

Avaliação do nível de satisfação discente de uma instituição de Ensino Superior: uma análise dos métodos da Teoria Clássica da Medida e da Teoria da Resposta ao Item

Anna Cecilia Amaral Petrassi ^a

Antonio Cezar Bornia ^b

Dalton Francisco Andrade ^c

Resumo

O presente trabalho compara as informações geradas pelo processo de autoavaliação de uma Instituição Federal de Educação Superior (Ifes), no contexto do Sistema Nacional de Avaliação da Educação Superior (Sinaes), analisadas pela Comissão Própria de Avaliação da Ifes, ciclo de 2017, com as mesmas informações tratadas a partir do uso da Teoria da Resposta ao Item (TRI) no lugar da Teoria Clássica da Medida, com objetivo de demonstrar como o processo de análise pode se beneficiar com a utilização da TRI, sendo um instrumento mais robusto para a análise e o desenho de ações de melhorias na gestão institucional. Para isso, foi selecionado o extrato “categoria discente” da avaliação sobre a instituição, no segundo ciclo de 2017. Foram descritas e comparadas as duas teorias, a clássica e a TRI. Com os dados coletados da instituição, foram calculadas e explicadas as curvas de informação dos itens e do teste, por meio da análise propiciada pela TRI foi demonstrado o ganho que se pode obter com a sua utilização na autoavaliação institucional.

Palavras-chave: Autoavaliação Institucional. Teoria da Resposta ao Item. Gestão Universitária.

^a Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.

^b Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.

^c Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.

Recebido em: 11 fev. 2019

Aceito em: 07 out. 2020

1 Introdução

A avaliação do Ensino Superior possui duas dimensões: a avaliação do aprendizado e a avaliação institucional. Nesse campo, atua o Sistema Nacional de Avaliação da Educação Superior (Sinaes), instituído em abril de 2004, pela Lei nº 10.086 (BRASIL, 2004). Dentro do Sinaes, a avaliação institucional é composta por uma autoavaliação, realizada pelas instituições de Ensino Superior (IES) através de suas Comissões Próprias de Avaliação (CPA) e por avaliações externas conduzidas por comissões indicadas pelo Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (Inep).

A autoavaliação institucional não deve ser apenas instrumento de coleta de informações, mas, sim, uma ferramenta para a melhoria do desempenho institucional, promovendo o planejamento com base no diagnóstico sobre os pontos críticos. A autoavaliação deve ser tratada como instrumento facilitador para o planejamento e para a tomada de decisões, promovendo a autocritica e o aprendizado coletivo, de forma que aconteça um processo de contínua melhoria e o alcance da missão da instituição (GALDINO, 2011).

Conforme nota técnica do Inep/MEC, o relatório de autoavaliação institucional, elaborado pela CPA, deve conter cinco partes: introdução, metodologia, desenvolvimento, análise dos dados e das informações e ações previstas com base nessa análise (INEP, 2014). Como metodologia de autoavaliação, as instituições costumam usar questionários, análise de documentos e, às vezes, entrevistas. A análise dos dados dos questionários, respondidos com escala *Likert*, normalmente é feita por meio de estatística descritiva, cálculo de médias, desvio padrão e alguma medida de correlação (GROHMANN *et al.*, 2013).

Um dos objetivos da autoavaliação institucional é capturar a percepção dos interessados quanto à qualidade e ao desempenho da instituição para planejar ações que possam aprimorar a gestão (BRASIL, 2004). A percepção quanto à qualidade é considerada um traço latente, pois não é possível medi-la diretamente. A maneira como se tratam os dados coletados, usualmente, aplicando técnicas da Teoria Clássica da Medida (TCM), deixa a interpretação dos dados sujeita a vieses, porque se apoia demasiadamente na subjetividade dos membros da Comissão. Além disso, Sartres e Souza-Formigoni (2013) citam o fato de que todas as medidas dependem da amostra de indivíduos que responderam ao instrumento, ou seja, as avaliações do teste são válidas somente se a amostra for representativa de fato da população do estudo. Outra limitação é que “[...] testes diferentes com índices de dificuldade e discriminação diferentes geram resultados diferentes para os mesmos indivíduos” (p. 243). Se o mesmo construto for medido

por dois testes diferentes, os resultados não estarão expressos na mesma escala, impedindo uma comparação direta e tornando muito difícil uma análise menos subjetiva. O que significa que o resultado medido pela TCM é dependente do teste usado (*test-dependent*) “Seria como dizer em Física que o comprimento do objeto é o que o metro mede ou a massa o que a balança mede” (PASQUALI; PRIMI, 2003, p. 100).

A Teoria da Resposta ao Item (TRI) é uma proposta de instrumento de análise complementar à TCM, que proporciona uma análise centrada nos itens como forma de superar os limites já citados. A TRI permite, ainda, posicionar os itens, de acordo com o seu parâmetro de dificuldade, e os sujeitos, considerando seus escores em uma mesma escala, representando o traço latente que se deseja avaliar. Por isso, acredita-se que a utilização da TRI para a criação de uma escala de avaliação, no lugar do tratamento estatístico atual usado pelas comissões, resultaria em um instrumento mais robusto para a análise e para o desenho de ações de melhorias na gestão institucional.

Em vista do exposto, este trabalho objetiva comparar as informações geradas a partir do uso da TRI em um processo de autoavaliação, já realizado, de uma Ifes com a análise feita pela sua CPA, com uso da TCM. Para isso, foi selecionado o extrato “categoria discente” da avaliação sobre a instituição, no segundo ciclo de 2017.

Além desta introdução, apresentam-se o referencial teórico utilizado, o método desenvolvido na pesquisa, bem como a apresentação dos resultados encontrados e a conclusão, que também traz as limitações da pesquisa e as indicações para futuros estudos.

2 Referencial teórico

A fim de obter embasamento para as comparações do estudo, utilizou-se como referencial as definições de Satisfação Percebida, TCM e TRI.

2.1 Satisfação Percebida

A qualidade do Ensino Superior está relacionada, além de outros fatores, à satisfação dos alunos (CAVALHEIRO; TAVARES; NASCIMENTO FILHO, 2011). As ações para a melhoria da qualidade devem envolver, em seu planejamento, a compreensão das expectativas dos discentes e da diferença que pode ocorrer entre o desejado (expectativas) e a experiência (percepção). A qualidade em uma Instituição de Ensino Superior é um fenômeno complexo. Entre outras coisas, a literatura menciona a importância da melhoria da qualidade

do serviço para a gestão no Ensino Superior. Tal ação deve ser planejada a partir de um diagnóstico feito sobre a definição da qualidade e de como ela é percebida pelos alunos (BRAUM *et al.*, 2015).

A definição de satisfação, descrita por Linder-Pelz (1982), tem a teoria sociopsicológica como base, segundo a qual, a satisfação é a expressão de uma atitude, ou uma resposta afetiva, relacionada com a crença do sujeito de que a sua experiência em dada situação, como a vida universitária, deve possuir certos atributos e, assim, a satisfação é definida como avaliações positivas do indivíduo acerca da sua experiência. Esse atributo é estabelecido de acordo com as expectativas do sujeito (ESPERIDIÃO; TRAD, 2006; LINDER-PELZ, 1982). Sendo um atributo que não pode ser medido diretamente, uma variável latente, torna-se necessário a construção de uma escala de medida para a avaliação desse atributo.

A escala de medida, segundo a qual o atributo pode ser avaliado, é produzida com o uso de instrumentos (como questionários) que se constituem de itens (questões do teste estruturadas de acordo com um modelo), que, segundo a fundamentação teórica utilizada e de acordo com os especialistas consultados, se associam diretamente ao atributo, nesse caso, a satisfação percebida pelos discentes. Assim, as respostas ao instrumento são variáveis indicadores do constructo que se deseja medir (SOARES, 2005).

