

ANÁLISE DO INSTRUMENTO DE MEDIDA UTILIZADO NA AVALIAÇÃO DOCENTE DA UFSM BASEADO NA TEORIA DA RESPOSTA AO ITEM

*Prof. Dr. Ivan Henrique Vey¹
Téc. Adm. Ricardo da Silva Barcelos²*

Introdução

No primeiro semestre de 2016 da UFSM aconteceu a aplicação do questionário de avaliação docente proposto pela Comissão Própria de Avaliação (CPA) em conjunto com a Coordenadoria de Planejamento e Avaliação Institucional (COPLAI). A avaliação docente pelos discentes é uma exigência do MEC fazendo parte do grande processo de avaliação institucional. Esta, por sua vez, está inserida no Sistema Nacional de Avaliação da Educação Superior – o SINAES.

A aplicação aconteceu em caráter de teste, sendo que a partir do segundo semestre de 2016 sua aplicação será de caráter permanente e definitivo. Neste sentido, este breve estudo procurou analisar de maneira global os resultados da avaliação docente e verificar a validade do instrumento de medida (questionário). Para atingir este objetivo, o estudo utilizou-se além da Teoria Clássica dos Testes (TCT) a Teoria da Resposta ao Item (TRI).

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 *et al.*, 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.

¹ Professor Associado no Depto. de Ciências Contábeis da Universidade Federal de Santa Maria.

² Técnico-administrativo em educação lotado no Centro de Tecnologia da Universidade Federal de Santa Maria, pós-graduando do Programa de Pós-Graduação em Gestão de Organizações Públicas da Universidade Federal de Santa Maria.

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.

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

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.

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.

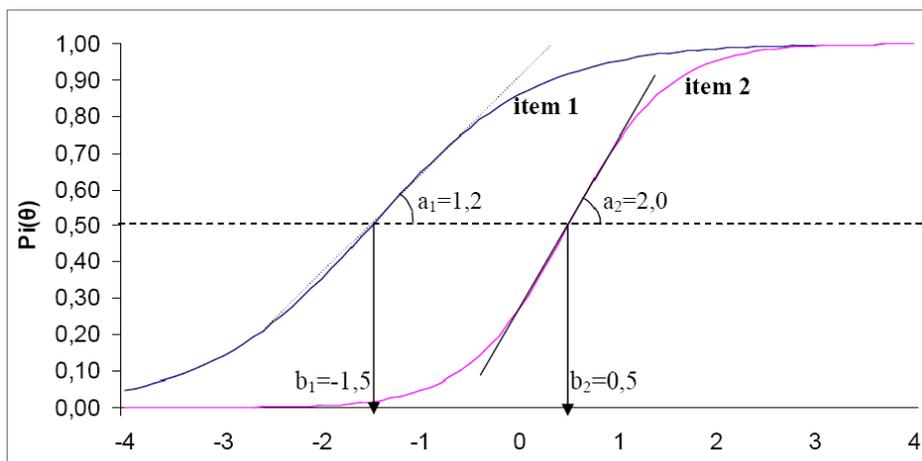


Figura 1 – Curva característica do item

Fonte: 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 no 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 possuírem um b_i menor serão itens de fácil aprovação, ou seja, itens que o respondente não necessita de uma elevada percepção 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.

Tratamento dos Dados

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.

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 (2004) coloca que quando o resultado do coeficiente se aproxima de 1 (um), pode-se afirmar que o teste possui um coeficiente de precisão.

Tabela 1 – Número de respondentes

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

Fonte: elaborado pelos autores.

Foram analisadas 23.871 respostas aos itens, destas, 20.543 foram consideradas na análise do alfa de Cronbach. 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

Fonte: elaborado pelos autores.

Tabela 1 – 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

Fonte: elaborado pelos autores.

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 n^o 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. Na análise utilizou-se a análise fatorial exploratória tendo sido utilizado o método de componentes principais para estimação. Na sequência são apresentados os resultados extraídos com a utilização do *software* SPSS.

