



ANÁLISE DO INSTRUMENTO DE MEDIDA UTILIZADO NA AUTOAVALIAÇÃO INSTITUCIONAL DA UFSM APLICADO AOS SEUS DISCENTES, UTILIZANDO A TEORIA DA RESPOSTA AO ITEM (TRI)

ANALYSIS OF THE MEASUREMENT INSTRUMENT USED IN INSTITUTIONAL SELF-EVALUATION OF UFSM APPLIED TO ITS DISCENTS, USING THE ITEM RESPONSE THEORY (IRT)

Ricardo da Silva Barcelos, Universidade Federal de Santa Maria, ricardo.rsb@gmail.com;

Ivan Henrique Vey, Universidade Federal de Santa Maria, ivanvey@hotmail.com

RESUMO

A autoavaliação institucional é um importante componente do SINAES, constituindo-se em um instrumento para que as Instituições de Ensino Superior identifiquem potencialidades e fraquezas. Uma das atividades inerentes à autoavaliação institucional é a realização da avaliação docente. Sob essa perspectiva, o presente trabalho analisou o instrumento de medida da avaliação docente, aplicado aos discentes dos cursos de graduação da UFSM. O estudo se baseou na técnica estatística da Teoria da Resposta ao Item (TRI) para avaliar confiabilidade, consistência interna e unidimensionalidade do construto, e estimar os parâmetros de dificuldade e discriminação do item, a partir da aplicação do Modelo Logístico de 2 Parâmetros. Os testes estatísticos foram aplicados a uma base de dados contendo mais de 20.000 avaliações, realizadas no primeiro semestre de 2016. Os resultados atestaram a consistência interna e boa qualidade de construção do conjunto de itens. O teste de dimensionalidade demonstrou que o instrumento de medida é unidimensional, medindo o desempenho docente. A partir dos resultados obtidos, concluiu-se que o instrumento de medida aplicado na avaliação docente pelos discentes da instituição apresentou validade estatística quando submetido aos testes da TRI.

Palavras-chave: Avaliação de desempenho docente. Autoavaliação institucional. Teoria da Resposta ao Item.

ABSTRACT

Institutional self-assessment is an important component of the SINAES, which consists in an instrument for higher education institutions to identify their capacities and weaknesses. One of the activities inherent to institutional self-assessment is the performance evaluation of its professors. From this perspective, this study aimed to analyze the measuring instrument used on professor performance evaluation, applied to students of undergraduate courses at the UFSM. The study was based on the statistical technique of Item Response Theory (IRT) to evaluate the construct reliability, its internal consistency and unidimensionality, and to estimate the items difficulty and discrimination parameters, by using the 2-Parameter Logistic Model. The data treatment and statistical tests were applied to a database containing over 20,000 evaluations, carried out in the first half of 2016. The results showed that the set of items has internal consistency, so as good build quality. The dimensionality test showed that the measuring instrument is one-dimensional, measuring a single latent trait, the professor's performance. The results also demonstrated that the measuring instrument applied to



professor performance evaluation is statistically valid when subjected to the Item Response Theory analysis.

Keywords: Professor performance evaluation. Institutional self-assessment. Item Response Theory.



1. Introdução

A educação, tida como motor do crescimento e desenvolvimento das potências mundiais, constitui-se em um amplo campo de estudo para temas relacionados à inovação, inclusive no que diz respeito à gestão de organizações públicas. Na esfera Federal, o ano de 2007 foi de grande valia para a educação superior, uma vez que o Decreto nº 6.096/2007 instituiu o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais – REUNI. O programa teve por objetivo “criar condições para a ampliação do acesso e permanência na educação superior, no nível de graduação, pelo melhor aproveitamento da estrutura física e de recursos humanos existentes nas universidades federais” (BRASIL, 2007). A vigência do programa iniciou no ano de 2008 e se estendeu até o final de 2012. Apesar disso, a crise político-econômica vivenciada no país nos últimos dois anos, tem resultado em contingenciamento de despesas orçamentárias.

Nesse cenário e sob a perspectiva de que as organizações não são estanques, as instituições devem identificar novas e melhores formas de executar seus processos, e melhor atender seu público-alvo. A adoção de ferramentas de diagnóstico e de avaliação podem auxiliar o processo de tomada de decisão e promover melhorias de gestão, sendo componentes fundamentais para o desenvolvimento das organizações públicas. Dentre elas, no campo da educação superior, está a avaliação institucional, regulamentada pela Lei nº 10.861/2004, que instituiu o Sistema Nacional de Avaliação da Educação Superior (SINAES).

Como componente do SINAES, a autoavaliação institucional está relacionada a quatro aspectos fundamentais, em consonância com a referida Lei: (1) melhoria da qualidade da educação superior; (2) orientação da expansão de sua oferta; (3) aumento da eficácia institucional e efetividade acadêmica e social; e (4) aprofundamento dos compromissos e responsabilidades sociais das instituições (INEP, 2015).

A autoavaliação é operacionalizada internamente pelas Instituições de Ensino Superior (IES) através de grupos de trabalho criados para este fim específico, as Comissões Próprias de Avaliação (CPA). No ano de 2016, atendendo a exigências legais, a CPA da Universidade Federal de Santa Maria (UFSM), por meio das Comissões Setoriais de Avaliação, incluiu na avaliação institucional um instrumento de coleta de dados que permitiu aos alunos dos cursos de graduação avaliarem o trabalho realizado pelos docentes. A avaliação *in loco* por parte dos discentes pode se configurar em uma importante ferramenta de diagnóstico em relação às práticas pedagógicas (BARBOSA, FREIRE e CRISÓSTOMO, 2011).



Para viabilizar a avaliação docente, a Pró-Reitoria de Planejamento da UFSM, por meio da Coordenadoria de Planejamento e Avaliação Institucional (COPLAI), elaborou um instrumento de coleta de dados na forma de questionário, que foi aplicado durante o primeiro semestre letivo, especificamente no período entre 16 de junho e 16 de julho de 2016 em fase de teste. Este instrumento contou com 17 questões visando avaliar o desempenho docente, e foi respondido via Portal do Aluno, sendo disponibilizado a todos os alunos de cursos de graduação da Instituição. A aplicação em caráter oficial ocorreu entre os dias 14 de dezembro de 2016 e 6 de março de 2017, também via Portal do Aluno, não sendo obrigatória a participação dos discentes.

