	0,93	-1,94	0,88	3,56	0,84
27	1,32	-1,99	0,40	2,99	0,47
28	1,66	-1,57	0,44	2,26	0,38
29	1,77	-1,65	0,32	2,33	0,33
30	1,23	-1,70	0,60	2,79	0,56
31	1,25	-2,31	-0,42	1,40	-0,44
34	1,01	-3,49	-0,68	2,10	-0,69
35	1,49	-1,75	0,60	2,59	0,48

Observa-se na Tabela 2 que os parâmetros de discriminação variaram de 0,79 a 2,17. Isto indica que todos os itens possuem um poder de discriminação satisfatório. Os três itens com maior grau de discriminação foram, em ordem crescente, 29, 22 e 24. Já os itens com menor poder de dificuldade, também em ordem crescente, foram os itens 19, 9 e 7. Os itens 32 e 33 apresentaram um baixo parâmetro de discriminação ($<0,7$); logo, são itens que não discriminam indivíduos satisfeitos dos que não estão satisfeitos. Estes itens foram eliminados da análise para a obtenção dessa calibração final.

A Figura 2 apresenta a curva característica e a função de informação do item 1. Este item possui a seguinte descrição: “Qualidade do curso de graduação

que realiza”. O parâmetro de discriminação é igual a 1,30, enquanto que o parâmetro de localização deste item, na escala (0, 1), é igual a -0,28. Indivíduos com satisfação nesta escala menor que -2,95 têm maior probabilidade de responder a categoria 1 (insatisfeito), indivíduos com grau de satisfação entre -2,95 e -0,40 têm maior probabilidade de responder a categoria 2 (pouco insatisfeito), indivíduos com grau de satisfação entre -0,40 e 2,52 têm maior probabilidade de responder a categoria 3 (satisfeito) e indivíduos com grau de satisfação maior que 2,52 têm maior probabilidade de responder a categoria 4 (muito satisfeito). É possível verificar os valores dos $b_{i,k}$ nesse gráfico: o $b_{1,1}$ (-2,94) é o ponto de inflexão da categoria 1 (insatisfeito), o $b_{1,2}$ é o ponto médio na

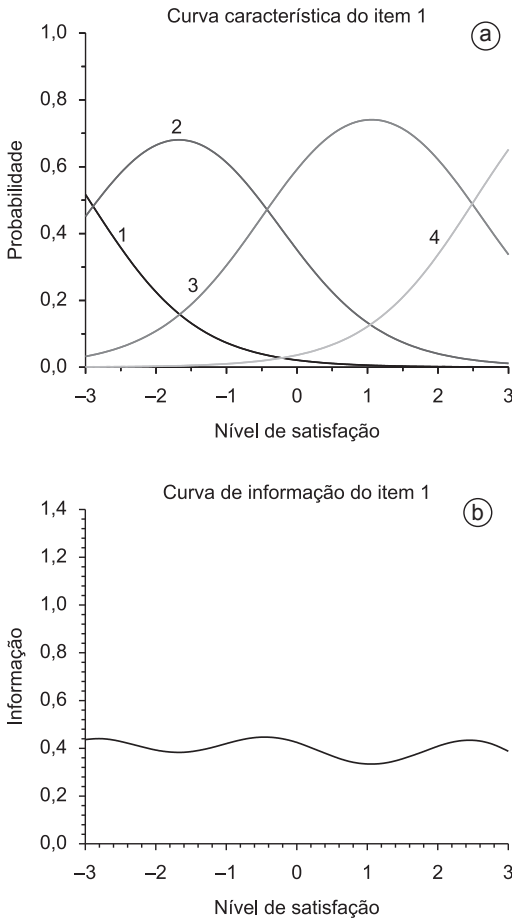


Figura 2. Função característica (a) e Função de informação (b) do item 1.

escala entre o valor máximo da categoria 2 (-1,6) e o máximo da categoria 3 (1,0), e o $b_{1,3}$ (2,52) é o ponto de inflexão da categoria 4 (muito satisfeito). Observa-se que esse item fornece informação praticamente constante no intervalo entre -3 e 3.

A Figura 3 apresenta a curva de informação do instrumento. Observa-se que o instrumento de medida tem maior informação no intervalo de -2,5 a 3. Isso significa que ele é mais adequado para medir a satisfação de indivíduos que possuem nível de satisfação entre -2,5 e 3, ou seja, ele é adequado para praticamente todos os indivíduos. A linha pontilhada representa o erro padrão. Nota-se que, nas regiões do nível de satisfação em que há quantidades maiores de informações, o erro padrão é menor.