A autoavaliação no ambiente universitário é necessária para que o processo de ensino se insira na melhoria contínua (MOREIRA JUNIOR *et al.*, 2015). Segundo Souza e Reinert (2010), a satisfação tem grande relevância na avaliação discente, por resultar de um julgamento formulado a partir da realidade percebida. A avaliação pode ser entendida como um processo que gera conhecimento e informação para a tomada de decisões e, ao mesmo tempo, ser um indicador para se avaliar o alcance de metas estabelecidas (MOREIRA JUNIOR, 2014).

No Sinaes, instituído pelo MEC em 2004, está inserida a autoavaliação feita por uma comissão de avaliação interna e com independência em relação à gestão da universidade. Na origem do sistema de avaliação, a autoavaliação seria a porção qualitativa do processo, não sujeitando o processo avaliativo das universidades apenas às pontuações objetivas e burocráticas, por meio de mecanismos baseados em índices. Sendo assim, esta faceta do sistema de avaliação é um de seus principais desafios (LACERDA; FERRI; DUARTE, 2016).

De acordo com Dias Sobrinho (2008):

A avaliação institucional, interna e externa, é o processo central do Sinaes. Consiste em um amplo balanço que cada instituição deve fazer para conhecer-se mais profundamente, refletir sobre suas responsabilidades, seus problemas e potencialidades, enfim, planejar e estabelecer metas para melhorar a qualidade em todas as dimensões institucionais e educativas.

No documento original do Sinaes, percebe-se uma tentativa de aproximar modelos distintos de avaliação, em especial, o qualitativo: “não se baseará exclusivamente em pontuações, pesos e resultados quantitativos; repousará também na avaliação qualitativa” (BRASIL, 2004, p. 131).

A autoavaliação é necessária tanto do ponto de vista de gestão da universidade, para a inserção na melhoria contínua, como também por ser uma exigência legal do MEC. O instrumento utilizado pela CPA para a autoavaliação é um questionário com alternativas de resposta do estilo *Likert* e os dados obtidos são tratados com a TCM, com uso de média, desvio padrão, correlação e, às vezes, análise fatorial. A seguir, apresentam-se a TCM e a TRI, com destaque para as diferenças entre elas.

2.2 Teoria Clássica da Medida - TCM

A TCM compreende um conjunto de conceitos e técnicas relacionadas que serviram de base para numerosos instrumentos de medição. Na TCM, cada item tem sua pontuação independente de outro. O princípio básico dessa teoria é que, quanto mais certos, acredita-se que o domínio de conhecimento que está sendo avaliado seja maior. Fazer inferências sobre coisas que não são diretamente observáveis é um processo imperfeito, as aproximações para variáveis não observáveis, ou traço latente que se quer medir tendem a ser propensos a algum grau de erro (DEVELLIS, 2006).

De acordo com DeVellis (2006), a TCM mantém-se popular apesar do surgimento de novas abordagens de medição, devido a vantagens como familiaridade com seus conceitos básicos, facilidade de tratar os dados computacionais, pois existem diversos programas disponíveis e relativamente fáceis de operar, é um modelo que se ajusta bem a vários instrumentos e os itens “ruins” não necessariamente afetam os demais. No entanto, uma das principais limitações, segundo DeVellis (2006), é que as estimativas de parâmetros sob a TCM dependem da amostra de indivíduos, ou seja, o instrumento de medida depende das características

dos respondentes que atenderam ao teste ou ao questionário (DEVELLIS, 2006; PASQUALI; PRIMI, 2003; REISE; AINSWORTH; HAVILAND, 2005). Não é possível a comparação entre grupos que não tenham feito exatamente o mesmo teste, ou o que se denomina de formas paralelas de testes (ANDRADE; TAVARES; VALLE, 2000).

Na TCM, a avaliação é feita baseada no escore total de um teste, assim, supõe-se que todos os itens sejam adequados, o que nem sempre é verdade (TEZZA; BORNIA, 2009). Uma das grandes vantagens da TRI sobre a TCM é que ela permite a comparação de resultados de grupos diferentes que tenham respondido provas diferentes, desde que as mesmas contenham alguns itens comuns (BORTOLOTTI *et al.*, 2012). Isto porque uma das principais características da TRI é que ela tem como elementos centrais os itens, e não a prova, como um todo (ANDRADE; TAVARES; VALLE, 2000).

Especialmente com relação a esses dois últimos tópicos, a TRI apresenta-se como alternativa, possibilitando a comparação de grupos, pois cria uma escala única para o teste.

2.3 Teoria da Resposta ao Item (TRI)

A TRI é uma ferramenta estatística que possui capacidade para suprir as necessidades decorrentes das limitações da TCM, como a impossibilidade de comparar grupos submetidos a diferentes testes, já citado, bem como de comparar a evolução de um grupo (MOREIRA JUNIOR, 2014).

Pode-se considerar a TRI como um conjunto de modelos matemáticos e estatísticos que (a) analisam itens e escalas, (b) criam medidas e (c) medem indivíduos ou organizações em um construto (traço latente) de interesse. A TRI usa modelagem estatística para relacionar o traço latente estudado com as respostas dos itens, sendo possível avaliar o ajuste do modelo nos níveis global, item ou pessoa (REISE; WALLER, 2009).

Assim, a TRI constitui-se de um conjunto de modelos matemáticos e estatísticos que visa a representar a probabilidade de que um indivíduo, com certo nível de habilidade(s), responda corretamente um item em função de determinados parâmetros (ANDRADE; TAVARES; VALLE, 2000).

Apesar dos primeiros trabalhos a respeito da TRI remontarem ao início do Século XX, por falta de robustez computacional, pois os cálculos para estimação

dos parâmetros exigem muito esforço dessa área, os avanços na TRI só ocorreram mais recentemente, em paralelo com os avanços computacionais (TEZZA; BORNIA, 2009). No Brasil, em 1995, a TRI foi aplicada pela primeira vez na análise dos resultados do Sistema Nacional de Ensino Básico (Saeb) e no Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (Saresp) (ANDRADE; TAVARES; VALLE, 2000).

Apesar de ser largamente aplicada na área de avaliação educacional, a exemplo do Enem, a TRI também vem sendo aplicada a outras áreas, como serviços, gestão pela qualidade total, qualidade de vida, avaliação de intangíveis nas organizações e usabilidade em sites de *e-commerce* (TEZZA; BORNIA, 2009).

Apesar de ter desvantagens, como a sua complexidade e a necessidade de grandes amostras, os benefícios apresentados pela TRI para a mensuração e a avaliação de traços latentes nas áreas da Educação, da saúde e da gestão são capazes de contornar tais desvantagens (TEZZA; BORNIA, 2009).

As principais vantagens com o uso da TRI, apontadas por Tezza e Bornia (2009) são: (a) poder de posicionar indivíduos de diferentes grupos em uma única escala, mesmo que estes tenham respondido a itens diferentes, (b) permitir uma avaliação mais precisa das propriedades dos itens e seus resultados, (c) compreender adequadamente as propriedades psicométricas dos instrumentos, (d) possibilidade de desenvolvimento de indicadores mais eficientes para avaliar diferenças individuais de processo, de práticas, de sistemas ou de indivíduos e (e) maior robustez dos resultados.

Segundo Tezza e Bornia (2009), a TRI traz ao menos cinco avanços em comparação com a psicometria tradicional: (1) o cálculo do nível de traço latente do sujeito independe da amostra de itens utilizados. Ao contrário do que ocorria na psicometria clássica, na qual o escore do sujeito dependia e variava segundo o grau de dificuldade e de precisão do instrumento aplicado, a TRI considera que o indivíduo possui um traço latente “verdadeiro” específico que não irá variar segundo os itens utilizados, desde que os itens estejam medindo o mesmo traço latente; (2) Na TRI, o cálculo dos parâmetros dos itens independe da amostra utilizada, enquanto que, na teoria clássica, os parâmetros dependem de os indivíduos da amostra possuírem maior ou menor nível no traço latente. Isto significa que, mesmo que a amostra não seja representativa, os parâmetros dos itens serão estimados corretamente pela TRI (PASQUALI; PRIMI, 2003); (3) A TRI permite ainda posicionar os itens (considerando seu parâmetro de dificuldade) e os sujeitos (considerando seus escores) em uma mesma escala do

traço latente que está sendo avaliado. Desta forma, é possível estabelecer uma relação entre itens e níveis do traço latente dos indivíduos, identificando os itens que mais bem avaliam cada nível do traço, particularmente útil para se construir a análise do teste; (4) Outro avanço se refere ao fato de que a TRI constitui um modelo que não precisa fazer suposições improváveis, tais como a de que os erros de medida são iguais para todos os indivíduos e (5) A TRI não exige que se trabalhe com testes estritamente paralelos, como exige a teoria clássica (PASQUALI; PRIMI, 2003). Mais uma vantagem para a análise com base na TRI é que ela permite que itens que “não se aplicam” a alguns respondentes não sejam considerados idênticos à resposta nula, distinguindo-se, dessa forma, dos testes clássicos (SARTES; SOUZA-FORMIGONI, 2013).