A consolidação de um instrumento de medida passa invariavelmente por algum tipo de teste de validação, seja validade de conteúdo, validade de critério ou outra técnica. Esses apontamentos demonstraram a necessidade de submeter o questionário a testes de validação para, inclusive, melhorar sua aceitação por parte do segmento docente. Nesse contexto, a Teoria da Resposta ao Item (TRI) se apresenta como uma técnica estatística que pode auxiliar a validação do construto, ao testar elementos como a confiabilidade, consistência interna, correlação e parâmetros dos itens que compõem o instrumento de medida aplicado na avaliação docente, a partir das respostas emitidas pelos avaliadores, no caso os discentes.

Utilizada inicialmente em testes na área de Psicologia (PASQUALI, 2013), a TRI foi difundida quando passou a ser aplicada em testes de conhecimentos e habilidades, e em avaliações de desempenho. Trata-se de uma ferramenta usada na medição de traços latentes (como inteligência, depressão e ansiedade, por exemplo), os quais não podem ser observados diretamente. Para determinação de como se processa um traço latente em um indivíduo, a TRI gera uma escala, formada a partir dos itens elaborados com a finalidade de montar um construto.

Considerando a avaliação como prática necessária ao desenvolvimento como é destacado no Projeto de Avaliação Institucional da Universidade Federal de Santa Maria, elaborado em junho de 2008, que preceitua que o tema deve ser encarado como exigência de controle de qualidade e deve possuir caráter permanente, este trabalho buscou responder ao seguinte problema de pesquisa: qual a confiabilidade e validade do instrumento de medida aplicado aos discentes na avaliação dos docentes da UFSM?

Os objetivos da pesquisa foram traçados a partir do detalhamento do problema de pesquisa, e levando em consideração seus fatores determinantes. Este trabalho teve por objetivo geral analisar o comportamento do conjunto de itens utilizado no instrumento de



medida da avaliação docente pelos discentes de graduação da UFSM, utilizando a Teoria da Resposta ao Item.

Os objetivos específicos surgiram a partir de desdobramentos do objetivo geral, com vistas a responder ao problema de pesquisa. No caso do presente trabalho, os objetivos específicos delimitados foram:

- Verificar a consistência interna, confiabilidade e dimensionalidade do instrumento de medida aplicado aos discentes;
- Realizar a análise fatorial de informação plena para verificar a unidimensionalidade do instrumento de medida;
- Analisar a correlação bisserial dos itens que compuseram o instrumento de medida;
- Verificar o comportamento dos parâmetros de discriminação (parâmetro a) e dificuldade (parâmetro b) dos itens para a Instituição;
- Analisar o comportamento dos parâmetros a e b nos Centros de Ensino, comparando-o com o resultado geral;
- Propor melhorias para o instrumento de avaliação, caso sejam verificadas inconsistências.

A motivação para o presente trabalho envolveu questões de ordem teórica e prática. De ordem teórica, a relevância do estudo envolve a avaliação do desempenho docente na percepção dos discentes, uma vez que estes podem ser considerados o público-alvo da atividade de docência. Portanto, é fundamental que existam mecanismos de avaliação que permitam a aferição do desempenho docente, e que estes mecanismos sejam testados, validados e adequados para que o desempenho seja devidamente medido, um dos objetivos deste trabalho.

Como relevância prática, espera-se que o trabalho contribua para a difusão do tema e para a realização de novas pesquisas e, principalmente, que os resultados aqui obtidos, assim como os resultados da avaliação sejam efetivamente utilizados pelos gestores da Instituição na busca constante pela excelência do ensino. Espera-se ainda que o estudo contribua para o aperfeiçoamento do processo de avaliação, ao buscar a validação do instrumento de medida através da Teoria da Resposta ao Item.

2. Referencial teórico

Os dois temas basilares para a fundamentação teórica deste trabalho foram a avaliação de desempenho docente e a Teoria da Resposta ao Item. Em relação à avaliação, buscou-se



construir o referencial contextualizando a avaliação de desempenho e o Sistema Nacional de Avaliação da Educação Superior.

2.1 Avaliação docente no contexto do SINAES

Os dezesseis artigos da Lei do SINAES norteiam os princípios e fatores que devem ser considerados na promoção da avaliação da educação superior, em sua maioria em linhas gerais. A avaliação de desempenho docente não é tratada diretamente na referida lei, porém, subentende-se sua necessidade, tendo em vista que o docente é o principal componente na difusão do conhecimento e de posturas críticas e reflexivas.

Alguns indícios disso são encontrados, por exemplo, no inciso V do art. 3º da Lei nº 10.861/2004, que trata das políticas de pessoal e das carreiras do corpo docente e técnico-administrativo, com vistas a propiciar seu aperfeiçoamento, desenvolvimento profissional e condições de trabalho (BRASIL, 2004). Reforçando este posicionamento, merece transcrição na íntegra o art. 4º da referida lei:

Art. 4º – A avaliação dos cursos de graduação tem por objetivo identificar as condições de ensino oferecidas aos estudantes, em especial as relativas ao perfil do corpo docente, às instalações físicas e à organização didático-pedagógica (BRASIL, 2004).

A partir desse trecho, pode-se inferir que ao tratar das condições de ensino, em especial relativas ao perfil do corpo docente, tem-se um primeiro vislumbre da importância da avaliação deste grupo. Nesse sentido, Rodrigues (2012) em trabalho intitulado “Políticas de Avaliação Docente: tendências e estratégias”, elenca algumas das finalidades da avaliação docente. Dentre elas estão o objetivo de melhorar o desempenho dos professores, assim como as práticas e procedimentos, e também como forma de prestação pública de contas, ou seja, transparência. Apesar de o trabalho da autora ter seu foco voltado para a educação básica, os pressupostos citados anteriormente também se aplicam à educação superior.

No contexto da avaliação docente, Castanheira e Ceroni (2007) trazem uma reflexão que reforça a importância do assunto, ao afirmarem:

Neste contexto, evidencia-se a necessidade de os educadores considerarem a avaliação importante meio de diagnóstico de seu trabalho, bem como a elaboração de instrumentos mais diretos de aferição da qualidade de ensino e aprendizagem do aluno que quer se formar (CASTANHEIRA e CERONI, 2007, p. 722).