4.4 Teste de ajuste do modelo

A Tabela 3 apresenta a distribuição de frequências dos 35 itens em relação à quantidade $\chi^2.gL^{-1}$ (a razão entre o coeficiente qui-quadrado de ajuste da distribuição e os graus de liberdade), obtida por meio

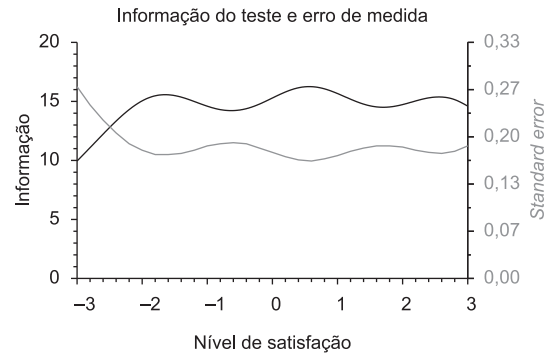


Figura 3. Função de informação do instrumento.

do programa *MODFIT Web Demo* (DRASGOW et al., 1995). Segundo Drasgow et al. (1995), para se ter um bom ajuste do modelo, deve-se obter $\chi^2.gL^{-1} < 3$. A primeira linha (*Singlets*) refere-se ao ajuste de cada um dos itens individualmente. A segunda linha (*Doublets*) refere-se ao ajuste de pares de itens, definidos dentro de grupos de itens de tamanho igual a três, os quais contêm itens de diferentes graus de dificuldade, definidos pelo *MODFIT Web Demo*. A terceira linha (*Triplets*) refere-se ao ajuste de trios de itens, definidos dentro dos mesmos grupos utilizados nos *Doublets*. Os resultados do teste de ajuste dos modelos, de acordo com Drasgow et al. (1995), indicam que o ajuste de modelo da Resposta Gradual foi bom. Os valores médios foram baixos para $\chi^2.gL^{-1} (<3)$ em todos os tipos de ajustes testados (*Singlets*, *Doublets* e *Triplets*).

4.5 Construção da escala de medida de satisfação

Os parâmetros dos itens e os parâmetros dos indivíduos são estimados na mesma métrica, ou seja, média 0 (zero) e desvio padrão 1 (um). É possível, para efeitos práticos, transformar os valores dos parâmetros tanto dos itens como indivíduos em outros valores, mas que mantenham as mesmas relações de ordem entre seus pontos. Neste sentido, a escala para a satisfação com relação ao curso foi construída com média 100 e desvio padrão 10 e utilizaram-se as transformações apresentadas nas Equações 7, 8, 9, 10

$$\theta^* = 10 \times \theta + 100 \tag{7}$$

$$b^* = 10 \times b + 100 \tag{8}$$

$$a^* = \frac{a}{10} \tag{9}$$

$$P(U_i = 1\theta) = P(U_i = 1\theta^*) \tag{10}$$

A escala da satisfação foi construída com base nos níveis âncora e quase âncora definidos na seção 2. Os

cálculos das probabilidades para $b_{i,1}$, $b_{i,2}$, $b_{i,3}$ estão apresentados nos Apêndices B, C e D, respectivamente. As probabilidades para $b_{i,1}$ representam a probabilidade de um respondente, com aquele nível de satisfação, estar situado na categoria pouco insatisfeito ou em uma categoria mais elevada. As probabilidades para $b_{i,2}$ representam a probabilidade de determinado respondente, com aquele nível de satisfação, estar localizado na categoria satisfeito ou em uma categoria mais elevada. Já as probabilidades calculadas para $b_{i,3}$ representam a probabilidade de determinado respondente, com aquele nível de satisfação, estar na categoria muito satisfeito, que é a categoria mais elevada. A probabilidade de obter a categoria mais baixa ou superior (insatisfeito ou mais elevada) é igual a 1,0.

Com base nas probabilidades acumuladas, obtiveram-se os níveis âncoras e quase âncoras. Foram identificados seis níveis âncoras: níveis: 80, 90, 110, 110, 120 e 130. Nenhum item foi identificado para o nível 70. As Tabelas 4, 5 apresentam os níveis âncoras e quase âncoras, respectivamente, com seus relativos itens âncoras para as categorias pouco satisfeito, satisfeito e muito satisfeito. Fundamentando-se no conteúdo dos itens, os níveis da escala de satisfação podem ser assim definidos:

Nível 80 – Caracteriza-se por identificar respondentes que estão pouco insatisfeitos com a

organização na exposição de conteúdos, com as formas de avaliação utilizadas para avaliar os níveis de aprendizagem, com a seriedade demonstrada pelos docentes e dos alunos, com a qualidade do curso e com o nível de formação dos alunos que concluem o curso.