3 Metodologia e análise dos resultados

Na UFSC, o processo de autoavaliação está instituído desde 2004 e é coordenado e executado pela CPA. Desde 2014, a CPA está vinculada à Reitoria da UFSC e é composta por 12 membros, dentre eles: servidores técnico administrativos, docentes, representantes discentes, representantes dos egressos e representantes da sociedade civil organizada (UNIVERSIDADE FEDERAL DE SANTA CATARINA, 2017). Em 2017, a autoavaliação ocorreu em dois ciclos, cada um abrangendo temas específicos. A coleta de dados ocorreu por meio de plataforma digital, de questionários respondidos pelos grupos de interesse: discentes, docentes, técnicos administrativos e a categoria gestores, que apesar de ser composta por docentes e técnicos administrativos, por estarem na posição de gestores, considera-se que possuem uma relação diferenciada com os temas (UNIVERSIDADE FEDERAL DE SANTA CATARINA, 2017). A CPA, de posse dos dados, coletados por meio das respostas aos questionários e das informações solicitadas a outros setores da Universidade, bem como de pesquisa em documentos institucionais, elabora o Relatório Integral de Autoavaliação. A maneira como vêm sendo tratados os dados coletados pela CPA da UFSC deixa sua interpretação sujeita a vieses dos membros da comissão, porque está demasiadamente baseada na subjetividade.

Para a utilização da TRI, é preciso que se defina antes o traço latente, ou seja, aquilo que se pretende medir. A TRI faz a análise dos itens e, a partir daí, constrói a escala padronizada. Na maioria das vezes, a elaboração dos itens é realizada com auxílio de especialistas ou extraída de literatura específica; no caso deste estudo, os itens já haviam sido elaborados pela CPA. O instrumento de coleta da CPA é dividido entre as categorias (docentes, técnicos e discentes); para este estudo, isolou-se a avaliação discente sobre a instituição. A coleta foi realizada no período de 27 de novembro de 2017 a 09 de março de 2018 (103 dias) e os questionários foram respondidos por 4.850 estudantes.

Foram respondidos 16 itens na categoria dos discentes. As alternativas de resposta para os itens eram: “Excelente”, “Muito Boa”, “Suficiente”, “Insuficiente”, “Inexistente”, “Desconheço” e “Não se Aplica”. Ou outra forma adaptada às perguntas, como por exemplo: “Concordo Fortemente”, “Concordo”, “Indiferente”, “Discordo” e “Discordo Fortemente”. Para cada alternativa foi atribuída uma nota, respectivamente 5, 4, 3, 2 e 1, restando “Não se Aplica” e “Desconheço”, sem valores correspondentes. Os dados obtidos foram processados por meio do *software* R (pacotes *MIRT*; *Psych*; *MSM* e *DirectLabels*). Uma consideração importante para a escolha do modelo da TRI é a dimensionalidade do traço latente. Quando o conjunto de itens de um instrumento mede um único traço latente, utiliza-se um modelo unidimensional da TRI. Por outro lado, quando os itens são explicados por mais de um traço latente, o mais adequado é utilizar um modelo multidimensional ou analisar as dimensões em separado. Normalmente, utiliza-se a Análise Fatorial para estudar a dimensionalidade do traço latente (COSTA, 2015).

Foi utilizado o método de análise fatorial de informação completa para a verificação de dimensionalidade, esse método é adequado para modelos politômicos e foi desenvolvido por Bock e Aitkin (1981). A análise fatorial de informação completa baseia-se nas estimativas de máxima verossimilhança marginal e em um algoritmo chamado “EM” (BOCK; GIBBONS; MURAKI, 1988). Com auxílio do *software* R, foram ajustados modelos considerando uma a quatro dimensões, que foram comparados utilizando-se os critérios de informação de Akaike (AIC e AICc) e Bayesiano (BIC). A Tabela 1 apresenta as estatísticas AIC, AICc e BIC para os quatro modelos.

Tabela 1 - Definição do número de dimensões

Dimensões	AIC	AICc	BIC
1	131736.2	131738.5	132225.3
2	127888.9	127892.2	128469.3
3	126892.1	126896.4	127557.3
4	126900.8	126906.2	127644.3

Fonte: Elaboração dos autores (2018).

Observa-se, na Tabela 1, que, para os três critérios, o menor valor corresponde ao modelo com três dimensões, o que indica que essa é a melhor opção. Na análise dos itens, também se identificam três dimensões: a percepção de “satisfação” do discente, a percepção de adequação das ações de “assistência estudantil” e a percepção de alinhamento da universidade com a “sustentabilidade”. No Quadro 1 tem-se os itens identificados com as suas dimensões.

Quadro 1 - Os 15 itens separados pelas dimensões identificadas

F1 SUSTENTABILIDADE	F2 SATISFAÇÃO	F3 ASSISTÊNCIA ESTUDANTIL
2 - Como você considera a contribuição da UFSC para a inclusão social?	1 - A Universidade tem superado minhas expectativas.	11 - Avalie a política de benefícios, assistência e apoio financeiro e/ou institucional.
3 - Como você considera a contribuição da UFSC para a preservação do meio ambiente?	4 - Ter escolhido estudar na UFSC foi a decisão certa.	12 - Avalie as ações da Prae.
5 - Como você considera a contribuição da UFSC para o desenvolvimento local e regional?	6 - Eu acredito que a Universidade me possibilita uma boa formação profissional e humana.	13 - Avalie o horário e o atendimento da Prae.
15 - Como você considera a atuação interna da UFSC com relação à sustentabilidade?	7 - Estou feliz com a minha experiência universitária.	14 - Avalie o horário e o atendimento do Restaurante Universitário.
16 - Eu tenho atitudes mais sustentáveis por influência das campanhas educativas veiculadas pela UFSC.	8 - Eu indicaria a outras pessoas que estudassem na UFSC.	
	9 - De um modo geral estou satisfeito com a Universidade.	

Fonte: Elaboração dos autores (2018).

Para análise de instrumentos multidimensionais, alguns autores utilizam múltiplas análises unidimensionais, uma para cada dimensão (Sartes & Souza-Formigoni, 2013). Para os objetivos deste estudo, foi escolhida a dimensão “Satisfação”. Portanto, no presente estudo, foram consideradas apenas as questões que procuram medir o nível de satisfação dos estudantes (os itens 1, 4, 6, 7, 8 e 9 do Quadro 1).

Primeiramente, foi usada a TCM. Segundo Andrade, Tavares e Valle (2000), essa fase serve para verificar se a leitura dos dados está correta e essas métricas são usadas como valores iniciais para os processos de estimação realizados nas fases seguintes. A Tabela 2 apresenta as estatísticas utilizadas para a análise clássica.

Tabela 2 - Estatísticas descritivas obtidas na TCM

	1	4	6	7	8	9
Média	3,3812	4,2820	4,1220	3,8059	4,3251	3,8103
Pearson	0,7497	0,7651	0,7233	0,7845	0,7701	0,7806
Variância	1,0832	0,6210	0,7121	1,0089	0,6207	0,8971
Alfa de Cronbach	0,9353					

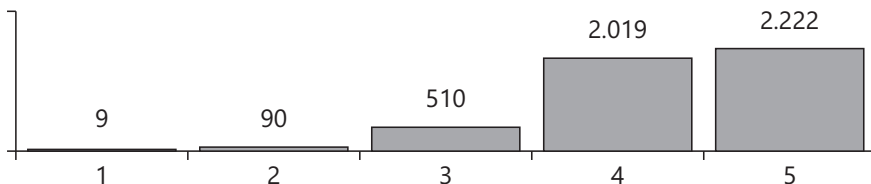
Fonte: Elaboração dos autores (2018).