A afirmação das autoras evidencia a necessidade de adoção de formas alternativas de avaliação docente. O desenvolvimento de ferramentas de avaliação mais efetivas, que



permitam ao docente um *feedback* de sua atuação nas atividades de ensino, pesquisa e extensão, vai ao encontro das dimensões do SINAES e de suas finalidades, principalmente a de propiciar a melhoria da qualidade da educação superior, como corrobora o posicionamento de Embiruçu, Fontes e Almeida (2010):

A avaliação do desempenho docente em Instituições de Ensino Superior (IES), de forma similar à avaliação do desempenho de qualquer profissional nos diversos campos de atuação e nos diversos sistemas e processos produtivos, é de fundamental importância para o desenvolvimento profissional dos docentes e para o crescimento institucional, ajudando a garantir que as metas qualitativas e quantitativas esperadas pela sociedade possam ser alcançadas (EMBIRUÇU et al., 2010, p. 797).

Reforçando a importância da avaliação para melhoria da qualidade do ensino, em especial a avaliação docente por parte do discente, Boclin (2004) destaca que “no processo de ensino tudo acontece na razão direta das relações entre professores e alunos e cujo palco e cenário vem a ser as salas de aulas, laboratórios, bibliotecas...” (BOCLIN, 2004, p. 959). Ainda conforme o autor, o melhor entendimento destas relações e os efeitos sobre a qualidade do ensino e da aprendizagem, é ao mesmo tempo fundamental e desafiador para a continuidade das ações educativas, que se encontram em constante transformação. Seguindo o mesmo raciocínio, Sousa e Andriola (2014) apregoam que a avaliação docente deve interagir fortemente com a gestão de recursos humanos, principalmente em razão da lógica de formação continuada, uma vez que somente a concessão de oportunidades para que o professor possa se qualificar e adquirir novas competências poderá contribuir para a melhoria da aprendizagem dos alunos.

Dentre os benefícios que podem advir do processo, além de identificar o perfil do docente, bem como os aspectos didáticos ou acadêmicos nos quais ele deva se aperfeiçoar, Embiruçu et al. (2010) apontam: fundamentação da avaliação em estágios probatórios, avaliação de desempenho de departamentos e Unidades Universitárias; e avaliação de desempenho da instituição como um todo. Nesse sentido, entende-se que a avaliação de desempenho docente impacta toda a instituição, dada sua importância. Castanheira e Ceroni (2007) destacam ainda que o levantamento da opinião discente em relação às atividades didático-pedagógicas dos professores é um dos componentes do processo de avaliação institucional. Destacam, porém, se tratar de tarefa complexa, “pois ao professor cabe, por vezes, não satisfazer seus alunos e suas expectativas imediatistas” (CASTANHEIRA e CERONI, 2007, p. 724). Tem-se aqui um indício de que a construção de um instrumento adequado deve atentar para possíveis distorções que podem ocorrer, principalmente



decorrente de respondentes que interpretem a avaliação docente como forma de retaliação a determinado professor.

2.2 A Teoria da Resposta ao Item

A TRI é um conjunto de modelos matemáticos que busca representar a probabilidade de um indivíduo dar uma determinada resposta a um item como função dos parâmetros do item e da sua habilidade (ou habilidades). Em geral, a maioria dos modelos é expressa de modo acumulativo. Porém, para determinados tipos de traços latentes (atitude e comportamentos), os modelos de desdobramento são mais apropriados (ANDRADE, TAVARES e VALLE, 2000).

A proposta da TRI é a de apresentar modelos probabilísticos para variáveis que não são medidas diretamente, tendo como característica principal o item, podendo se estender por item, tarefas ou ações empíricas que constituem a representação do traço latente, ou seja, a habilidade que se pretende medir. No caso em estudo o traço latente que se procurou mensurar foi o desempenho dos docentes da UFSM na percepção dos discentes. O modelo utilizado para análise com a TRI foi o modelo logístico de dois parâmetros (ML2).

Birnbaum (1968) desenvolveu a equação que serve para avaliar dois parâmetros do item: dificuldade e discriminação. Assim, o modelo logístico de dois parâmetros utiliza, além do parâmetro de dificuldade do item, o parâmetro de discriminação do item. Desta forma, considera-se que U_{ij} seja uma variável aleatória assumindo valores 0 ou 1, em que o valor 0 é associado a uma resposta errada, e o valor de 1, a uma resposta correta por parte do indivíduo. O modelo de 2 parâmetros expressa a relação entre a variável latente θ e a resposta ao item como pode ser verificado na Equação 1.

$$P\left(U_{ij} = \frac{1}{\theta_j}\right) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (\text{Equação 1})$$

Desta forma, a_i é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional à inclinação da curva característica do item (CCI) no ponto b_i . D é um fator de escala constante e igual a 1. Utiliza-se o valor 1,7 quando se deseja que a função logística forneça resultados semelhantes ao da função ogiva normal, diz-se, então, que o modelo está na métrica normal. O índice i representa o número do item e j o respondente.

Em outras palavras, o modelo representa a probabilidade de um indivíduo “j” dar uma resposta certa a um item “i” de um instrumento de medição (conjunto de itens ou testes) em função dos parâmetros dos itens (a_i , b_i) e do conhecimento θ_j (habilidade, proficiência). Em suma, quanto maior o traço latente, maior a probabilidade de resposta correta.

Uma das características da TRI enquanto técnica estatística é que o parâmetro de dificuldade do item (b_i) se encontra na mesma escala do traço latente do indivíduo (θ_i), permitindo que “a dificuldade dos itens possa ser interpretada em termos de variações padronizadas na habilidade dos sujeitos” (GARCIA et al., 2001 apud COUTO e PRIMI, 2011).

A Figura 1 apresenta a representação gráfica da curva característica do item (CCI) em um modelo logístico de dois parâmetros com dois itens hipotéticos, e na sequência a interpretação do gráfico.