Nível 90 – Além das características anteriores, este nível identifica os respondentes que estão pouco insatisfeitos com a capacidade e qualificação manifestada pelos alunos para leitura e escrita de textos científicos, com a oportunidade de iniciação dos alunos na pesquisa, com as medidas adotadas para aprimorar a metodologia das aulas e avaliação dos alunos nas disciplinas, com os procedimentos adotados pelo curso para a qualificação dos conteúdos desenvolvidos nas disciplinas, com a dinâmica das aulas para manter a atenção dos alunos, com os mecanismos de atendimento e orientação acadêmica dos alunos no dia a dia dos cursos, com o comprometimento efetivo dos docentes com a qualificação do curso, do conhecimento da situação dos alunos que concluíram o curso, com o nível de formação dos alunos que ingressam no curso e com a oportunidade de treinamento e inserção no mercado de trabalho oferecido pelo curso.

Nível 100 – Caracteriza-se por identificar respondentes que estão satisfeitos com relação ao nível de formação atingido pelos alunos que concluem o curso e com a qualidade do curso e com a dinâmica

Tabela 3. Frequência dos itens em relação ao teste de ajuste.

$\chi^2 \cdot gL^{-1}$	<1	1 < 2	2 < 3	3 < 4	4 < 5	5 < 7	>7	Média	Desvio padrão
<i>Singlets</i>	35	0	0	0	0	0	0	0,017	0,015
<i>Doublets</i>	25	11	2	0	1	0	0	1,012	0,750
<i>Triplets</i>	10	6	1	0	0	0	0	1,058	0,572

Tabela 4. Distribuição dos itens âncora de acordo com as categorias.

Categorias	Nível de satisfação						
	70	80	90	100	110	120	130
b1 – categoria pouco insatisfeito	-	1,3,10,11	15,17,23,24,28,29,30,35	-	-	-	-
b2 – categoria satisfeito	-	-	-	11,35	3,4,8,17,25,28,29,30	21,22	-
b3 – categoria muito satisfeito	-	-	-	-	-	-	1,29,35

Tabela 5. Distribuição dos itens quase âncora de acordo com as categorias.

Categorias	Nível de satisfação						
	70	80	90	100	110	120	130
b1 – categoria pouco insatisfeito	-	6,12	13,14,18,22	-	-	-	-
b2 – categoria satisfeito	-	-	-	1	2	14,15	-
b3 – categoria muito satisfeito	-	-	-	-	-	-	25

das aulas para manter a atenção dos alunos, além das características dos níveis anteriores .

Nível 110 – Caracteriza-se por identificar respondentes que estão satisfeitos quanto à organização de conteúdos pelos docentes, com a qualidade dos planos de ensino, bem como com a estrutura curricular do curso, com a criatividade demonstrada, oportunidade de iniciação dos alunos na pesquisa, com as condições para a dedicação ao curso, com o atendimento da orientação acadêmica, com o comprometimento dos docentes e conhecimento deles quanto à qualificação do curso, além das características dos níveis anteriores .

Nível 120 – Caracteriza-se por identificar respondentes que estão satisfeitos com as medidas para superar as dificuldades dos alunos com deficiências nas disciplinas, com as medidas adotadas para aprimorar a metodologia das aulas, bem como a avaliação e com a capacidade dos alunos para a leitura e escrita de trabalhos científicos, além das características dos níveis anteriores .

Nível 130 – Caracteriza-se por identificar respondentes que estão muito satisfeitos quando a qualidade do curso que realiza, bem como com as condições dos alunos quanto à dedicação ao curso e com o comprometimento dos docentes com a qualificação do curso e dinâmica das aulas para manter a atenção dos alunos, além das características dos níveis anteriores .

4.6 Localização dos respondentes na escala de medida de satisfação

A Figura 4 apresenta a distribuição dos respondentes nos níveis da escala de satisfação.

Nota-se que a maioria dos respondentes (aproximadamente 94%) está concentrada nos níveis 100 e 110, ou seja, 52,38% dos respondentes encontram-se no nível 100 e 41,90% situam-se no nível 110, indicando que eles estão satisfeitos com os

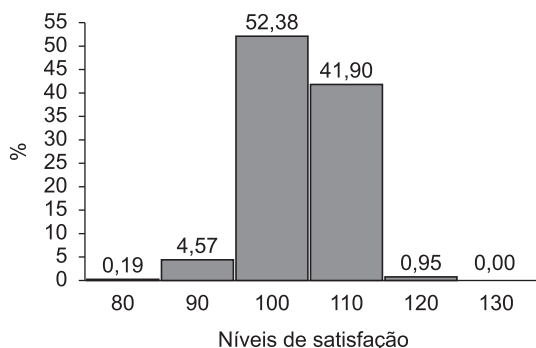


Figura 4. Distribuição dos respondentes nos níveis da escala de satisfação.

itens relativos a estes níveis. Apenas 0,95% situam-se no nível 120. Nota-se uma porcentagem muito pequena de respondentes localizados nos níveis 80 e 90, isto é, 0,19% e 4,5%, respectivamente, sugerindo que eles estão pouco satisfeitos (ou insatisfeitos) com os itens relativos a estes níveis. Não há nenhum respondente no nível 130.