Tabela 2 – Variância total explicada

Componente	Valores próprios iniciais			Somadas de extração de carregamentos ao quadrado		
	Total	% de variância	% cumulativa	Total	% de variância	% cumulativa
1	7,178	55,219	55,219	7,178	55,219	55,219
2	0,837	6,440	61,659			
3	0,801	6,161	67,820			
4	0,702	5,401	73,221			
5	0,529	4,073	77,294			
6	0,487	3,743	81,036			
7	0,433	3,333	84,370			
8	0,422	3,248	87,618			
9	0,386	2,967	90,585			
10	0,367	2,823	93,409			
11	0,330	2,541	95,949			
12	0,270	2,080	98,030			
13	0,256	1,970	100			

Fonte: elaborado pelos autores.

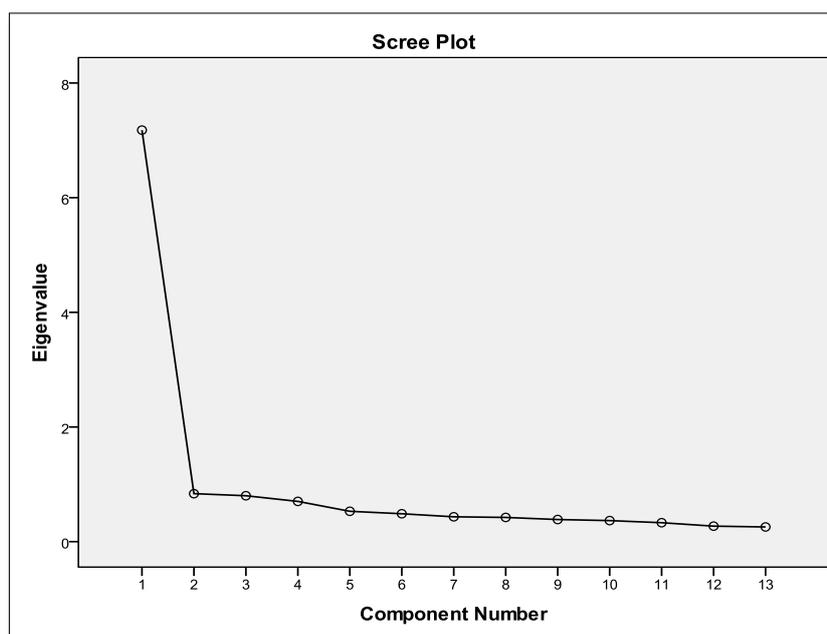


Figura 2 – Variância total explicada

Fonte: elaborado pelos autores.

Com base nos resultados apresentados pode-se afirmar que existe um fator (variável) determinante responsável por 55,21% da variabilidade dos dados. Assim, pode-se afirmar que o instrumento de medida é unidimensional, ou seja, está medindo um único traço latente, o desempenho dos docentes.

Desta forma, 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.

Análise baseada na Teoria da Resposta ao Item (TRI)

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

Tabela 5 – Escala de concordância

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

Fonte: CPA/UFSM.

Após análise da média e mediana dos itens decidiu-se cortar todos os itens no 3, ou seja, 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 completa (específica para itens dicotomizados) utilizando-se o *software* TESTFACT.

O resultado obtido novamente 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. Assim, a correlação bisserial fornece uma análise preliminar do conjunto de itens.

Tabela 3 – 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

Fonte: elaborado pelos autores.

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. Assim, diante dos resultados da correlação bisserial pode-se dizer que todos os itens apresentam correlação significativa com o escore bruto do conjunto de itens.

Na figura 3 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.