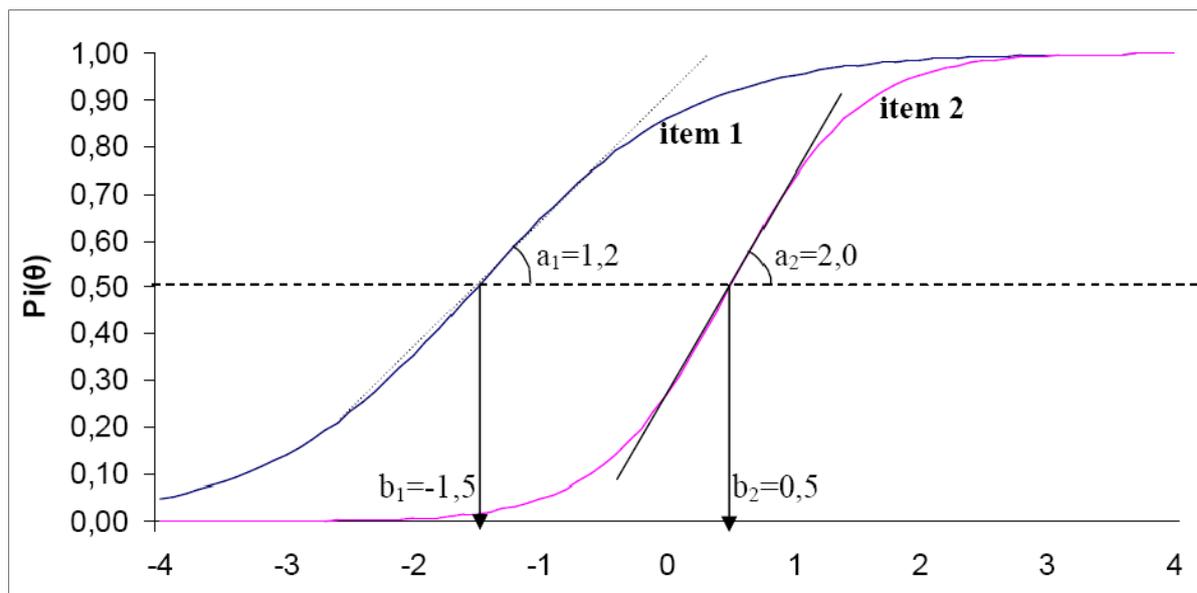


Figura 1 – Curva característica do item. Adaptado de Pasquali e Primi (2003).

A Figura 1 mostra o exemplo de uma CCI (Curva Característica do Item) de dois itens hipotéticos (item 1 e 2) e a identificação dos parâmetros dos itens. Utilizando o exemplo e analisando sob o ponto de vista do presente estudo, o eixo x representa a dificuldade dos itens e está na mesma escala de avaliação do desempenho dos docentes, ou seja, quanto maior a dificuldade maior o grau de aprovação do desempenho (θ) requerido para aquele item. Sendo assim, o parâmetro b_i , que representa matematicamente a dificuldade do item i , é definido como o valor onde a probabilidade de resposta ao item é de 0,5.

O parâmetro a_i , que representa a discriminação do item, é proporcional à derivada da tangente da curva no ponto de inflexão e indica a inclinação da curva, ou seja, quanto maior for o valor de a_i , mais inclinada vai ser a curva e, conseqüentemente mais estreito será o intervalo de discriminação do item i . Quanto à escala, normalmente supõem-se que a escala assuma uma métrica normal (0, 1), ou seja, média 0 (zero) e desvio padrão 1 (um). Por exemplo, alunos posicionados na escala (eixo x) mais à direita (1, 2, 3...) são os que melhor avaliam o desempenho dos docentes em estudo, ou seja, possuem um traço latente maior que os outros que se situam abaixo na escala. Quanto maior o θ (*theta*), maior a avaliação (percepção) positiva do desempenho do item pelo respondente (discente).

Por outro lado, também se pode fazer a seguinte análise: itens que possuem um b_i maior serão itens mais difíceis, ou seja, itens do desempenho docente os quais necessitam de uma percepção de desempenho alta (θ), por parte do respondente (discente), para que avaliem positivamente seu desempenho. Itens que possuem um b_i menor serão itens de fácil aprovação, ou seja, itens que o respondente não necessita de uma elevada avaliação de desempenho (θ) para que avalie positivamente o mesmo.

Desta forma, no exemplo anterior, o item 2 é mais difícil que o item 1, ou seja, apenas discentes com alta avaliação (percepção) de desempenho docente (θ) irão avaliar o item positivamente, ou seja, aprovar o seu desempenho. Pode-se dizer que alunos posicionados mais à direita na escala serão aqueles que estarão mais satisfeitos com o desempenho docente em estudo, ou seja, consideraram o desempenho docente elevado, aprovam seu desempenho.

3. Método de pesquisa

Foram analisados os 13 primeiros itens do instrumento de medida aplicado, os quais estão diretamente ligados ao desempenho docente, pois os itens restantes (14, 15, 16 e 17) tratam de questões peculiares e individuais, relativas à docência orientada (14 a 16), e ao comprometimento do aluno com as atividades propostas pelo professor (17).

Inicialmente os dados foram tabulados e analisados utilizando a Teoria Clássica dos Testes (TCT) utilizando os programas MS-Excel e SPSS. Posteriormente os dados foram dicotomizados e analisados com base na Teoria da Resposta ao Item (TRI) utilizando os programas BILOG-MG e TESTFACT, específicos para a TRI. A avaliação ocorreu no primeiro semestre do ano de 2016. Foram avaliados todos os docentes pertencentes ao quadro da UFSM. Considerando os dados tabulados no MS-Excel, em que cada linha preenchida



representa a avaliação dada por um acadêmico ao professor de uma disciplina, foram respondidos mais de 23.000 questionários somente no campus sede.

Na sequência são apresentados os resultados e análises.

4. Apresentação e discussão dos resultados

Para a análise dos dados, adotou-se inicialmente a Teoria Clássica dos Testes e, a seguir, a Teoria da Resposta ao Item. A análise integrada a partir das duas correntes visa reforçar e aumentar a consistência dos resultados, tendo em vista o caráter complementar que uma teoria exerce sobre a outra.

4.1 Análise com base na Teoria Clássica dos Testes (TCT)

Uma das propriedades na Teoria Clássica dos Testes (TCT) utilizada para avaliar a qualidade do instrumento de medida é dada pela fidedignidade. A consistência interna do conjunto de itens que medem o desempenho docente pode ser verificada através da determinação do alfa (α) de Cronbach. Para Nunnally (1978), um valor superior a 0,75 é considerado satisfatório. Nesse sentido, Pasquali (2013) coloca que quando o resultado do coeficiente se aproxima de 1 (um), pode-se afirmar que o teste possui um coeficiente de precisão.

Na Tabela 1 são apresentados os dados descritivos da amostra em que este estudo se baseou. Após, são apresentados os resultados do teste alfa de Cronbach.

Tabela 1
Número de respondentes

	N	%
Válidos	20543	86,1
Excluídos	3328	13,9
Total	23871	100,0

Nota. Fonte: elaborado pelos autores com resultados extraídos do *software SPSS*.