5 Discussão e conclusão

A utilização da Teoria da Resposta ao Item (TRI) em avaliações é uma tendência universal. A TRI tem sido utilizada na área educacional, proporcionando uma avaliação mais precisa do fenômeno estudado e também está encontrando espaço em outras áreas, avaliando outros tipos de traços latentes, principalmente os relacionados com fenômenos organizacionais.

Embora a TRI possua muitas vantagens (EMBRETSON; REISE, 2000), ela apresenta também algumas limitações. Uma delas está relacionada quanto ao tamanho da amostra. Para se obter estimativas dos parâmetros dos itens com baixos erros padrões, deve-se ter uma amostra significativa de respondentes. O número de respondentes depende do número de parâmetros a serem estimados e do número de categorias de respostas no item. No presente estudo, foram estimados 140 parâmetros (Tabela 1), com uma amostra de 525 respondentes. A análise de dimensionalidade mostrou que o instrumento de medida avalia um traço latente dominante que, no caso, é o nível de satisfação, satisfazendo o critério de unidimensionalidade do MRG.

Os resultados deste estudo mostram que o modelo de Escala Gradual foi adequado para a análise e construção de uma escala para avaliar satisfação quando os itens possuem mais de duas categorias de respostas ordenadas.

As estatísticas de ajuste obtidas pelo programa *MODFIT Web Demo* mostraram que os dados se ajustaram bem ao modelo MRG para os três tipos de ajustes testados (*Singlets*, *Doublets* e *Triplets*).

O instrumento utilizado neste trabalho para avaliar o curso continha 35 itens. Na calibração dos itens, dois foram eliminados por apresentarem um valor baixo de discriminação, restando 33 itens para a construção da escala. A Teoria de Resposta ao Item permitiu a criação de uma escala de medida de satisfação baseada em níveis âncoras e quase âncoras. Nesta escala, é possível caracterizar o nível no qual cada respondente se encontra, quanto a sua satisfação em relação ao curso. Neste trabalho, foram criados seis níveis padronizados. Cada nível continha características relativas aos itens nele especificados.

Os resultados mostraram que a maioria dos respondentes situou-se em dois níveis na escala, indicando que os alunos estão satisfeitos com os itens

investigados, bem como com as práticas desenvolvidas no curso.

O uso do Modelo de Resposta Gradual (MRG) e um instrumento de medida com respostas graduadas permite obter mais informação das respostas dos indivíduos, uma vez que leva em consideração a probabilidade de ele responder cada categoria de resposta.

Ressalta-se, ainda, que a TRI possibilita, por meio de um processo denominado equalização, que novos itens sejam adicionados a este instrumento de medida e sejam calibrados na mesma escala, permitindo a criação de um banco de itens para avaliar satisfação. Uma vez que os itens já estão calibrados, também é possível conhecer o nível de satisfação de outra amostra de respondentes.

Com esta aplicação, verificou-se que a TRI fornece estimativas consistentes para medidas de traços latentes. Na medida de satisfação, a aplicação da TRI permite uma avaliação do fenômeno em estudo com mais precisão e consistência e, conseqüentemente, fornece informações relevantes na tomada de decisões e no planejamento estratégico, podendo trazer uma grande contribuição na área de gestão.

Os resultados da TRI suprimem a intuição e fortalecem a racionalidade para a tomada de decisão. A importância da TRI reside no método aplicado e nos resultados obtidos, para posterior aplicação de medidas corretivas ou alternativas.