Os itens com a maior média de escore são os itens 4 e 8. Esses itens apresentam o maior percentual de respostas 5, o nível mais satisfeito (4 - Ter escolhido estudar na UFSC foi a decisão certa; e 8 - Eu indicaria a outras pessoas que estudassem na UFSC). Os itens com os menores graus de satisfação foram os itens 1 e 7, que correspondem às menores médias de escore (1 - A Universidade tem superado minhas expectativas; e 7 - Estou feliz com a minha experiência universitária).

Os coeficientes de correlação são altos, indicando que todos os itens são coerentes com o conjunto de itens que medem o desempenho docente. A consistência interna pode ser verificada por meio da determinação do alfa (α) de Cronbach. Ele mede o grau de covariância de uma série de itens e varia de 0 a 1 e, quanto mais elevada a contagem, maior a confiabilidade da escala. De acordo com Pasquali (2003), quando o resultado do coeficiente se aproxima de um, pode-se afirmar que o teste possui um coeficiente de precisão. O valor de 0,93 indica consistência do conjunto.

A distribuição dos respondentes pelos níveis de satisfação foi feita pelo escore médio de cada discente, ela indica que a maioria dos respondentes se encontra no nível 5 (“muito boa”) representando, aproximadamente, 50% do total de respondentes.

Figura 1 - Distribuição dos respondentes pelas categorias de respostas (TCM)



Fonte: Elaboração dos autores (2018).

Ressalta-se que o principal interesse da TCM é o escore total, explicação do resultado final de um teste. O foco da TRI é o traço latente e não o escore final do teste.

Para a análise por meio da TRI, foi utilizado o Modelo de Resposta Gradual de Samejima da TRI, no qual a discriminação de uma categoria específica de resposta depende do parâmetro de discriminação “a”, comum a todas as categorias do item, e, necessariamente, as categorias são ordenadas quanto ao nível de dificuldade, ou seja, $b_{i,1} \leq b_{i,2} \leq b_{i,3} \dots \leq b_{i,m}$ (ANDRADE; TAVARES; VALLE, 2000; MOREIRA JUNIOR, 2014). A probabilidade de um indivíduo j receber um escore k no item i é dada pela seguinte expressão:

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

Onde:

$i = 1, 2, \dots, I$ (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 ;

θ_j representa a habilidade (traço latente) do j -ésimo indivíduo

$P_{i,k}(\theta_j)$ é a probabilidade de um indivíduo j com habilidade θ escolher a categoria k do item i .

$P_{i,k}^+(\theta_j)$ é a probabilidade de um indivíduo j com habilidade θ escolher a categoria k ou uma categoria superior do item i .

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

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

Observa-se que, em um item com $(m_i + 1)$ categorias, m_i valores de dificuldade necessitam ser estimados. Assim, para cada item, o número de parâmetros b a ser estimado será dado pelo seu número de categorias de resposta menos um (MOREIRA JUNIOR *et al.*, 2015). Cada item possui cinco possibilidades de resposta; dessa maneira, para cada item, foram estimados quatro parâmetros “ b ” e um parâmetro “ a ”. Na Tabela 3, pode-se ver os parâmetros estimados dos itens 1, 4, 6, 7, 8 e 9.

Tabela 3 - Parâmetros dos itens e respectivos erros padrões

	a	EP(a)	b1	EP(b1)	b2	EP(b2)	b3	EP(b3)	b4	EP(b4)
X1	2,1970	0,0589	-2,2509	0,0543	-0,9299	0,0277	-0,1521	0,0224	1,5578	0,0373
X4	2,9362	0,0904	-2,9792	0,0920	-2,0799	0,0470	-1,3226	0,0305	0,1551	0,0210
X6	2,3028	0,0663	-2,8692	0,0806	-1,8818	0,0443	-1,3408	0,0329	0,5121	0,0245

Continua

Continuação

X7	2,5612	0,0702	-2,3525	0,0567	-1,3202	0,0313	-0,7561	0,0242	0,8434	0,0264
X8	3,3778	0,1116	-2,6450	0,0693	-2,0524	0,0449	-1,3521	0,0298	0,0549	0,0200
X9	2,6868	0,0757	-2,4590	0,0600	-1,3578	0,0313	-0,8307	0,0245	0,9721	0,0272

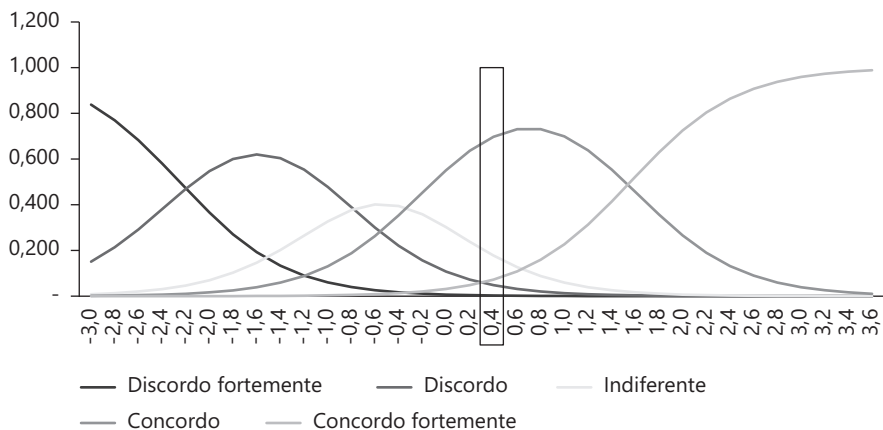
Fonte: Elaboração dos autores (2018).

Conforme se percebe pelos erros padrões, presentes na Tabela 3, o processo de estimação pode ser considerado confiável, pois os erros padrões das estimativas se mantêm consistentes, na mesma magnitude, para todos os parâmetros, não havendo valores discrepantes entre eles. A métrica utilizada no modelo e nas análises foi (0,1), média = 0 e desvio padrão = 1 (ANDRADE; TAVARES; VALLE, 2000).

Observando a Tabela 3, percebe-se que os itens mais fáceis são o 4 e o 8, e os mais difíceis, que necessitam de um *theta* maior, são os itens 1 e o 9. Pela análise feita com a TCM, os maiores escores, itens mais fáceis, foram o 4 e 8, mas os menores escores foram os itens 1 e 7, divergindo um pouco da análise pela TRI. Pelas médias, os itens 7 e 9 são mais próximos do que quando analisados pelos parâmetros “b”, pois, análise com a TRI gera mais informação para analisar o item e as suas nuances.

Na Figura 2, estão representadas as curvas de probabilidade calculadas no Excel para as categorias de respostas do item 1 (“A Universidade tem superado minhas expectativas”). Os valores de θ aparecem no eixo X e as probabilidades, no eixo Y.

Figura 2 - Curva Característica do item 1

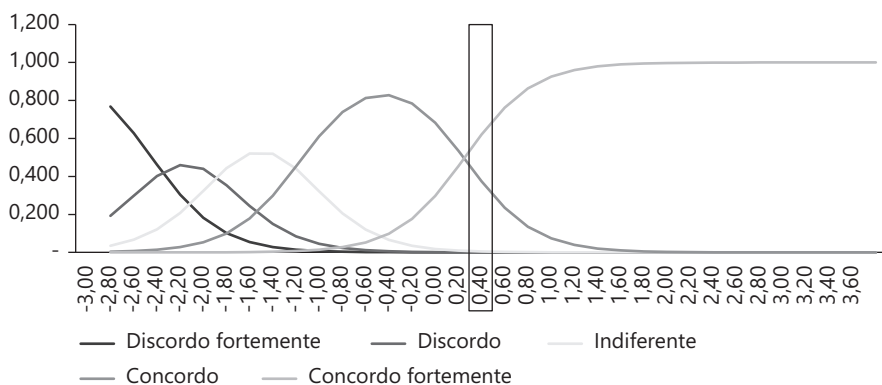


Fonte: Elaboração dos autores (2018).