Foram analisadas 23.871 respostas aos itens, destas, 20.543 foram consideradas na análise do alfa de Cronbach. Apesar de o número total de casos considerados no teste ser de 23.871, dezessete deles foram excluídos nas demais análises em razão da atual lotação de exercício dos docentes avaliados ser a Reitoria Universitária. Isto impossibilitou que, no banco de dados, fosse identificado o departamento vinculado à Unidade Universitária a qual o

docente está oficialmente lotado. Na sequência são apresentadas as estatísticas individuais de cada item e o resultado final do teste.

Tabela 2
Alfa de Cronbach dos itens

Item	Alfa de Cronbach
01	0,940
02	0,943
03	0,936
04	0,939
05	0,938
06	0,938
07	0,936
08	0,938
09	0,937
10	0,940
11	0,939
12	0,942
13	0,939

Nota. Fonte: elaborado pelos autores com resultados extraídos do *software SPSS*.

Tabela 3
Alfa de Cronbach para o conjunto de itens

Alfa de Cronbach	Alfa de Cronbach com base em itens padronizados	N de itens
0,943	0,944	13

Nota. Fonte: elaborado pelos autores com resultados extraídos do *software SPSS*.

Considerando que o resultado final do alfa de Cronbach para o conjunto de itens foi de 0,94, pode-se afirmar que o instrumento é confiável e possui consistência interna.

Outro procedimento adotado neste estudo foi a utilização da análise fatorial para testar a dimensionalidade do conjunto de itens. A proposta primordial da análise fatorial é descrever, se possível, a estrutura de covariâncias entre as variáveis em termos de um número menor de variáveis (não observáveis) chamadas fatores. Por outras palavras, a análise fatorial estuda os inter-relacionamentos entre as variáveis, num esforço para encontrar um conjunto de fatores (em menor nº que o conjunto de variáveis originais) que exprima o que as variáveis originais partilham em comum. Assim, para testar a dimensionalidade, Spencer (2004) sugere a utilização da análise fatorial que, para o autor, é um conjunto de ferramentas estatísticas utilizadas para reduzir o número de variáveis, assim como para avaliar a estrutura de dados, podendo ser exploratória ou confirmatória. Mello (2014) complementa este entendimento ao

afirmar que a técnica se propõe a analisar a estruturas das inter-relações entre um grande número de variáveis, ao passo que define um conjunto de dimensões latentes – os fatores. Na análise utilizou-se do método de componentes principais como método de estimação. Na sequência são apresentados os resultados extraídos com a utilização do *software* SPSS.

Tabela 4
Variância total explicada

Componente	Valores próprios iniciais			Somadas de extração de carregamentos ao quadrado		
	Total	% de variância	% cumulativa	Total	% de variância	% cumulativa
1	7,813	60,102	60,102	7,813	60,102	60,102
2	,865	6,654	66,756			
3	,682	5,246	72,003			
4	,580	4,461	76,464			
5	,524	4,034	80,498			
6	,459	3,528	84,026			
7	,396	3,047	87,073			
8	,365	2,805	89,877			
9	,317	2,438	92,315			
10	,279	2,148	94,463			
11	,262	2,019	96,482			
12	,239	1,836	98,318			
13	,219	1,682	100,000			

Nota. Fonte: elaborado pelos autores com resultados extraídos do *software* SPSS.

Com base nos resultados apresentados pode-se afirmar que existe um fator (variável) determinante responsável por 60,10% da variabilidade dos dados. Na concepção de Hair et al. (2005), é comum em ciências sociais que seja considerada satisfatória uma solução que explique 60% da variância total. Nesse sentido, estes valores atestam que o instrumento de medida é unidimensional, ou seja, está medindo um único traço latente, o desempenho dos docentes. Assim, o instrumento de medida é unidimensional, ou seja, está medindo um único traço latente, o desempenho dos docentes.

Assim, com base nos resultados estatísticos desta primeira análise a qual testou a confiabilidade e dimensionalidade do instrumento de medida aplicado junto aos discentes, pode-se concluir que o instrumento de medida possui confiabilidade e consistência interna. Por outro lado, o conjunto de itens que compõe o instrumento de medida estão medindo um único traço latente, o desempenho docente, ou seja, o instrumento de medida é unidimensional.

4.2 Análise baseada na Teoria da Resposta ao Item (TRI)

Nesta parte da análise o estudo utilizou-se da Teoria da Resposta ao Item. Para tanto foi necessário realizar a dicotomização das respostas. O instrumento aplicado utilizou-se de uma escala Likert de 5 pontos.

Tabela 5
Escala de concordância

1	2	3	4	5
Discordo totalmente	Discordo em parte	Não concordo nem discordo	Concordo em parte	Concordo totalmente

Nota. Fonte: CPA/UFSM.

Após análise da média e mediana dos itens decidiu-se cortar todos os itens no 3, ou seja, a discentes que responderam 1, 2 ou 3 foram considerados “não aprova o desempenho do docente” naquele item, discentes que responderam 4 ou 5 foram considerados “aprova o desempenho do docente”.

Com os dados dicotomizados os itens novamente foram testados para verificar a dimensionalidade do instrumento de medida. A análise da dimensionalidade foi realizada através da análise fatorial de informação plena utilizando-se o *software* TESTFACT.

O resultado obtido apresentou o primeiro fator responsável por explicar 63,70% da variabilidade geral dos dados e o segundo fator 4,52%. Assim, a análise comprova a suposição de que o conjunto de itens é unidimensional, ou seja, está medindo um único traço latente, o desempenho docente. Isso quer dizer que o resultado justifica a utilização de um modelo unidimensional da TRI, no caso, o modelo logístico de dois parâmetros.

Na Tabela 6 são apresentados os dados da correlação bisserial. No entendimento de Soares e Mendonça (2003), a correlação bisserial é uma medida estatística da capacidade de discriminação do item, na medida em que verifica se um determinado item binário apresenta correlação significativa com o escore bruto resultante do conjunto de itens.

Tabela 6
Correlação bisserial do conjunto de itens.

Item	Correlação de Pearson	Correlação bisserial
01	0,5750	0,9420
02	0,5040	0,8330
03	0,7020	1,0160
04	0,5870	0,8350
05	0,6560	1,0000



06	0,6460	0,9280
07	0,7300	1,0750
08	0,6690	1,0630
09	0,6950	0,9810
10	0,5960	0,9780
11	0,6290	0,9760
12	0,5800	0,7860
13	0,6340	1,0290

Nota. Fonte: elaborado pelos autores com resultados extraídos do *software Bilog-MG*.