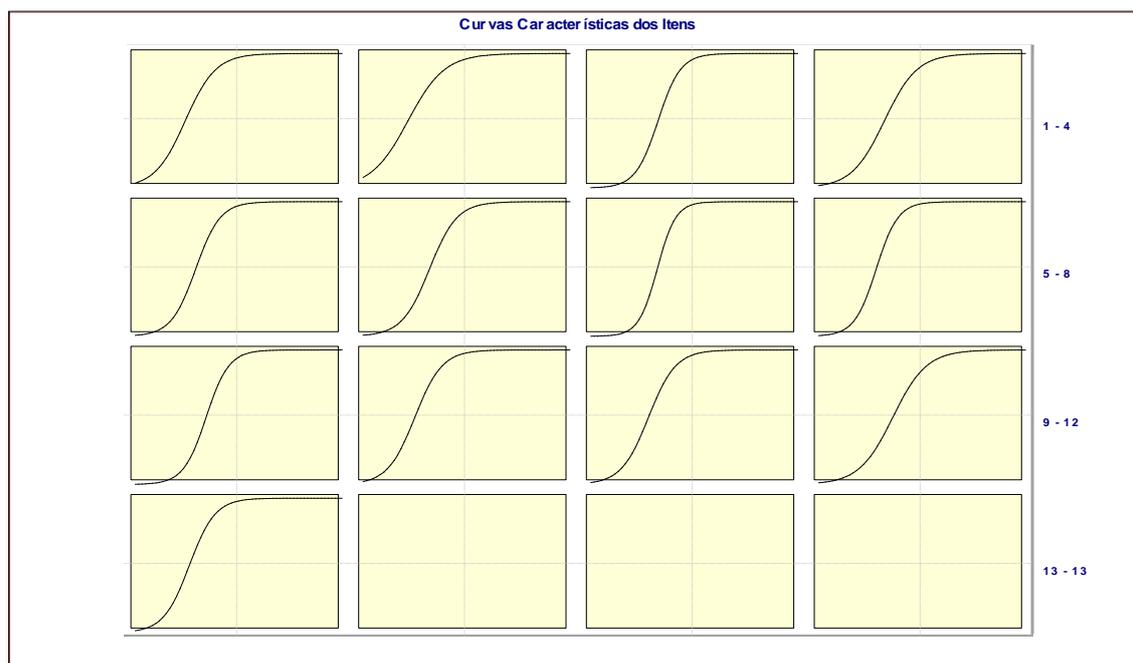


Figura 3 – Curvas características dos itens (CCI)

Fonte: elaborado pelos autores.

Tabela 7 – Parâmetros a (discriminação) e b (dificuldade) dos itens

Item	Parâmetro a	Parâmetro b
01	2,257	-1,552
02	1,881	-1,698
03	3,072	-1,053
04	2,107	-1,094
05	2,812	-1,249
06	2,509	-1,075
07	3,723	-1,077
08	3,177	-1,333
09	3,045	-0,945
10	2,532	-1,499
11	2,537	-1,330
12	2,049	-0,850
13	2,680	-1,433

Fonte: elaborado pelos autores.

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 apresentou um elevado 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. Assim, considerando que todos os itens apresentaram valor superior a 1 (um) para o parâmetro a , pode-se dizer que todos os itens possuem uma boa discriminação.

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 parâmetro de dificuldade. Sendo assim, este aspecto poderá ser trabalhado para que o desempenho seja melhorado em futuras avaliações. Por outro lado, o item 02 (“O professor comparece às aulas e cumpre os horários de início e de término das mesmas”) foi o que apresentou menor valor no parâmetro b , indicando que na percepção dos discentes é um item em que os docentes tiveram avaliação satisfatória.

Conclusões

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 (COPLAI), 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 dimensionalidade, 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 dimensionalidade 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 indica que os itens se ajustam bem ao escore bruto produzido pelo teste.

Do mesmo modo, foi novamente testada a dimensionalidade do construto através da análise fatorial de informação completa, específica para itens dicotomizados. O resultado obtido novamente 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. Por fim, a análise do parâmetro a revelou que o conjunto de itens é consistente, pois todos apresentaram capacidade de discriminação superior a um.

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

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. Teoria de resposta ao item: conceitos e aplicações. ABE — Associação Brasileira de Estatística, 4^o SINAPE, 2000.

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.

NUNNALLY, J. C. Psychometric theory. 2. ed. New York: McGraw-Hill, 1978.

PASQUALI, L. Psicometria: teoria dos testes na psicologia e na educação. Petrópolis, RJ: Vozes, 2004.

PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item. Avaliação Psicológica, v. 2, n. 2, p. 99-110, 2003.

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

SOARES, T. M.; MENDONÇA, M. C. M. Construção de um modelo de regressão 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.

VEY, I. H. Avaliação de desempenho logístico no serviço ao cliente baseada na Teoria da Resposta ao Item. Florianópolis; 2011. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina.