

A UTILIZAÇÃO DA TEORIA DA RESPOSTA AO ITEM COMO FERRAMENTA PARA A CONSTRUÇÃO DE ESCALAS NA ÁREA DE GESTÃO*

USE OF ITEM RESPONSE THEORY AS A TOOL TO CONSTRUCTION SCALES IN MANAGEMENT AREA

FERNANDO DE JESUS MOREIRA JUNIOR**

RESUMO

A Teoria da Resposta ao Item (TRI) é uma metodologia que sugere formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta a um item (questão), os traços latentes do indivíduo e as características dos itens, por meio de modelos matemáticos. Um dos grandes avanços da TRI é a possibilidade de criar escalas onde itens e respondentes são posicionados simultaneamente, o que permite a interpretação dessa escala. O objetivo desse artigo é apresentar uma discussão sobre a Teoria da Resposta ao Item como método para a construção de escalas, mostrando como a TRI pode contribuir para a criação de escalas na área de Gestão e apresentando as escalas criadas por meio da TRI no Brasil nas áreas relacionadas com a Gestão. Esse trabalho enfatiza os modelos unidimensionais da TRI (um único traço latente), que têm sido os modelos utilizados pelos pesquisadores da área da Gestão no Brasil. Os resultados mostraram que ainda existe pouca aplicação da TRI na área de Gestão no país, em comparação com outras metodologias empregadas, no entanto, essa metodologia deve ganhar espaço, principalmente pelas vantagens apresentadas. O acompanhamento longitudinal dos indivíduos e a manutenção e atualização da escala são os desafios para o futuro.

Palavras-chave: Teoria da Resposta ao Item. Construção de Escalas. Gestão.

ABSTRACT

Item Response Theory (TRI) is a method that suggests ways to represent the relationship between the probability of an individual to response to a certain item, the latent trait and characteristics of the individual items by mathematical models. The major advancement of IRT is the possibility of creating scales where items and respondents are positioned simultaneously, which allows interpretation of the scale. The aim of this paper is to present a discussion on Item Response Theory as a method for the construction of scales, showing how IRT can contribute to the creation of scales in the area of Management and presenting the scales created by IRT in Brazil in areas related to management. This article emphasizes the unidimensional IRT models (a single latent trait), which are the models used by researchers in the field of Management in Brazil. Results showed there is little application of TRI in the management area in the country, compared to other methods used, however, this methodology should gain space, mainly due to the advantages presented. Longitudinal follow-up of individuals and the maintenance and update of the scale are the challenges for the future.

Keywords: Item Response Theory. Construction of Scales. Management.

* Data de submissão: 31/07/2013. Data de aceite: 18/03/2014.

** Professor do Departamento de Estatística / UFSM

1 INTRODUÇÃO

A Teoria da Resposta ao Item (TRI) é uma metodologia que sugere formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta a um item (questão), os traços latentes do indivíduo e as características dos itens, por meio de modelos matemáticos (Andrade; Tavares; Valle, 2000). Traços latentes são características do indivíduo que não podem ser medidas diretamente, porém, são medidas por meio de um questionário com variáveis secundárias relacionadas com o traço latente. São exemplos de traços latentes: proficiência, grau de satisfação, importância atribuída, nível de ansiedade, grau de usabilidade, etc.