A partir da análise dos dados da correlação bisserial, pode-se afirmar que há forte correlação entre cada item e o conjunto como um todo, tendo em vista que é usual a aceitação de itens que apresentem valores superiores a 0,3 para este parâmetro (SOARES, 2005). Considerando que o item que apresentou menor correlação (item 12) atingiu 0,7860, tem-se um primeiro indício de que o conjunto de itens do instrumento de coleta de dados foi bem construído, pois apresentam consistência interna e correlação com o escore bruto.

Na Figura 2 são apresentadas as curvas características dos itens. Estas são analisadas em conjunto com a Tabela 7, em que constam os valores dos parâmetros a e b – discriminação e dificuldade, respectivamente – de cada item integrante do teste. Na Tabela 7 também estão contidas as assertivas associadas a cada um dos itens do teste. A estimação dos parâmetros a e b considera que os dados seguem uma distribuição normal, com média igual a zero e desvio padrão igual a um ($\mu=0$ e $\sigma=1$). Conforme atesta Vey (2011), esta etapa dos testes estatísticos é conhecida na TRI como calibração dos itens. Para tanto, utilizou-se o método de Máxima Verossimilhança Marginal (MVM) na estimação dos parâmetros. Ademais, a convergência dos dados também foi testada pelo algoritmo *Expectation – Maximization* (EM) e Newton-Raphson (NR), no intuito de garantir a boa estimação dos parâmetros (MELLO, 2014).

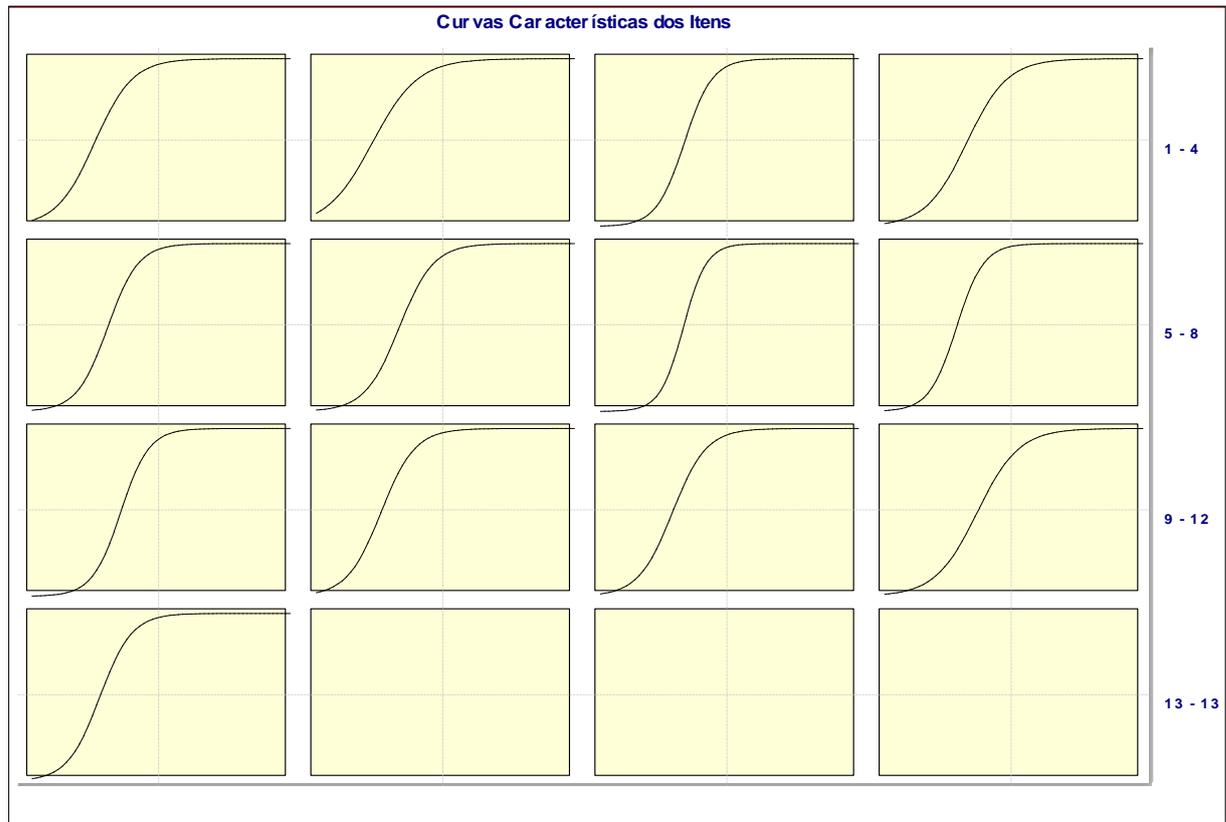


Figura 2 – Curvas características dos itens (CCI). Elaborado pelos autores.

Tabela 7

Parâmetros *a* (discriminação) e *b* (dificuldade) dos itens

Item	Assertiva	Parâmetro <i>a</i>	Parâmetro <i>b</i>
01	<i>O programa da disciplina foi apresentado pelo professor (objetivos, conteúdo a ser desenvolvido e bibliografia)</i>	2,257	-1,552
02	<i>O professor comparece às aulas e cumpre os horários de início e de término das mesmas</i>	1,881	-1,698
03	<i>O professor utiliza de forma adequada os recursos didáticos disponíveis, favorecendo a aprendizagem</i>	3,072	-1,053
Item	Assertiva	Parâmetro <i>a</i>	Parâmetro <i>b</i>
04	<i>O professor estimula a utilização de materiais complementares (livros, sites, periódicos online, áudio, vídeos, entre outros)</i>	2,107	-1,094
05	<i>O professor demonstra domínio sobre o conteúdo apresentado, tratando-o com clareza e objetividade</i>	2,812	-1,249
06	<i>O professor estabelece relações entre os conteúdos da sua disciplina com os conteúdos das demais disciplinas, contribuindo com a formação profissional</i>	2,509	-1,075
07	<i>O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes</i>	3,723	-1,077
08	<i>O professor mostra-se disponível para tirar dúvidas sobre a disciplina e o conteúdo ministrado</i>	3,177	-1,333
09	<i>O professor ouve críticas, opiniões e sugestões referentes às suas aulas, mostrando-se aberto ao diálogo</i>	3,045	-0,945
10	<i>O professor apresenta uma postura de respeito mútuo, ou seja, preserva a imagem da Instituição, dos colegas, dos acadêmicos e não manifesta atitudes preconceituosas</i>	2,532	-1,499
11	<i>O professor elabora avaliações compatíveis com os conteúdos ministrados em aula</i>	2,537	-1,330

12	<i>Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula</i>	2,049	-0,850
13	<i>O professor cumpre o programa da disciplina apresentado</i>	2,680	-1,433

Nota. Fonte: elaborado pelos autores com resultados extraídos do *software SPSS*.