Pode-se ler a Figura 2 como: respondentes que possuem *theta* inferior a -2,2 possuem maior probabilidade de responder a categoria 1 (“discordo fortemente”); respondentes com *thetas* entre -2,2 e -0,8 possuem maior probabilidade de responder a categoria 2 (“discordo”); respondentes com *thetas* entre -0,8 e -0,3, possuem maior probabilidade de responder a categoria 3 (“indiferente”); respondentes com *thetas* entre -0,3 e 1,6 apresentam probabilidade maior de responder a categoria 4 (“concordo”); e, por fim, respondentes que apresentam *thetas* superiores a 1,6 têm maior probabilidade de responder a categoria 5 (“concordo fortemente”).

Ou, ainda, tomando como exemplo um indivíduo com *theta* igual a 0,4, como assinalado na Figura 2, no item 1, esse indivíduo tem aproximadamente 7,3% de probabilidade de estar entre os que marcam “concordo fortemente” com a afirmação do item. De modo semelhante, ele teria 69,8% de probabilidade de estar entre os que escolheriam “concordo”, 1,78% de marcar “indiferente”, 4,8% de escolher “discordar” e 0,3% chance de “discordar fortemente”.

Figura 3 - Curva de probabilidade do item 8

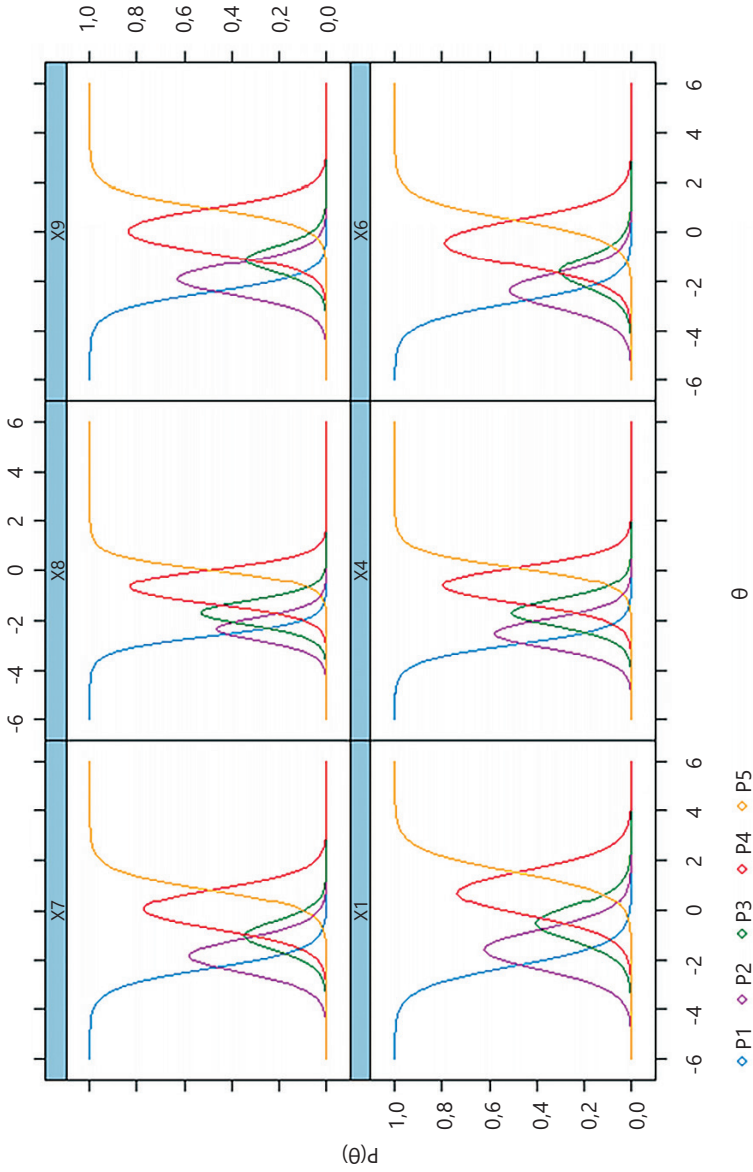


Fonte: Elaboração dos autores (2018).

Seguindo o mesmo exemplo, porém para o item 8, na Figura 3, o mesmo indivíduo de *theta* igual a 0,4 teria 76,2% de probabilidade de “concordar fortemente” com a afirmação do item e 23,5% de “concordar”, 0,24% de marcar “indiferente”, 0,022% e 0,003% de escolher “discordo” e “discordo fortemente” respectivamente.

Na Figura 4, pode-se ver as curvas de probabilidade para cada um dos itens.

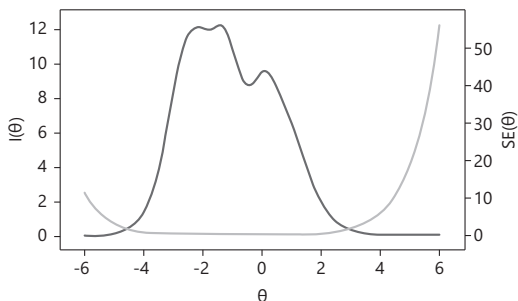
Figura 4 - Curva característica de todos os itens



Fonte: Elaboração dos autores (2018).

Na Figura 5, está a Curva de Informação do Teste. Esta curva é uma representação gráfica da função de informação do teste. Esta função é representada pela soma das informações do grupo de itens que compõem o teste. O erro padrão das estimativas também é apresentado na Figura 5.

Figura 5 - Curva de informação do teste e erro padrão



Fonte: Elaboração dos autores (2018).

Pode-se ver que o teste possui mais informação (menor erro padrão para as estimativas das satisfações) entre -4 e 2.

O próximo passo é interpretar a escala (0,1) criada (ANDRADE; TAVARES; VALLE, 2000). Para isso, utilizam-se os chamados “níveis âncora”, definidos por Andrade, Tavares e Valle (2000) como pontos selecionados na escala da habilidade para serem interpretados pedagogicamente, sendo a ligação entre os valores mensurados e os conteúdos envolvidos na avaliação, no presente caso, a satisfação percebida pelo discente. Para um item ser considerado âncora em um determinado nível, ele precisa ser respondido positivamente por um grande percentual de indivíduos (por exemplo, acima de 65%) com este nível de habilidade e por um pequeno percentual de indivíduos (por exemplo, abaixo de 50%) com o nível de habilidade imediatamente anterior. A identificação de níveis âncora e os itens que os representam auxiliam na construção e na interpretação da escala.

Com os itens âncoras identificados para os níveis da escala, os especialistas no traço latente estudado caracterizam os níveis de acordo com o conteúdo abordado pelos itens posicionados em cada nível. Dessa forma, um sujeito situado em um determinado nível possui habilidade caracterizada pelos itens posicionados neste nível e nos níveis inferiores e não possui habilidade para os conteúdos caracterizados pelos itens posicionados nos níveis superiores a esse (VALLE, 2001). Como os itens deste estudo são ordinais, o raciocínio citado aplica-se para as categorias dos itens, denominadas “categorias âncora”. A Tabela 4 apresenta as probabilidades de ocorrência das categorias ou categorias superiores dos itens ($P_{ik}^+(\theta_j)$).