Em relação à avaliação tradicional da Teoria Clássica dos Testes (TCT) ou Teoria Clássica da Medida (TCM), em que o traço latente é mensurado por meio de um somatório ponderado, ou não, dos pontos atribuídos a cada questão do questionário, a TRI apresenta algumas vantagens (Embretson; Reise, 2000). Entre elas, destacam-se: (1) a TRI fornece informações mais precisas do desempenho dos respondentes, já que o traço latente do indivíduo não depende da dificuldade das questões do teste, enquanto que na TCT o escore do indivíduo depende essencialmente dos itens que compõe o teste (Andrade; Tavares; Valle, 2000; Vendramini; Silva; Canale, 2004); (2) a TRI permite obter índices de precisão do item (função de informação do item - FII) e do teste (função de informação do teste - FIT) mais ricos do que os índices utilizados pela TCT (Andrade; Tavares; Valle, 2000; Baker, 2001); (3) a TRI permite utilizar modelos que consideram a possibilidade de acerto casual (Andrade; Tavares; Valle, 2000), muito útil nas avaliações educacionais; (4) a TRI permite, sob certas condições, a comparação através do escore entre os indivíduos que responderam questionários com itens diferentes para medir o mesmo traço latente, pois os itens e os indivíduos são colocados numa mesma escala (Andrade; Tavares; Valle, 2000; Embretson; Reise, 2000); (5) na TRI, uma vez estimada a proficiência do indivíduo, é possível verificar qual a probabilidade de dar certa resposta a um determinado item que ele não respondeu (Vendramini; Silva; Canale, 2004); (6) na TRI, cada respondente tem seu próprio erro padrão, relacionado à sua habilidade ou traço latente, em que a estimação desse erro é mais precisa (Embretson; Reise, 2000); (7) na TRI, testes curtos podem ser mais confiáveis que os testes longos (Embretson; Reise, 2000); (8) a TRI permite a utilização de formatos mesclados de itens (por exemplo, dicotômicos e politômicos nominais e graduais) sem causar um impacto desequilibrado nos escores total do teste (Embretson; Reise, 2000).

A TRI foi utilizada pela primeira vez no Brasil em 1995 no SAEB (Sistema de Avaliação da Educação Básica), para montagem de instrumentos, tratamento de dados e construção de escalas a partir de resultados apresentados por alunos em provas de rendimento (Souza, 2005), o que permitiu que alunos de diferentes séries pudessem ser comparados e colocados na mesma escala, que não era possível fazer através da TCT. A maioria das aplicações têm sido na avaliação educacional (Moreira Junior, 2010), em que o ENEM (Exame Nacional do Ensino Médio) tem sido, atualmente, o melhor exemplo que mostra os benefícios da TRI. Nos últimos anos, a TRI tem sido aplicada em diversas áreas no Brasil, inclusive na área de gestão, como por exemplo, na Gestão da Qualidade Total (Alexandre et al., 2001; 2002a; 2002b; 2003a; 2003b; Batista; Alexandre, 2004; Batista et al., 2002; Oliveira, 2010; Vasconcelos et al., 2002), na avaliação da prática da melhoria contínua (Milnitz; Tubino, 2012), na gestão organizacional (Morais, 2009; Santos et al., 2009), na gestão do conhecimento (Almeida, 2009), na avaliação de intangíveis nas organizações (Vargas, 2007; Vargas et al., 2008; Moreira Junior; Vargas; Andrade, 2010), na avaliação do nível de satisfação (Bornia et al., 2009; Bortolotti, 2003; Bortolotti; Andrade, 2007; Bortolotti; Moreira Junior; Souza Junior, 2010; Bortolotti; Souza Junior; Andrade, 2009; Bortolotti et al., 2012; Costa; Chaves Neto, 2002; Cunha; Sena Junior; Matos, 2002; Moreira Junior et al., 2010), na avaliação da maturidade empresarial (Castro Junior, 2007; Pereira, 2007; Roriz Junior, 2008), na avaliação do potencial empreendedor (Alves; Bornia, 2011), e na avaliação de risco de crédito (Mafra, 2011).

Um dos grandes avanços da TRI, em relação à TCT, é a possibilidade de criar escalas nas quais itens e respondentes são posicionados simultaneamente, o que permite a interpretação dessa escala (Andrade; Tavares; Valle, 2000). O objetivo desse artigo é apresentar uma discussão sobre a Teoria da Resposta ao Item como método para a construção de escalas, mostrando como a TRI pode contribuir para a criação de escalas na área de Gestão e mostrando algumas escalas criadas por meio da TRI no Brasil nas áreas relacionadas com a Gestão. Embora seja mencionado que existem diversos tipos de modelos da TRI, esse trabalho enfatiza os modelos unidimensionais da TRI (um único traço latente), que são os modelos utilizados pelos pesquisadores da área da Gestão no Brasil.