Referências

- ALEXANDRE, J. W. C. et al. Uma proposta de análise de um construto para a medição dos fatores críticos da gestão pela qualidade através da teoria da resposta ao item. **Gestão & Produção**, v. 9, n. 2, p. 129-141, 2002.
- ANDRADE, D. F.; BORTOLOTTI, S. L. V. Aplicação de um modelo de desdobramento graduado generalizado - GGUM da teoria da resposta ao item. **Estudos em Avaliação Educacional**, v. 18, n. 37, 2007.
- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria de Resposta ao Item: conceitos e aplicações**. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA - SINAPE, 4., 2000, Caxambu. **Anais...** Caxambu: Associação Brasileira de Estatística, 2000.
- ARAUJO, E. A. C.; ANDRADE, D. F.; BORTOLOTTI, S. L. V. Teoria da Resposta ao item. **Revista da Escola de Enfermagem USP**, v. 43, n. especial, p. 1000-1008, 2009.
- BAKER, F. B. **Item response theory parameter estimation techniques**. New York: Marcel Dekker, 1992.
- BEATON, A. E.; ALLEN, N. L. Interpreting scales through scale anchoring. **Journal of Educational Statistics**, v. 17, p. 191-204, 1992. <http://dx.doi.org/10.2307/1165169>
- DRASGOW, F. et al. Fitting Polytomous Item Response Theory Models to Multiple-Choice Tests. **Applied Psychological Measurement**, v. 19, n. 2, p. 143-165, 1995. <http://dx.doi.org/10.1177/014662169501900203>
- EMBRETSON, S.; REISE, S. P. **Item Response Theory for Psychologists**. New Jersey: Lawrence Erlbaum Associates, Inc. Publishers, 2000.
- EVRRARD, Y. **A satisfação dos consumidores: situação das pesquisas**. Tradução de Ana Maria Machado Toaldo. Porto Alegre: PPGA/UFRGS, 1995. Tradução para fins acadêmicos.
- FORNELL, C. et al. The American Customer Satisfaction Index: Nature, Purpose, and Findings. **Journal of Marketing**, v. 60, p. 7-18, 1996. <http://dx.doi.org/10.2307/1251898>
- HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. **Fundamentals of item response theory**. Newbury Park: Sage, 1991.
- HASTREITER, S. T. **Satisfação do consumidor em Shopping Centers de Curitiba**. 1998. Dissertação (Mestrado em Ciências Sociais)-Universidade Federal do Paraná, Curitiba, 1998.
- KOTLER, P. **Administração de Marketing: Análise, Planejamento, Implementação e Controle**. 5. ed. São Paulo: Atlas, 1998.
- LORD, F. M. **Applications of item response theory to practical testing problems**. Hillsdale: Lawrence Erlbaum, 1980.
- 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.
- MATTAR, F. N. **Pesquisa de marketing**. 6. ed. São Paulo: Atlas, 2005. vol. 1.
- MATTEUCCI, M.; STRACQUALURSI, L. Student assessment via Graded Response Model. **Statistica**, v. 66, n. 4, p. 435-447, 2006.
- MEDEIROS, M. F.; FARIA, E. T. (Org.). **Educação a distância: cartografias pulsantes em movimento**. Porto Alegre: EDIPUCRS, 2003.
- MOWEN, J. C. **Consumer Behavior**. 5th ed. Englewood Cliffs: Prentice-Hall, 1997.
- 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.
- OLIVER, R. L. **Satisfaction: a behavioral perspective on the consumer**. 2nd ed. New York: M. E. Sharpe, 2009.
- PERILLO, R. D. **Índice de Satisfação Ponderado pela Importância: Um modelo para avaliar a satisfação de consumidores**. 2000. 162 f. Dissertação (Mestrado em Administração)-Universidade de Brasília, Brasília, 2000.
- RECKASE, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. **Journal of Educational Statistics**, v. 4, p. 207-230, 1979. <http://dx.doi.org/10.2307/1164671>
- REISE, S. P.; YU, J. Parameter recovery in the graded response model using MULTILOG. **Journal of Educational Measurement**, n. 27, p. 133-144, 1990. <http://dx.doi.org/10.1111/j.1745-3984.1990.tb00738.x>
- ROSSI, C. A. V.; SLOGO, L. A. Pesquisa de satisfação de consumidores: o estado-da-arte e proposição de

- um método brasileiro. **Revista de Administração Contemporânea**, v. 1, n. 1, 1998.
- SAMEJIMA, F. Estimation of latent ability using a response pattern of graded scores. **Psychometrika Monograph Supplement**, n. 17, 1969.
- SAMEJIMA, F. A general model for free response data. **Psychometrika Monograph Supplement**, n. 18, 1972.
- SAMEJIMA, F. Graded response model. In: VAN DER LINDEN, W. J.; HAMBLETON, R. K. (Eds.). **Handbook of modern item response theory**. New York: Springer, 1997. p. 85-100.
- SPRENG, R. A.; MacKenzie, S. B.; OLSHAVSKY, R. W. A Reexamination of the Determinants of Consumer Satisfaction. **Journal of Marketing**, v. 6, n. 3, p. 15-32, 1996. <http://dx.doi.org/10.2307/1251839>
- THISSEN, D. **Multilog User's Guide: Multiple, Categorical Item Analysis and Testing Score using Item Response Theory**. Chicago: Scientific Software, Inc, 1991.
- THISSEN D. **Multilog 7. 0. Multiple: categorical item analysis and test scoring using item response theory**. Lincolnwood: Scientific Software International, 2003.
- TOIT, M. **IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT**. Chicago: Scientific Software, Inc.; 2003.
- VARGAS, V. C. C. et al. Avaliação dos intangíveis: uma aplicação em capital humano. **Gestão & Produção**, v. 15, n. 3, p. 619-634, 2008.