Tabela 4 - Probabilidades de ocorrência das categorias ou categorias superiores dos itens.

	a	b	-3,40	-3,15	-2,90	-2,65	-2,40	-2,15	-1,90	-1,65	-1,40	-1,15	-0,90	-0,65	-0,40	-0,15	0,10	0,35	0,60	0,85	1,10	1,35	1,60	1,80	
X4	2,936	b1 -2,979	0,225	0,377	0,558	0,724	0,846	0,919	0,960	0,980	0,990	0,995	0,998	0,999	0,999	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
X7	2,561	b1 -2,979	0,254	0,392	0,551	0,699	0,815	0,893	0,941	0,968	0,983	0,991	0,995	0,997	0,999	0,999	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
X6	2,303	b1 -2,869	0,228	0,344	0,482	0,624	0,747	0,840	0,903	0,943	0,967	0,981	0,989	0,994	0,997	0,998	0,999	0,999	1,000	1,000	1,000	1,000	1,000	1,000	1,000
X8	3,378	b1 -2,645	0,072	0,154	0,297	0,496	0,696	0,842	0,925	0,966	0,985	0,994	0,997	0,999	0,999	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
X9	2,687	b1 -2,459	0,074	0,135	0,234	0,374	0,540	0,696	0,818	0,898	0,945	0,971	0,985	0,992	0,996	0,998	0,999	0,999	1,000	1,000	1,000	1,000	1,000	1,000	1,000
X1	2,197	b1 -2,251	0,074	0,122	0,194	0,294	0,419	0,555	0,684	0,789	0,866	0,918	0,951	0,971	0,983	0,990	0,994	0,997	0,998	0,999	0,999	1,000	1,000	1,000	1,000
X4	2,936	b2 -2,080	0,020	0,041	0,083	0,158	0,281	0,449	0,629	0,779	0,880	0,939	0,970	0,985	0,993	0,997	0,998	0,999	1,000	1,000	1,000	1,000	1,000	1,000	1,000
X7	2,561	b2 -2,080	0,033	0,061	0,109	0,188	0,306	0,455	0,613	0,750	0,851	0,915	0,954	0,975	0,987	0,993	0,996	0,998	0,999	0,999	1,000	1,000	1,000	1,000	1,000
X8	3,378	b2 -2,052	0,010	0,024	0,054	0,117	0,236	0,418	0,626	0,796	0,901	0,955	0,980	0,991	0,996	0,998	0,999	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
X6	2,303	b2 -1,882	0,029	0,051	0,087	0,146	0,233	0,350	0,490	0,630	0,752	0,844	0,906	0,945	0,968	0,982	0,990	0,994	0,997	0,998	0,999	0,999	1,000	1,000	1,000
X9	2,687	b2 -1,358	0,004	0,008	0,016	0,030	0,057	0,106	0,189	0,313	0,472	0,636	0,774	0,870	0,929	0,962	0,980	0,990	0,995	0,997	0,999	0,999	1,000	1,000	1,000
X8	3,378	b3 -1,352	0,001	0,002	0,005	0,012	0,028	0,063	0,136	0,268	0,460	0,664	0,822	0,915	0,961	0,983	0,993	0,997	0,999	0,999	1,000	1,000	1,000	1,000	1,000
X6	2,303	b3 -1,341	0,009	0,015	0,027	0,047	0,080	0,134	0,216	0,329	0,466	0,608	0,734	0,831	0,897	0,939	0,965	0,980	0,989	0,994	0,996	0,998	0,999	0,999	0,999
X4	2,936	b3 -1,323	0,002	0,005	0,010	0,020	0,041	0,081	0,155	0,277	0,443	0,624	0,776	0,878	0,938	0,969	0,985	0,993	0,996	0,998	0,999	1,000	1,000	1,000	1,000
X7	2,561	b3 -1,323	0,005	0,009	0,017	0,032	0,060	0,107	0,186	0,302	0,451	0,609	0,747	0,848	0,914	0,953	0,975	0,986	0,993	0,996	0,998	0,999	0,999	1,000	1,000
X1	2,197	b2 -0,930	0,004	0,008	0,013	0,022	0,038	0,064	0,106	0,171	0,263	0,381	0,516	0,649	0,762	0,847	0,906	0,943	0,966	0,980	0,989	0,993	0,996	0,998	0,998
X9	2,687	b3 -0,831	0,001	0,002	0,004	0,007	0,015	0,028	0,054	0,100	0,178	0,298	0,454	0,619	0,761	0,862	0,924	0,960	0,979	0,989	0,994	0,997	0,999	0,999	0,999
X1	2,197	b3 -0,152	0,001	0,001	0,002	0,004	0,007	0,012	0,021	0,036	0,061	0,100	0,162	0,251	0,367	0,501	0,635	0,751	0,839	0,900	0,940	0,964	0,979	0,986	0,986
X8	3,378	b4 0,055	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,003	0,007	0,017	0,038	0,085	0,177	0,334	0,538	0,730	0,863	0,936	0,972	0,988	0,995	0,997	0,997
X4	2,936	b4 0,155	0,000	0,000	0,000	0,000	0,001	0,001	0,002	0,005	0,010	0,021	0,043	0,086	0,164	0,290	0,460	0,639	0,787	0,885	0,941	0,971	0,986	0,992	0,992
X7	2,561	b4 0,155	0,000	0,000	0,000	0,001	0,001	0,003	0,005	0,010	0,018	0,034	0,063	0,113	0,194	0,314	0,465	0,622	0,758	0,856	0,918	0,955	0,976	0,985	0,985
X6	2,303	b4 0,512	0,000	0,000	0,000	0,001	0,001	0,002	0,004	0,007	0,012	0,021	0,037	0,064	0,109	0,179	0,279	0,408	0,550	0,685	0,795	0,873	0,925	0,951	0,951
X9	2,687	b4 0,972	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,003	0,003	0,006	0,013	0,024	0,047	0,088	0,158	0,269	0,419	0,585	0,734	0,844	0,902	0,902
X1	2,197	b4 1,558	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,002	0,003	0,004	0,008	0,013	0,023	0,039	0,066	0,109	0,174	0,268	0,388	0,523	0,630	0,630

Fonte: Elaboração dos autores (2018).

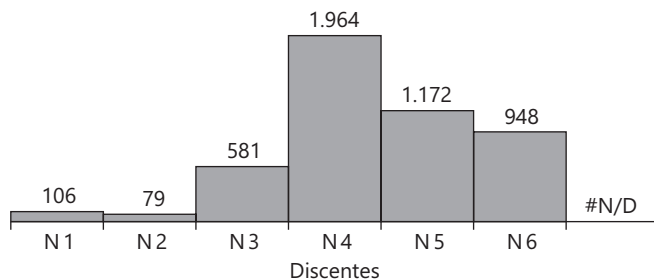
Observando-se as probabilidades e realizando-se a leitura, com análise dos itens, foram definidos seis níveis para a interpretação, conforme Quadro 2.

Quadro 2 - Interpretação da escala

Nível de satisfação	Interpretação
Nível 1 – N1 Muito Insatisfeitos (<i>theta</i> até -1,9)	Discordam de todos os itens.
Nível 2 – N2 Insatisfeitos (<i>theta</i> de -1,9 a -1,65)	Ainda discordam de que a decisão de estudar na UFSC foi acertada, de que estão satisfeitos com a Universidade e de a UFSC esteja superando as expectativas. Passam a indiferentes quanto a formação profissional e humana da UFSC ser boa, de indicar a UFSC a outras pessoas e de estarem felizes com a experiência universitária.
Nível 3 – N3 Indiferentes (<i>theta</i> de -1,65 a 0,9)	Discordam ainda de que a Universidade tem superado suas expectativas. São indiferentes quanto a estarem satisfeitos com a UFSC. E passam a concordar quanto a: ter escolhido estudar na UFSC ter sido a decisão certa, acreditar que a Universidade possibilita uma boa formação profissional e humana, estarem felizes com a experiência universitária e indicarem a outras pessoas que estudem na UFSC.
Nível 4 – N4 (<i>theta</i> de 0,9 a 0,1) Pouco Satisfeitos	Permanecem concordando quanto a: ter escolhido estudar na UFSC ter sido a decisão certa, acreditar que a Universidade possibilita uma boa formação profissional e humana, estarem felizes com a experiência universitária e indicarem a outras pessoas que estudem na UFSC. E passam a concordar quanto a estarem satisfeitos com a UFSC e quanto a UFSC superar suas expectativas.
Nível 5 – N5 Satisfeitos (<i>theta</i> de 0,1 a 0,85)	Permanecem concordando quanto a estarem satisfeitos com a UFSC e quanto a UFSC superar suas expectativas. Passam a concordar fortemente que a escolha da UFSC foi a certa, que estão felizes com a experiência universitária, que a UFSC possibilita boa formação profissional e humana e indicariam a UFSC a outras pessoas. (satisfeito)
Nível 6 – N6 Muito Satisfeitos (<i>theta</i> acima de 0,85)	Continuam concordando fortemente que a escolha da UFSC foi a certa, que estão felizes com a experiência universitária, que a UFSC possibilita boa formação profissional e humana e indicariam a UFSC a outras pessoas. Passam a concordar fortemente que a Universidade tem superado suas expectativas e admitem estarem satisfeitos com a UFSC.