Em uma análise inicial, para fins de interpretação do parâmetro a , considera-se que itens com valores de $a < 1$ indicam que o item apresenta pouco poder de discriminação (VEY, 2011). Nas CCI, isto pode ser verificado a partir da inclinação obtida no ponto médio da curva em relação ao eixo x , em que os itens com maior poder de discriminação apresentam curvas mais íngremes, enquanto os demais apresentam curvas com ângulo de inclinação menor. Para o estudo em questão, o item 07 (“*O professor demonstra satisfação em ensinar e interesse pelo aprendizado dos estudantes*”) foi o que resultou em melhor poder de discriminação – 3,723. Em contrapartida, o item 02 (“*O professor comparece às aulas e cumpre os horários de início e de término das mesmas*”) apresentou o valor de 1,881 para o referido parâmetro. Apesar disso, considerando que todos os itens apresentaram valor superior a um para o parâmetro a , de maneira geral o conjunto de itens mostra-se bastante equilibrado.

Para interpretação do parâmetro b , não há referência a valores absolutos. Considera-se que, quanto maior o valor de b na escala, maior será o grau de dificuldade para aprovação do desempenho docente na percepção dos discentes. Com base nos dados da tabela, o item 12 (“*Após as avaliações, o professor apresenta os itens avaliados, esclarecendo a nota atribuída e discutindo as questões em sala de aula*”) foi o que apresentou maior grau de dificuldade para ser bem avaliado. Sendo assim, este aspecto poderá ser trabalhado para que o desempenho seja melhorado em futuras avaliações. Por outro lado, o item 02 foi o que apresentou menor valor no parâmetro b , indicando que a percepção dos discentes para melhoria neste aspecto possui grau de dificuldade menor em relação aos demais itens do conjunto.

A Figura 3 representa a curva de informação total do teste. O gráfico apresenta o somatório da informação de todos os itens que compõem o teste, e é obtido a partir da função de informação do teste (FIT), demonstrando para qual região da escala o construto é mais indicado (VEY, 2011), se para avaliar docentes com desempenho baixo, médio ou elevado.

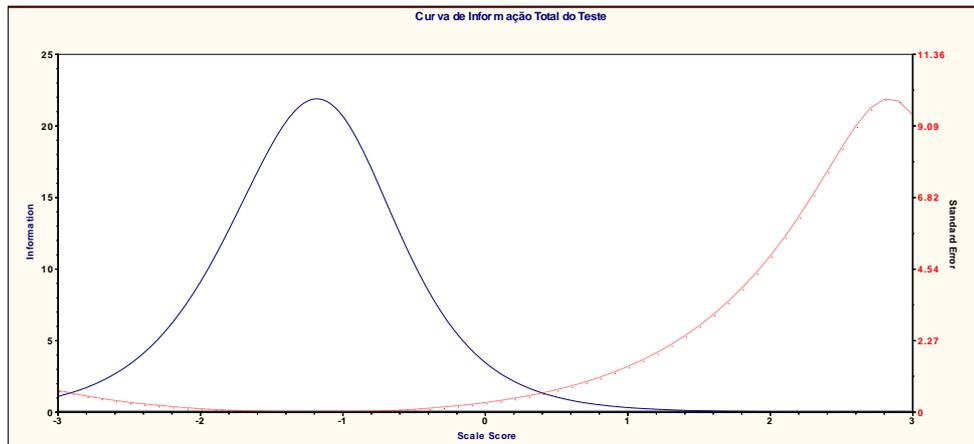


Figura 3 – Curva de informação total do teste. Elaborado pelos autores.

No gráfico são apresentadas a função de informação do teste e o erro padrão de medida (EPM) para o conjunto de itens. A FIT é representada pela linha contígua, devendo ser analisada pela escala vertical situada à esquerda do gráfico. A linha pontilhada representa o EPM, e sua análise é feita a partir das informações constantes na escala vertical à direita.

Em relação à FIT, verificou-se que a maior quantidade de informação – acima de 10 – para os níveis de desempenho docente situa-se no intervalo de -1,9 a -0,5. Em relação ao EPM (linha pontilhada), interpreta-se que quanto maior a quantidade de informação, menor será o erro padrão e maior a precisão. No caso, o menor erro padrão também se situou no intervalo entre -1,9 e -0,5. O EPM aumenta para valores mais elevados na escala de desempenho docente, atingindo valor superior a 9 no extremo direito, para $\theta = 2,8$. Já no extremo esquerdo, o EPM atingiu valor superior a 1,5 para $\theta = -3$. Considerando os resultados obtidos, verificou-se que o conjunto de itens do construto é mais adequado para medir itens entre os níveis -2,7 e 0,2 na escala (0, 1). Além disso, o formato da curva também é um indicador de validade do construto. Tendo em vista sua forma próxima a de uma distribuição normal, pode-se reforçar a validade do conjunto de itens.

5. Considerações finais

A autoavaliação institucional é fundamental para as instituições de ensino superior identificarem suas potencialidades e fraquezas, e dessa forma tomarem melhores decisões acerca de seus rumos. Dentro da autoavaliação está incluída a avaliação de desempenho docente, processo este que foi retomado no ano de 2016 pela UFSM através da Comissão de Planejamento e Avaliação Institucional, vinculada à Pró-Reitoria de Planejamento.



O presente estudo teve por objetivo analisar o instrumento de medida aplicado aos discentes dos cursos de graduação na avaliação do desempenho docente, utilizando os pressupostos da Teoria da Resposta ao Item. Buscou-se, através da realização de testes estatísticos e da aplicação dos modelos matemáticos, verificar a consistência interna do conjunto de itens que compõem o teste, a correlação bisserial, bem como proceder à análise dos parâmetros a e b (discriminação e dificuldade, respectivamente). Também foi realizada uma breve análise à luz da Teoria Clássica dos Testes.