Apêndice A. Itens avaliados na medida de satisfação.

Item	Descrição do item
1	Qualidade do curso de graduação que realiza
2	Estrutura curricular (de disciplinas) do curso
3	Organização na exposição de conteúdos pelos docentes
4	Qualidade dos Planos de Ensino apresentados pelos professores
5	Metodologia para o desenvolvimento do ensino-aprendizagem nas aulas
6	Formas de avaliação utilizadas nas disciplinas para “medir” os níveis de aprendizagem dos alunos
7	Alternativas oferecidas aos alunos para a complementação de sua formação global
8	Criatividade demonstrada pelos docentes no desempenho das atividades de ensino
9	Notas obtidas nas disciplinas em relação à aprendizagem alcançada
10	Seriedade acadêmica manifestada pelos docentes do curso
11	Nível de formação atingido pelos alunos que concluem o curso
12	Seriedade acadêmica dos alunos do curso
13	Nível de formação dos alunos quando ingressam no curso
14	Capacidade manifestada pelos alunos para a leitura de textos científicos durante o curso de graduação
15	Qualificação dos alunos para a escrita de trabalhos científicos durante o curso de graduação
16	Qualificação manifestada pelos alunos para a elaboração de monografia e/ou trabalho de conclusão de curso
17	Oportunidade de iniciação dos alunos na pesquisa no curso que realiza
18	Oportunidade de treinamento e inserção no mercado de trabalho oferecido pelo curso
19	Número de alunos que concluem o curso a cada ano
20	Medidas adotadas para a melhoria da qualidade do ensino no curso
21	Medidas adotadas para superar as dificuldades dos alunos com deficiências nas disciplinas
22	Medidas adotadas para aprimorar a metodologia das aulas nas disciplinas do curso
23	Medidas adotadas para aprimorar a avaliação dos alunos nas disciplinas do curso
24	Procedimentos adotados pelo curso para a qualificação dos conteúdos desenvolvidos nas disciplinas
25	Condições dos alunos para a dedicação ao curso de graduação
26	Tempo dedicado ao estudo das disciplinas que cursa
27	Iniciativa dos alunos para a complementação de sua formação acadêmica
28	Mecanismos de atendimento e orientação acadêmica dos alunos no cotidiano dos cursos
29	Comprometimento efetivo dos docentes com a qualificação do curso de graduação que realiza
30	Conhecimento da situação dos alunos que já concluíram o curso no mercado de trabalho
31	Satisfação em relação ao curso que está realizando
32	Limpeza e estado de conservação da sala de aula
33	Pontualidade e assiduidade dos docentes nas aulas
34	Conhecimento demonstrado pelos docentes nas matérias que lecionam
35	Dinâmica das aulas para manter a atenção dos alunos

Apêndice B. Probabilidades acumuladas para a categoria 1 (pouco insatisfeito) ou mais elevada e definição de itens âncoras e quase âncoras.