2 TEORIA DA RESPOSTA AO ITEM

2.1 CONCEITOS BÁSICOS

Como mencionado anteriormente, a TRI é uma metodologia que propõe uma maneira de representar a relação entre a probabilidade de uma determinada resposta a um item, o traço latente, e as características dos itens, por meio de modelos de probabilidade. Em geral, temos: (1) um conjunto de itens, (2) uma amostra de respondentes e (3) um modelo de probabilidade associado, também chamado de Modelo de Resposta ao Item (MRI).

Uma análise de TRI normalmente segue o procedimento descrito a seguir. O conjunto de itens deve ser elaborado da forma tradicional, ou seja, por especialistas sobre o traço latente analisado, consultando a literatura existente sobre o assunto, abrangendo todo o conteúdo necessário (tendo o cuidado para que não fique demasiadamente extenso), fazendo pré-teste e análise de juízes, quando for conveniente. O conjunto de itens resultantes irá compor o questionário, ou seja, o instrumento de avaliação. Esse será submetido a uma amostra suficiente de respondentes. Esses respondentes podem ser indivíduos ou não (empresas, objetos, animais, etc.), dependendo da situação. As respostas obtidas serão analisadas por meio de um MRI, o qual irá fornecer estimativas para os parâmetros dos itens desse modelo. Os itens serão analisados por meio dessas estimativas e por meio da Função de Informação do Item (FII), que mede a qualidade do item para o construto. Se algum item tiver qualidade duvidosa, poderá ser eliminado da análise. Outros itens, com boas qualidades, irão se tornar “itens âncoras”, conceito que será explorado na Seção 2.3. Nesse momento, será criada a escala do traço latente, onde os itens serão posicionados. Finalizada a avaliação dos itens, a TRI irá avaliar os respondentes, atribuindo uma “nota” (um valor) para cada respondente, que é a estimativa para o parâmetro dos respondentes. Nesse momento, os respondentes poderão ser posicionados na escala criada. Com itens e respondentes posicionados na mesma escala, pode-se fazer a análise e interpretação da escala.

Tanto para estimação dos parâmetros dos itens quanto para a estimação do traço latente, há vários métodos estatísticos sofisticados que podem ser utilizados, tais como o método de Máxima Verossimilhança Marginal (MVM), o método bayesiano da Moda a Posteriori (MAP), o método bayesiano da Média a Posteriori (EAP) e o método da Máxima Verossimilhança Conjunta (MVC). Esses métodos não possuem solução explícita, o que torna necessária

a utilização de algum método numérico iterativo, como o Algoritmo Newton-Raphson (Issac; Keller, 1966), o Método Scoring de Fisher (Rao, 1973) e o Algoritmo EM (Dempster; Laird; Rubin, 1977). Essas soluções envolvem cálculos bastante complexos e, conseqüentemente, necessitam de programas computacionais específicos. Os principais softwares utilizados para análise de TRI são o BILOG-MG, o MULTILOG e o PARSCALE (Toit, 2003), o R (por exemplo, Partchev, 2013; Rizopoulos, 2013; Weeks, 2011) e o Xcalibre (Weiss; Guyer, 2010).

2.2 MODELOS DE RESPOSTA AO ITEM

Existem vários modelos matemáticos utilizados na TRI, diferentes quanto à sua função e à quantidade de parâmetros, e cada um deles é específico para uma (ou mais) situação. Esses modelos podem ser classificados quanto à sua dimensão (unidimensionais ou multidimensionais), quanto ao tipo de traço latente (cumulativo ou não cumulativo), quanto ao tipo de item (dicotômico ou politômico) e quanto ao número de populações envolvidas (Moreira Junior, 2011).