Fonte: Elaboração dos autores (2018).

A distribuição dos respondentes pelos níveis de satisfação está representada na Figura 6.

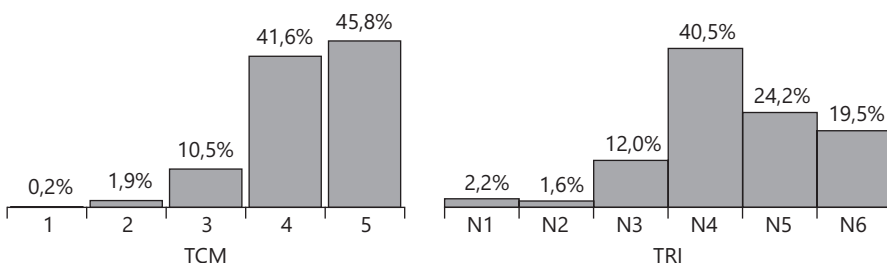
Figura 6 - Distribuição de frequência dos respondentes pelos níveis da escala (TRI)

Fonte: Elaboração dos autores (2018).

Pode-se notar que o número maior de estudantes encontra-se no nível “4”, concordando com os itens, parecendo pouco satisfeitos a satisfeitos. Observando os níveis superiores “5” e “6”, percebe-se que ainda há ações que se pode implantar para que os estudantes percebam que a UFSC proporciona uma boa formação, para que sintam que fizeram a escolha certa, e que sua satisfação de modo geral aumente.

Observa-se que, com a ancoragem das categorias dos itens, a interpretação do teste, como um todo, torna-se muito mais rica. Pequenas nuances podem ser observadas e, assim, ações estratégicas podem ser traçadas para se atingir, com mais precisão, os objetivos de melhoria na satisfação dos discentes.

Ao se comparar a distribuição dos respondentes pela TCM na Figura 1 com a distribuição na Figura 7, elaborada com a TRI, nota-se que os níveis mais altos do traço latente ganham mais detalhamento. Assim, por meio da leitura e da análise dos itens que compõem os níveis N4, N5 e N6, os gestores são capazes de planejar melhor as ações futuras.

Figura 7 - Comparação da TCM e TRI

Fonte: Elaboração dos autores (2018).

4 Conclusão

O estudo presente buscou a comparação de dois métodos de tratamento de dados estatísticos. Foram trabalhados os dados coletados por meio do instrumento executado pela CPA da UFSC, em 2017. Aplicou-se a TCM, que a CPA utiliza em seu relatório, e a TRI nos mesmos dados coletados. Ambos tratamentos estatísticos apresentam coerência com o que se buscou medir.

Os itens avaliados mostraram-se adequados, uma vez que possuem bom poder discriminatório, os parâmetros de dificuldade não apresentaram valores extremos e os erros padrões das estimativas estavam consistentes, não havendo nenhum que se destacasse dos demais. Esta análise foi possível com a TRI.

O questionário conseguiu estimar com eficiência o nível de satisfação de indivíduos que estavam posicionados entre -3,4 e 1,8 na escala do traço latente. Com a curva de informação do teste, pode-se visualizar o intervalo da escala que gerava estimativas mais precisas (com menores erros padrões). Esta análise permitiu acompanhar o ganho de informações obtido por itens extras inserido no formulário, bem como sinalizar quais intervalos careciam de informação e, portanto, novos itens seriam particularmente mais úteis para melhorar o instrumento.

O trabalho de definição dos níveis âncora, com seus respectivos itens, permitiu uma análise aprofundada para a interpretação da escala, levando em conta os conteúdos e analisando os itens e a informação trazida por eles, superando a subjetividade da TCM na divisão de níveis de categorização e sua distribuição de frequência matemática.

De acordo com o objetivo do trabalho, ficou demonstrado que o resultado do tratamento fornecido pela TRI, pelo MRG de Samejima, mostra-se mais eficiente para que o gestor perceba que pontos podem ser alvo de melhorias, pois a interpretação dos itens na mesma escala que os respondentes fornecem mais informação e detalhes de acordo com os itens e as categorias analisados possibilitando “ajustes finos” no plano de ação da gestão.

Não se pode ignorar as simplificações feitas em prol da comparação, nesse caso específico, pois os dados já estavam coletados e disponíveis para os pesquisadores. Sendo essa uma limitação a ser destacada no presente estudo, mas que não invalida a comparação entre a TCM e a TRI, e as vantagens da segunda para a tomada de decisões da gestão.

Estudos futuros que abranjam uma quantidade maior de itens devem ser realizados, a fim de que possam comparar, no tempo, o desenvolvimento do traço latente, e, também, que tratem as demais dimensões.

Student satisfaction evaluation of a higher education institution: an analysis of the methods of the Classical Theory of Measurement and of the Item Response Theory

Abstract

The present work compares the information generated by the Institutional Self-evaluation process of an Federal Institution of Higher Education (Ifes), in the context of the National Higher Education Evaluation System (Sinaes), and analyzed by the IFES Self-evaluation Committee, cycle of 2017, with the same information treated with the Item Response Theory (IRT) in place of the Classical Theory of Measurement, with the objective of demonstrating how the analysis process can benefit from the use of IRT, being a more robust instrument for the analysis and design of actions of improvements for institutional management. For this, we selected the "student category" extract from the evaluation about the institution, in the second cycle of 2017. The two theories, Classical and IRT, were described and compared. With the data collected from the institution, the information curves of the items and the test were calculated and explained, and through the analysis provided by the IRT was demonstrated the gain that can be obtained with its use in institutional management.

Keywords: Institutional Self-evaluation. Item Response Theory. University Management.

Evaluación de la satisfacción del estudiante de una institución de Educación Superior: un análisis de la Teoría Clásica de la Medición y la Teoría de la Respuesta al Ítem

Resumen

Este trabajo compara la información generada por el proceso de autoevaluación de una Institución Federal de Educación Superior (Ifes), en el contexto del Sistema Nacional de Evaluación de la Educación Superior (Sinaes), analizada por la Comisión de Evaluación Propia de Ifes, ciclo 2017, con la misma información tratada a partir del uso de la Teoría de Respuesta al Ítem (TRI) en lugar de la Teoría Clásica de la Medida, con el fin de demostrar cómo el proceso de análisis puede beneficiarse del uso de la TRI, siendo un instrumento más robusto para el análisis y la diseño de acciones de mejora en la gestión institucional. Para ello se seleccionó el extracto "categoría discente" de la evaluación de la institución, en el segundo ciclo de 2017. Se describieron y compararon las dos teorías, la clásica y la TRI. Con los datos recolectados de la institución se calcularon y explicaron las curvas de información del ítem y de la prueba, mediante el análisis proporcionado por el TRI se demostró la ganancia que se puede obtener con su uso en la autoevaluación institucional.

Palabras clave: Autoevaluación Institucional. Teoría de la Respuesta al Ítem. Gestión Universitaria.

Referências

- ANDRADE, D. F. D.; TAVARES, H. R.; VALLE, R. D. C. *Teoria da resposta ao item: conceitos e aplicações*. São Paulo: Associação Brasileira de Estatística, 2000.
- BOCK, R. D.; AITKIN, M. Marginal maximum likelihood estimation of item parameters: application of an em algorithm. *Psychometrika*, [s. l.], v. 46, n. 4, p. 443-459, Dec. 1981. <https://doi.org/10.1007/BF02293801>
- BOCK, R. D.; GIBBONS, R.; MURAKI, E. Full-information item factor analysis. *Applied Psychological Measurement*, v. 12, n. 3, p. 261-280, Sep.1988. <https://doi.org/10.1177/014662168801200305>
- BORTOLOTTI, S. L. V. *et al.* Avaliação do nível de satisfação de alunos de uma instituição de Ensino Superior: uma aplicação da teoria da resposta ao item. *Gestão e Produção*, São Carls, v. 19, n. 2, p. 287-302, 2012. <https://doi.org/10.1590/S0104-530X2012000200005>
- BRASIL. Lei nº 10.861, de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior -SINAES e dá outras providências. *Diário Oficial da União*, Brasília, DF, 15 abr. 2004.
- BRAUM, L. M. S. *et al.* Percepção dos alunos sobre qualidade no Ensino Superior: identificação das prioridades de melhoria no Curso de Ciências Contábeis. *Revista de Estudos Contábeis*, Londrina, v. 6, n. 11, p. 62-80, 2015.
- CAVALHEIRO, E. A.; TAVARES, C. E. M.; NASCIMENTO FILHO, M. O que determina a satisfação de nossos alunos? Um estudo comparativo entre o Curso de Administração e os demais Cursos do Centro de Ciências Sociais Aplicadas da Universidade de Cruz Alta. In: CONGRESSO UFSC DE CONTROLADORIA E FINANÇAS, 4., Florianópolis, 2011. *Anais[...]*. Florianópolis: UFSC, 2011.
- COSTA, C. E. S. *Análise da dimensionalidade e modelagem multidimensional pela tri no ENEM (1998-2008)*. Dissertação – Programa de Pós-Graduação em Métodos e Gestão em Avaliação, Universidade Federal de Santa Catarina, Florianópolis, 2015.
- DEVELLIS, R. F. Classical test theory. *Medical Care*, Hagerstown, v. 44, n. 11, p. S50-S59, Nov. 2006. 10.1097/01.mlr.0000245426.10853.30

DIAS SOBRINHO, J. Qualidade, avaliação: do Sinaes a índices. *Avaliação: Revista da Avaliação da Educação Superior*, Campinas, v. 13, n. 3, p. 817-825, nov. 2008. <https://doi.org/10.1590/S1414-40772008000300011>

ESPERIDIÃO, M. A.; TRAD, L. A. B. Avaliação de satisfação de usuários: considerações teórico-conceituais. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 22, n. 6, p. 1267-1276, Jun. 2006. <https://doi.org/10.1590/S0102-311X2006000600016>

GALDINO, M. N. D. A autoavaliação institucional no ensino superior como instrumento de gestão. In: CONGRESSO IBERO-AMERICANO DE POLÍTICA E ADMINISTRAÇÃO DA EDUCAÇÃO, 2., 2011 São Paulo. *Anais[...]* São Paulo: PUC-USP, 2011. Disponível em: http://www.unigranrio.com.br/_docs/cpa/autoav-inst-ensino-sup-instr-gestao-mary-galdino.pdf. Acesso em: 15 set 2018.

GROHMANN, M. Z. *et al.* Uma década de auto avaliação institucional: o que mudou na percepção dos docentes? *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, Madrid, v. 12, n. 1, p. 5-23, enero 2016.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA – INEP. *Nota técnica INEP/DAES/CONAES nº 065*. Roteiro para Relatório de Autoavaliação Institucional. Brasília, DF, 2014.

LACERDA, L. L. V. D.; FERRI, C.; DUARTE, B. K. D. C. SINAES: avaliação, accountability e desempenho. *Avaliação: Revista da Avaliação da Educação Superior*, Campinas, v. 21, n. 3, p. 975-992, ago./nov. 2016. <https://doi.org/10.1590/S1414-40772016000300015>

LINDER-PELZ, S. Toward a theory of patient satisfaction. *Social Science & Medicine*, Oxford, v. 16, n. 5, p. 577-582, 1982. [https://doi.org/10.1016/0277-9536\(82\)90311-2](https://doi.org/10.1016/0277-9536(82)90311-2)

MOREIRA JUNIOR, F. J. Contribuições da Teoria da Resposta ao Item nas avaliações educacionais. *Ciência e Natura*, Santa Maria, v. 36, n. ed. especial, p. 58-72, 2014. <https://doi.org/10.5902/2179460X13120>

MOREIRA JUNIOR, F. J. *et al.* Avaliação da satisfação de alunos por meio do Modelo de Resposta Gradual da Teoria da Resposta ao Item. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 23, n. 86, p. 129-158, jan./fev. 2015. <https://doi.org/10.1590/S0104-40362015000100005>

PASQUALI, L.; PRIMI, R. Fundamentos da Teoria da Resposta ao Item –TRI. *Avaliação Psicológica*, v. 2, n. 2, p. 99-110, 2003.

REISE, S. P.; AINSWORTH, A. T.; HAVILAND, M. G. Item Response Theory: fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, [s. l.], v. 14, n. 2, p. 95-101, Apr. 2005. <https://doi.org/10.1111/j.0963-7214.2005.00342.x>

REISE, S. P.; WALLER, N. G. Item Response Theory and clinical measurement. *Annual Review of Clinical Psychology*, Palo Alto, v. 5, p. 27-48, 2009. <https://doi.org/10.1146/annurev.clinpsy.032408.153553>

SARTES, L. M. A.; SOUZA-FORMIGONI, M. L. O. D. Avanços na psicometria: da teoria clássica dos testes à Teoria de Resposta ao Item. *Psicologia: Reflexão e Crítica*, Porto Alege, v. 26, n. 2, p. 241-250, 2013. <https://doi.org/10.1590/S0102-79722013000200004>

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, p. 83-112, jan./abr. 2005. <https://doi.org/10.1590/S0101-74382005000100006>

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*, Campinas, v. 15, n. 1, p. 159-176, 2010. <https://doi.org/10.1590/S1414-40772010000100009>

TEZZA, R.; BORNIA, A. C. Teoria da Resposta ao Item: vantagens e oportunidades para a engenharia de produção. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 29., Salvador, 2009. *A engenharia de produção e o desenvolvimento sustentável: integrando tecnologia e gestão*. Rio de Janeiro: Abepro, 2009. Disponível em: http://www.abepro.org.br/biblioteca/enegep2009_TN_STP_094_638_13156.pdf. Acesso em: out. 2018.

UNIVERSIDADE FEDERAL DE SANTA CATARINA. Comissão Própria de Avaliação. *Plano de avaliação institucional – 2017*. Florianópolis, 2017.

UNIVERSIDADE FEDERAL DE SANTA CATARINA. Comissão Própria de Avaliação. *Relatório integral de autoavaliação institucional de 2017*. Florianópolis, 2018.

VALLE, R. C. A construção e a interpretação de escalas de conhecimento: considerações gerais e uma visão do que vem sendo feito no SARESP. *Estudos em Avaliação Educacional*, Rio de Janeiro, n. 23, p. 71-92, 2001. <https://doi.org/10.18222/ea02320012208>



Informações sobre os autores

Anna Cecília Amaral Petrassi: Doutoranda no Programa de Pós-Graduação em Engenharia de Produção pela Universidade Federal de Santa Catarina. Técnico Administrativo em Educação da mesma Universidade. Contato: a.petrassi@ufsc.br.

 <http://orcid.org/0000-0003-0072-4837>

Antonio Cezar Bornia: Doutor em Engenharia de Produção pela Universidade Federal de Santa Catarina. Professor Titular do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina. Contato: cezar.bornia@ufsc.br.

 <http://orcid.org/0000-0003-3468-7536>

Dalton Francisco Andrade: Doutor em Biostatistics pela University of North Carolina System. Professor Titular do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina. Contato: dandrade@ufsc.br.

 <https://orcid.org/0000-0002-4403-980X>