Item	a	$b_{i,1}$	70	80	90	100	110	120	130
1	0,130	70,5	0,484	0,775	0,927	0,979	0,994	0,998	1,000
2	0,110	72,6	<i>0,429</i>	<i>0,693</i>	0,871	0,953	0,984	0,995	0,998
3	0,145	75,8	0,301	0,648	0,887	0,971	0,993	0,998	1,000
4	0,126	68,4	0,550	0,812	0,938	0,982	0,995	0,999	1,000
5	0,104	66,2	0,598	0,808	0,922	0,971	0,990	0,996	0,999
6	0,101	72,7	<i>0,432</i>	<i>0,676</i>	0,852	0,940	0,977	0,992	0,997
7	0,092	73,5	0,420	0,645	0,820	0,920	0,966	0,986	0,995
8	0,137	77,7	0,258	0,578	0,844	0,955	0,988	0,997	0,999
9	0,085	63,7	0,631	0,800	0,903	0,956	0,981	0,992	0,996
10	0,131	72,0	0,435	0,740	0,914	0,975	0,993	0,998	0,999
11	0,137	74,6	0,347	0,677	0,892	0,970	0,992	0,998	0,999
12	0,095	71,8	<i>0,457</i>	<i>0,685</i>	0,849	0,936	0,974	0,990	0,996
13	0,110	82,5	0,202	<i>0,432</i>	<i>0,695</i>	0,873	0,954	0,984	0,995
14	0,124	80,9	0,206	<i>0,472</i>	<i>0,756</i>	0,914	0,974	0,992	0,998
15	0,143	83,0	0,135	0,394	0,731	0,919	0,979	0,995	0,999
16	0,116	78,0	0,283	0,558	0,801	0,928	0,976	0,992	0,998
17	0,136	83,6	0,136	0,380	0,705	0,903	0,973	0,993	0,998
18	0,118	80,9	0,216	<i>0,473</i>	<i>0,745</i>	0,905	0,969	0,990	0,997
19	0,079	72,7	0,447	0,640	0,797	0,896	0,950	0,977	0,989
20	0,153	88,0	0,060	0,227	0,576	0,862	0,967	0,993	0,998
21	0,165	87,8	0,050	0,216	0,590	0,882	0,975	0,995	0,999
22	0,194	85,2	0,050	<i>0,267</i>	<i>0,717</i>	0,946	0,992	0,999	1,000
23	0,171	83,1	0,096	0,370	0,765	0,947	0,990	0,998	1,000
24	0,218	80,6	0,090	0,467	0,886	0,986	0,998	1,000	1,000
25	0,127	77,1	0,289	0,591	0,837	0,948	0,985	0,996	0,999
26	0,093	80,6	0,272	0,486	0,706	0,859	0,939	0,975	0,990
27	0,132	80,1	0,209	0,497	0,787	0,933	0,981	0,995	0,999
28	0,166	84,3	0,085	0,329	0,720	0,931	0,986	0,997	0,999
29	0,177	83,5	0,084	0,350	0,760	0,949	0,991	0,998	1,000
30	0,123	83,0	0,168	0,409	0,703	0,890	0,965	0,990	0,997
31	0,125	76,9	0,297	0,596	0,837	0,947	0,984	0,995	0,999
34	0,101	65,1	0,621	0,818	0,925	0,971	0,989	0,996	0,999
35	0,149	82,5	0,134	0,408	0,754	0,931	0,984	0,996	0,999

Nota: as células em itálico indicam o nível em que o item é quase âncora; as células em negrito indicam o nível em que o item é âncora.

Apêndice C. Probabilidades acumuladas para a categoria 2 (satisfeito) ou mais elevada e definição de itens âncoras e quase âncoras.

Item	a	$b_{i,2}$	70	80	90	100	110	120	130
1	0,130	96,0	0,033	0,111	<i>0,314</i>	<i>0,627</i>	0,861	0,958	0,988
2	0,110	101,5	0,030	0,086	0,220	<i>0,459</i>	<i>0,718</i>	0,884	0,958
3	0,145	100,0	0,013	0,052	0,190	0,500	0,810	0,948	0,987
4	0,126	102,9	0,016	0,053	0,164	0,410	0,710	0,896	0,968
5	0,104	99,7	0,044	0,114	0,267	0,508	0,745	0,892	0,959
6	0,101	102,3	0,037	0,095	0,224	0,442	0,685	0,857	0,943
7	0,092	99,5	0,062	0,143	0,294	0,511	0,724	0,868	0,943
8	0,137	104,4	0,009	0,034	0,122	0,354	0,683	0,894	0,971
9	0,085	104,7	0,050	0,109	0,223	0,401	0,611	0,786	0,896
10	0,131	100,2	0,019	0,066	0,208	0,493	0,783	0,930	0,980
11	0,137	95,4	0,030	0,108	0,323	0,653	0,881	0,967	0,991
12	0,095	97,8	0,067	0,156	0,323	0,552	0,761	0,892	0,955
13	0,110	101,5	0,030	0,086	0,220	0,459	0,718	0,884	0,958
14	0,124	110,9	0,006	0,021	0,070	0,206	<i>0,472</i>	<i>0,756</i>	0,914
15	0,143	109,4	0,004	0,015	0,059	0,207	<i>0,521</i>	<i>0,820</i>	0,950
16	0,116	108,3	0,012	0,036	0,107	0,276	0,549	0,795	0,925
17	0,136	105,0	0,008	0,032	0,115	0,336	0,664	0,885	0,968
18	0,118	107,5	0,012	0,038	0,113	0,292	0,573	0,814	0,934
19	0,079	102,0	0,074	0,150	0,279	0,461	0,653	0,806	0,901
20	0,153	107,1	0,003	0,016	0,068	0,252	0,609	0,878	0,971
21	0,165	110,4	0,001	0,007	0,033	0,152	0,484	0,830	0,962
22	0,194	110,8	0,000	0,003	0,017	0,110	0,461	0,856	0,976
23	0,171	108,0	0,002	0,008	0,044	0,203	0,585	0,886	0,977
24	0,218	107,4	0,000	0,003	0,022	0,166	0,638	0,940	0,993
25	0,127	104,4	0,013	0,043	0,138	0,364	0,671	0,879	0,963
26	0,093	105,0	0,037	0,089	0,199	0,386	0,614	0,801	0,911
27	0,132	108,8	0,006	0,022	0,077	0,238	0,540	0,814	0,943
28	0,166	104,0	0,004	0,018	0,089	0,340	0,730	0,934	0,987
29	0,177	104,4	0,002	0,013	0,073	0,315	0,729	0,941	0,989
30	0,123	103,2	0,017	0,054	0,165	0,403	0,698	0,888	0,964
31	0,125	106,0	0,011	0,037	0,119	0,321	0,622	0,852	0,953
34	0,101	95,8	0,069	0,169	0,358	0,604	0,808	0,920	0,969
35	0,149	93,2	0,031	0,123	0,383	0,734	0,924	0,982	0,996