Os modelos unidimensionais são utilizados quando se supõe que o objeto de estudo é composto por um único traço latente que está relacionado com a capacidade ou habilidade do indivíduo em responder aos itens do teste, ou quando o traço latente pode ser representado por uma única dimensão ou fator, o qual deve explicar mais de 20% da variância total, segundo Reckase (1979). Já os modelos multidimensionais são adequados quando se estuda mais de um traço latente ou quando o traço latente não pode ser representado por uma única dimensão.

Os modelos cumulativos devem ser utilizados quando o traço latente analisado possui característica acumulativa, ou seja, quando há uma relação de dominância entre itens e sujeitos (que são posicionados na mesma escala), de tal forma que se um indivíduo domina certo item, conseqüentemente ele domina também todos os itens que estão posicionados abaixo desse item na escala. Por outro lado, os modelos não cumulativos são adequados quando o traço latente não possui característica cumulativa, ou seja, quando há uma relação de dominância entre itens e sujeitos (que são posicionados na mesma escala), de tal forma que se um indivíduo os itens que estão posicionados próximos dele.

Os modelos dicotômicos são utilizados quando os itens são de natureza dicotômica ou quando eles podem ser dicotomizados ou avaliados de forma dicotomizada, enquanto que os modelos politômicos são utilizados quando os itens possuem

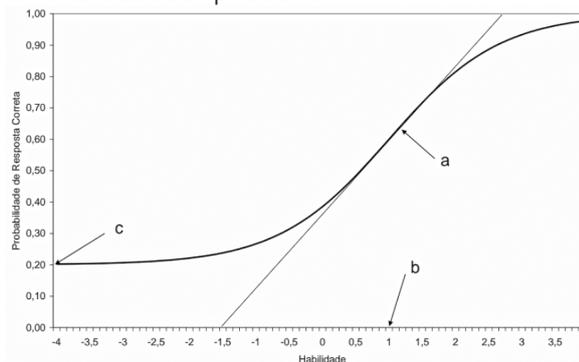
mais de duas categorias de respostas que podem ser ordinais ou nominais.

Os modelos também podem variar quanto ao número de populações e de amostras (da mesma população ou não). Nesses casos, deve-se fazer uma equalização, ou seja, colocar os parâmetros dos diferentes grupos numa mesma escala. Essa equalização pode ser feita por meio do método dos Grupos Múltiplos (Bock; Zimowski, 1997), se o software permitir, ou por meio de métodos de equalização a posteriori (Andrade; Tavares; Valle, 2000).

Os modelos mais utilizados na TRI são: os modelos logísticos com 1 parâmetro – ML1 (Wright, 1968), também conhecido como Modelo de Rasch (Rasch, 1960), com 2 parâmetros – ML2 (Lord, 1980; Birnbaum, 1968), e com 3 parâmetros – ML3 (Lord, 1980; Birnbaum, 1968), o modelo de resposta nominal – MRN (Bock, 1972), modelo de resposta gradual – MRG (Samejima, 1969), o modelo de escala gradual – MEG (Andrich, 1978), o modelo de crédito parcial – MCP (Masters, 1982), o modelo de crédito parcial generalizado – MCPG (Muraki, 1992) e o modelo de desdobramento graduado generalizado – GGUM (Roberts; Donoghue; Laughlin, 2000). A seguir, são apresentadas duas figuras para ilustrar alguns desses modelos.