Os resultados demonstraram de forma geral que o instrumento de medida aplicado aos discentes atinge sua finalidade, qual seja, medir o desempenho docente em sala de aula. Isto é atestado inicialmente pelo resultado do alfa de Cronbach, em que o valor atingido pelo conjunto de itens foi superior a 0,94, certificando a consistência e a confiabilidade do teste.

A unidimensionalidade do conjunto de itens foi atestada a partir da análise fatorial, com um fator respondendo por 55,21% da variabilidade dos resultados. Deste resultado depreende-se que apenas um traço latente está sendo medido, no caso, o desempenho docente. Os testes realizados a partir das técnicas da TRI, após a dicotomização do conjunto de itens, permitiram aferir a correlação bisserial dos itens, ao comparar a correlação dos itens em relação ao escore bruto. O fato de todos os itens apresentarem correlação superior a 0,78, além de evitar que algum item seja eventualmente excluído da análise, indica que o instrumento de medida possui boa consistência interna.

Por fim, a análise do parâmetro a revelou que o conjunto de itens é consistente, pois todos apresentaram capacidade de discriminação superior a um. Da mesma forma, o parâmetro de dificuldade (b) variou de -1,698 a -0,850. O fato de todos os itens apresentarem parâmetro b com sinal negativo indica que os itens são relativamente fáceis, ou seja, não há dificuldade do avaliador, no caso o discente, em concordar com o desempenho do docente.

Com base nos resultados obtidos, e a partir da metodologia utilizada, pode-se concluir que o instrumento de medida utilizado na avaliação docente por parte dos discentes atende aos objetivos a que se propõe. Deve-se considerar, porém, que novos testes e análises devem ser feitas para que tal instrumento seja constantemente aperfeiçoado, tendo em vista a qualificação do processo de avaliação de desempenho. Também, espera-se que a utilização destes dados, bem como dos resultados da avaliação se constituam em importante *feedback* para que os docentes possam aprimorar sua prática e atuação em sala de aula.

REFERÊNCIAS BIBLIOGRÁFICAS



- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. *Teoria de resposta ao item: conceitos e aplicações*. ABE — Associação Brasileira de Estatística, 4º SINAPE, 2000.
- BARBOSA, G. C.; FREIRE, F. S.; CRISÓSTOMO, V. L. Análise dos indicadores de gestão das IFES e o desempenho discente no ENADE. *Avaliação*, Campinas; Sorocaba, SP, v. 16, n. 2, p. 317-344, jul. 2011.
- BIRNBAUM, A. *Some latent trait models and their use in Inferring an examinee's ability*. In F. M. Lord and M. R. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley, 1968.
- BOCLIN, R.. Avaliação de docentes do ensino superior: um estudo de caso. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 12, n. 45, p. 959-980, out.-dez. 2004.
- BRASIL. Decreto n. 6.096 de 24 de abril de 2007. Institui o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais – REUNI. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/decreto/d6096.htm>. Acesso em: 15 mai. 2016.
- _____. Lei n. 10.861 de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior – SINAES e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Lei/L10.861.htm>. Acesso em: 15 set. 2015.
- CASTANHEIRA, A. M.; CERONI, M. R. Reflexões sobre o processo de avaliar docente contribuindo com sua formação. *Avaliação*, Campinas; Sorocaba, SP, v.12, n. 4, p. 719-737, dez. 2007.
- COUTO, G.; PRIMI, R. Teoria de Resposta ao Item (TRI): Conceitos elementares dos modelos para itens dicotômicos. *Boletim de Psicologia*, São Paulo, v. 61, n. 134, p. 1-15, 2011.
- EMBIRUÇU, M.; FONTES, C.; ALMEIDA, L. Um indicador para a avaliação do desempenho docente em Instituições de Ensino Superior. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 18, n. 69, p. 795-820, out.-dez. 2010.
- HAIR JUNIOR, J. F. et al. *Análise multivariada de dados*. 5. ed. Porto Alegre: Bookman, 2005.
- INEP. *Avaliação institucional*. Brasília, DF, 2015. Disponível em: <http://portal.inep.gov.br/superior-avaliacao_institucional>. Acesso em: 14 nov. 2015.
- MELLO, L. A. *Validação de um instrumento de avaliação de desempenho de coordenadores de curso pela Teoria da Resposta ao Item*. Santa Maria; 2014. Dissertação (Mestrado em Gestão Pública) – Universidade Federal de Santa Maria.
- NUNNALLY, J. C. *Psychometric theory*. 2. ed. New York: McGraw-Hill, 1978.



- PASQUALI, L. *Psicometria: teoria dos testes na Psicologia e na Educaçao*. 5. ed. Petrópolis: Vozes, 2013.
- PASQUALI, L.; PRIMI, R. *Fundamentos da teoria da resposta ao item*. *Avaliaçao Psicologica*, v. 2, n. 2, p. 99-110, 2003.
- RODRIGUES, S. S. Políticas de avaliaçao docente: tendências e estratégias. *Ensaio: Avaliaçao e Políticas Públicas em Educaçao*, Rio de Janeiro, RJ, v. 20, n. 77, p. 749-768, out.-dez. 2012.
- SOARES, T. M. Utilizaçao da teoria da resposta ao item na produçao de indicadores socioeconômicos. *Pesquisa Operacional*, Rio de Janeiro, v.25, n. 1, 2005.
- SOARES, T. M.; MENDONÇA, M. C. M. Construçao de um modelo de regressao hierárquico para os dados do SIMAVE-2000. *Pesquisa Operacional*, Rio de Janeiro, v. 23, n. 3, 2003.
- SPENCER, S. G. *The strength of multidimensional item response theory in exploring construct space that is multidimensional and correlated*. A dissertation submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, November 19, 2004.
- SOUSA, A. C. G.; ANDRIOLA, W. B. Avaliaçao docente na educaçao superior: expectativas acadêmicas acerca da influênci do feedback e das decisões institucionais. *Revista Eletrônica Acta Sapientia*, Fortaleza, CE, v.1, n. 1, 2014. Disponível em: <<http://www.poeduc.ufc.br/revista/index.php/actasap/article/view/2>>. Acesso em: 01 set. 2016.
- VEY, I. H. *Avaliaçao de desempenho logístico no serviço ao cliente baseada na Teoria da Resposta ao Item*. Florianópolis; 2011. Tese (Doutorado em Engenharia de Produçao) – Universidade Federal de Santa Catarina.