Nota: as células em itálico indicam o nível em que o item é quase âncora; as células em negrito indicam o nível em que o item é âncora.

Apêndice D. Probabilidades acumuladas para a categoria 3 (muito satisfeito) ou mais elevada e definição de itens âncoras e quase âncoras.

Item	a	$b_{i,3}$	70	80	90	100	110	120	130
1	0,130	125,2	0,001	0,003	0,010	0,036	0,122	0,337	0,651
2	0,110	134,7	0,001	0,002	0,007	0,022	0,062	0,166	0,374
3	0,145	100,0	0,013	0,052	0,190	0,500	0,810	0,948	0,987
4	0,126	128,5	0,001	0,002	0,008	0,027	0,089	0,255	0,547
5	0,104	128,7	0,002	0,006	0,018	0,048	0,125	0,288	0,534
6	0,101	133,6	0,002	0,004	0,012	0,032	0,084	0,202	0,410
7	0,092	131,4	0,004	0,009	0,022	0,053	0,123	0,259	0,468
8	0,137	134,1	0,000	0,001	0,002	0,009	0,036	0,127	0,363
9	0,085	128,1	0,007	0,016	0,038	0,084	0,177	0,334	0,540
10	0,131	140,2	0,000	0,000	0,001	0,005	0,019	0,066	0,208
11	0,137	119,1	0,001	0,005	0,018	0,068	0,223	0,531	0,817
12	0,095	123,3	0,006	0,016	0,041	0,099	0,220	0,422	0,654
13	0,110	130,9	0,001	0,004	0,011	0,032	0,091	0,232	0,475
14	0,124	140,8	0,000	0,001	0,002	0,006	0,021	0,070	0,208
15	0,143	135,6	0,000	0,000	0,001	0,006	0,025	0,097	0,310
16	0,116	135,6	0,000	0,002	0,005	0,016	0,049	0,141	0,343
17	0,136	129,9	0,000	0,001	0,004	0,017	0,063	0,206	0,503
18	0,118	127,3	0,001	0,004	0,012	0,038	0,115	0,297	0,579
19	0,079	122,9	0,015	0,033	0,069	0,141	0,265	0,443	0,637
20	0,153	140,3	0,000	0,000	0,000	0,002	0,010	0,043	0,171
21	0,165	130,5	0,000	0,000	0,001	0,006	0,033	0,150	0,479
22	0,194	130,2	0,000	0,000	0,000	0,003	0,019	0,121	0,490
23	0,171	128,9	0,000	0,000	0,001	0,007	0,038	0,179	0,547
24	0,218	134,7	0,000	0,000	0,000	0,001	0,005	0,039	0,264
25	0,127	125,3	0,001	0,003	0,011	0,039	0,125	<i>0,338</i>	<i>0,645</i>
26	0,093	129,4	0,004	0,010	0,025	0,061	0,141	0,294	0,514
27	0,132	135,6	0,000	0,001	0,002	0,009	0,033	0,113	0,323
28	0,166	129,9	0,000	0,000	0,001	0,007	0,035	0,162	0,504
29	0,177	122,6	0,000	0,001	0,003	0,018	0,097	0,387	0,787
30	0,123	123,3	0,001	0,005	0,016	0,054	0,163	0,400	0,695
31	0,125	127,9	0,001	0,003	0,009	0,030	0,096	0,271	0,565
34	0,101	114,0	0,012	0,031	0,081	0,196	0,400	0,647	0,834
35	0,149	121,0	0,001	0,002	0,010	0,042	0,163	0,463	0,793

Nota: as células em itálico indicam o nível em que o item é quase âncora; as células em negrito indicam o nível em que o item é âncora.