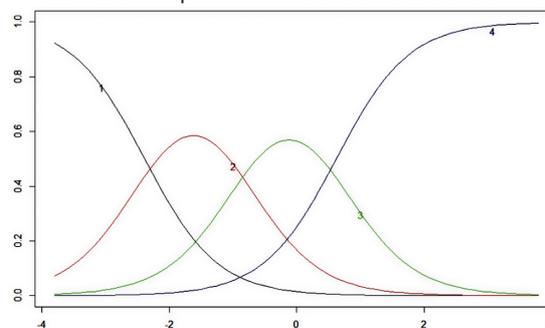
A Figura 1 apresenta um exemplo de uma Curva Característica do Item (CCI) de um modelo dicotômico, o ML3, e a relação existente com os parâmetros do item a_i (poder de discriminação), b_i (posição do item na escala) e c_i (probabilidade de acerto casual de indivíduos com baixa habilidade). Devido à existência do parâmetro c_i , o ML3 é adequado para avaliações educacionais. Quando o parâmetro c_i não é aplicável, ou seja, se não é adequado avaliar o acerto casual no traço latente analisado, o seu valor é igual a zero, o que equivale ao ML2. Além disso, se todos os itens tiverem a mesma discriminação com o parâmetro a_i igual a um, então o modelo se reduz a um ML1. O eixo X indica o valor do traço latente (habilidade), considerando uma escala com média 0 e desvio padrão 1, enquanto que o eixo Y indica a probabilidade do indivíduo j responder positivamente o item dicotômico i , dado que ele tem habilidade Θ_j . Por exemplo, pode-se observar que um indivíduo com habilidade igual a 1, nessa escala, tem probabilidade igual a 0,60 de responder positivamente o item dicotômico i .

Figura 1 – Exemplo gráfico de item dicotômico
Fonte: elaborado pelo autor.



A Figura 2 apresenta um exemplo de uma Curva Característica do Item (CCI) de um modelo politômico, o MRG, em que cada curva representa a probabilidade de resposta a uma das quatro categorias do item. Segundo as características desse modelo, há quatro parâmetros representado nessa figura: um parâmetro a_i (poder de discriminação), e três (o número de categorias menos um) parâmetros b_i (posição do item). O eixo X indica o valor do traço latente (habilidade), considerando uma escala com média 0 e desvio padrão 1, enquanto que o eixo Y indica a probabilidade do indivíduo j responder cada categoria do item politômico i , dado que ele tem habilidade Θ_j . Por exemplo, pode-se observar que um indivíduo com habilidade igual a 1, nessa escala, tem maior probabilidade de responder a categoria 4 desse item.

Figura 2 – Exemplo gráfico de item politômico
Fonte: elaborado pelo autor.



2.3 MÉTODOS PARA A CONSTRUÇÃO DE ESCALA NA TRI

Segundo Andrade, Tavares e Valle (2000), os valores da escala do traço latente (ou habilidade) podem assumir teoricamente qualquer valor real entre $-\infty$ e $+\infty$, diferentemente da TCT, em que a escala geralmente varia entre 0 e a quantidade total de questões do teste. No entanto, o importante nessa escala não é a sua magnitude, mas as relações de ordem existentes. Dessa forma, é preciso

estabelecer valores para a média e para o desvio padrão que representem a escala de habilidade dos indivíduos na população.

A calibração dos itens, ou seja, a estimação dos parâmetros dos itens, devido à facilidade computacional, geralmente é feita na escala (0,1), ou seja, numa escala com média igual a zero e desvio padrão igual a 1. Após essa etapa, é realizada a construção da escala do traço latente, que é colocada na mesma métrica dos itens. Conforme, Fontanive, Elliot e Klein (2007), as escalas de habilidade ordenam o desempenho dos indivíduos do menor para o maior de forma contínua e são cumulativas, isto é, os indivíduos que situam-se em um determinado nível da escala são capazes de demonstrar as habilidades descritas nesse nível e nos níveis anteriores dessa escala.

No entanto, na prática, existe uma dificuldade em compreender os valores negativos e decimais que existem na escala (0,1), na qual muitas vezes os valores das proficiências são interpretados inadequadamente (Valle, 2001). Nesse sentido, é bastante comum fazer uma transformação linear em todos os parâmetros envolvidos antes da construção das escalas, fazendo com que todos os valores fiquem positivos, de forma a representar a habilidade em valores numéricos que tornem mais fácil o entendimento pelas pessoas.

Conforme foi dito na seção 2.1, a construção da escala de habilidade é efetuada após a calibração (e equalização, se necessário) dos itens, com o objetivo de encontrar uma interpretação qualitativa dos valores obtidos pela aplicação do modelo da TRI, possibilitando, assim, a interpretação pedagógica dos valores das habilidades. Nesse sentido, surge a ideia dos níveis âncoras e a técnica conhecida como ancoragem (Beaton; Allen, 1992).

Andrade, Tavares e Valle (2000) definem níveis âncora como pontos selecionados pelo analista na escala da habilidade para serem interpretados pedagogicamente. Valle (2001) ressalta que esses níveis âncoras não podem ser muito próximos nem muito distantes, podendo-se tomar como base a média e o desvio padrão. Usualmente, os níveis âncoras são estabelecidos em função do desvio padrão em relação à média, obtendo-se, normalmente, cerca de até 7 níveis âncoras ($-3\sigma+\mu, -2\sigma+\mu, -1\sigma+\mu, \mu, 1\sigma+\mu, 2\sigma+\mu, 3\sigma+\mu$). Outras configurações de níveis âncoras podem ser definidas (por exemplo, Klein; Fontanive; Elliot, 2007), porém, não são muito utilizadas na prática. No entanto, não se sabe a priori quantos níveis âncoras serão representados, já que isso depende da quantidade de itens âncoras.

Kolen e Brennan (1995) definem item âncora da seguinte forma: considere dois níveis âncora

consecutivos Y e Z, sendo que $Y < Z$. Um determinado item é âncora para o nível Z se e somente se as 3 condições abaixo forem satisfeitas simultaneamente:

$$P(U=1 | 0=Z) \geq 0,65,$$

$$P(U=1 | 0=Y) \leq 0,50 \text{ e}$$

$$P(U=1 | 0=Z) - P(U=1 | 0=Y) \geq 0,30.$$

Isso significa que, para um item ser considerado âncora em um determinado nível âncora, ele precisa ser respondido positivamente por um grande percentual de indivíduos (pelo menos 65%) com este nível de habilidade e por um percentual menor de indivíduos (no máximo 50%) com o nível de habilidade imediatamente anterior. Além disso, a diferença entre essas duas proporções deve ser de pelo menos 0,30. Assim, para um item ser âncora em um nível, ele deve ser um item “típico” desse nível, ou seja, bastante respondido positivamente por indivíduos com aquele nível de habilidade e pouco respondido positivamente por indivíduos com um nível de habilidade imediatamente inferior (Andrade; Tavares; Valle, 2000).

Na prática, às vezes um item não se caracteriza âncora por violar “levemente” uma das três condições necessárias. Nessas situações, pode-se considerar esse item como sendo âncora, se ele for importante ou se existirem poucos itens no instrumento de pesquisa. Outra alternativa é dividir os itens em grupos, segundo a quantidade de condições satisfeitas. Valle (2001) salienta que alguns níveis âncoras extremos podem ser mal caracterizados por serem definidos por itens muito fáceis ou muito difíceis, os quais geralmente são poucos.

Depois de identificados os itens âncoras de cada nível âncora, os especialistas no traço latente estudado devem caracterizar os níveis âncoras, segundo o conteúdo abordado no conjunto de itens que compõem cada nível. Após essa etapa, a escala está pronta para ser utilizada, por exemplo, para o posicionamento das populações ou dos indivíduos a fim de verificar os conteúdos dominados ou para identificar o percentual de indivíduos em cada nível de habilidade (Valle, 2001). Assim, um sujeito que se situa acima de um determinado nível dominará os conteúdos caracterizados por esse nível e pelos níveis inferiores a esse, porém, não dominará os assuntos dos níveis superiores a esse.

Observa-se que as condições estipuladas por Kolen e Brennan (1995) são válidas para os Modelos Logísticos (dicotômicos) da TRI. No entanto, podem ser adaptadas aos modelos politômicos, segundo Mafra (2010), no qual não é verificado se um item é âncora, mas se uma categoria do item é âncora. Nesse sentido, algumas abordagens alternativas são utilizadas para a interpretação da escala, nas quais pode-se considerar, por exemplo,

as seguintes situações: (1) dicotomizar as categorias adjacentes dos itens, uma a uma e considerar todas as categorias como candidatas a “categoria âncora”, (2) dicotomizar as categorias dos itens em dois grupos, tratando o item como dicotômico e usando a abordagem de Kolen e Brennan (1995), ou (3) considerar apenas a categoria mais alta como candidata a “categoria âncora”.

3 CONSTRUÇÃO DE ESCALAS POR MEIO DA TRI NA ÁREA DA GESTÃO

Nessa seção serão apresentados alguns resultados obtidos por meio da aplicação da Teoria da Resposta ao Item como método para a construção de escalas na área de Gestão no Brasil.

Alexandre et al. (2001; 2002a; 2002b; 2003a; 2003b) investigaram as práticas da GQT (Gestão da Qualidade Total) no setor manufatureiro de portes médio e grande do estado do Ceará. Foram analisados 43 itens politômicos de escala Likert com cinco categorias variando de “discordo totalmente” a “concordo totalmente”. Numa primeira análise (Alexandre et al., 2001; 2002a), os itens foram dicotomizados e foi utilizado o ML2. Em outra análise (Alexandre et al., 2002b), os itens foram ajustados por meio do MRG. Num outro trabalho (Alexandre et al., 2003b), os itens foram ajustados por meio do MEG e foi feito um estudo sobre o número adequado de categorias. No entanto, somente num estudo (Alexandre et al., 2003a) foi criada uma escala por meio da TRI segundo o ML2. Os resultados mostraram que, dos 43 itens analisados, 32 foram caracterizados como itens âncoras. Quanto à escala criada (com média zero e desvio padrão 1), foram identificados três níveis âncoras:

1. Nível -1 (com 3 itens âncoras): empresas situadas nesse nível estão relacionadas às práticas de controle estatístico de processo (CEP);
2. Nível 0,5 (com 17 itens âncoras): empresas situadas nesse nível possuem tanto práticas de CEP quanto práticas mais avançadas da GQT, como por exemplo, práticas associadas ao comprometimento da alta administração e foco no consumidor;
3. Nível 2 (com 12 itens âncoras): empresas situadas nesse nível possuem tanto práticas de CEP quanto práticas mais avançadas da GQT, além de práticas mais complexas da GQT, tais como, envolvimento de funcionários, parceria com fornecedores e benchmarking.

Alexandre et al. (2003a) concluem o estudo dizendo que a escala parece refletir um caráter acumulativo na maturidade na GQT e o modelo proposto parece ser adequado para a construção

de uma métrica que analisa o nível de maturidade organizacional na GQT. Os autores argumentam que a criação dessa métrica pode contribuir para que as empresas conheçam sua posição com relação ao nível de maturidade na escala construída, respondendo o questionário elaborado. Além disso, pode-se comparar as empresas com relação à maturidade na gestão pela qualidade, os autores sugerem que essa escala poderia ser utilizada, por exemplo, na concessão do Prêmio Nacional da Qualidade (PNQ) para aquelas empresas que estão no nível superior da escala, o que reduziria sensivelmente o custo associado à concessão desse prêmio. No entanto, devido a pouca quantidade de empresas analisadas (75 empresas), os autores ressaltam que o estudo não pretende validar a escala construída.

Oliveira (2010) seguiu o mesmo objetivo de Alexandre et al. (2003a), no entanto, enfatizou a evolução do grau de maturidade, quanto à utilização de ferramentas e técnicas da Gestão pela Qualidade Total (GQT) nos últimos 10 anos, nas empresas de transformação e construção civil de portes médio e grande no estado do Ceará. Para tanto, foi utilizado o mesmo questionário de Alexandre et al. (2003a), no qual foram acrescentados alguns itens (devido à evolução na aplicação das práticas da GQT nos últimos anos), totalizando 61, e foram analisadas 166 empresas no total, 75 em 1998 e 91 em 2009, sendo 16 delas respondentes nos dois momentos. Nesse estudo foi utilizado o MRG e foi criada uma escala denominada Escala do Grau de Maturidade (EGM) com média 100 e desvio padrão 25. Foram considerados tanto os itens âncoras (22 itens) quanto os quase âncoras (18 itens), totalizando 40 itens, que estão nos níveis 50 (2 itens), 75 (6 itens), 100 (2 itens) e 125 (10 itens) na EGM. As empresas utilizadas na pesquisa de Alexandre et al. (2003a) tiveram seu traço latente reestimado na escala (100, 25) para que pudesse ser feita uma comparação com as estimativas atuais. Verificou-se um crescimento no grau de maturidade médio foi de apenas 8,43. No entanto, o nível de maturidade 150 foi atingido por 15,4% das empresas, quase seis vezes mais que a dez anos. Uma importante contribuição do estudo foi a criação de uma escala padronizada do grau de maturidade, que permite a comparação entre empresas e também um acompanhamento longitudinal dos avanços de determinada empresa em relação à GQT. Os resultados obtidos podem ser utilizados pelas empresas para eleger práticas prioritárias para aumentar sua maturidade e melhorar os resultados obtidos com a qualidade dos produtos. Os resultados mostraram que as práticas mais básicas estão relacionadas ao comprometimento da alta administração, sele-

ção do fornecedor tendo preço e qualidade como critério, resultados das avaliações utilizados como suporte para melhoria da qualidade, apoio aos funcionários para a solução de problemas e manutenção de programas sobre a prevenção de acidentes. Por outro lado, as práticas de parceria com fornecedor, fornecimento de resultados da qualidade aos funcionários, programa de melhoria contínua, *Benchmarking*, padrão de exigência ambiental a fornecedores e questões de ética e responsabilidade social mostraram-se de nível mais alto na escala de grau de maturidade.

Milnitz e Tubino (2012) avaliaram o nível de prática da melhoria contínua de funcionários de uma empresa. O questionário composto por 34 itens com três categorias de resposta (nunca, às vezes, sempre) foi aplicado a 519 funcionários de uma empresa do setor têxtil que já aplicava os conceitos de melhoramento contínuo. Os itens foram avaliados segundo o MRN e foi criada uma escala para o nível da prática de melhoria contínua nas empresas com média 100 e desvio padrão 10. Foram identificados cinco níveis âncoras, com 20 itens âncoras, descritos a seguir:

1. N1 (score 104) - Pré Melhoria Contínua (3 itens): o interesse no conceito já foi iniciado (por uma crise, visita a outra organização, etc.), mas a implementação é feita de forma pontual.

2. N2 (score 106) - Melhoria Contínua estruturada (4 itens): há um compromisso formal de construir um sistema que irá desenvolver a melhoria contínua por toda a organização.

3. N3 (score 108) - Melhoria Contínua orientada para metas (7 itens): nesse nível existe o compromisso de estabelecer relações no comportamento da melhoria contínua, relacionando o nível de atuação operacional com os objetivos estratégicos mais abrangentes da organização.

4. N4 (score 110) - Melhoria Contínua Pró-ativa (5 itens): Há uma tentativa de dar autonomia e poder aos indivíduos e grupos para gerenciar e dirigir seus próprios processos e iniciativas de ações de melhoria.

5. N5 (score 112) - Plena capacidade em Melhoria Contínua (1 item): Aproxima-se do modelo das organizações de aprendizagem. O aprendizado organizacional ocorre de modo extensivo e largamente distribuído, é comum a iniciativa de procura e solução sistemática de problemas, de forma mais proativa.

Milnitz e Tubino (2012) concluem que o modelo proposto serve de guia para ações de intervenção na medida do nível de prática da melhoria contínua nas organizações através da avaliação dos indivíduos que as compõem. No entanto, salientam que é necessário melhorar a escala proposta

por meio da elaboração de novos itens que possam representar melhor os níveis com poucos itens âncoras na escala desenvolvida.

Castro Junior (2007) avaliou a percepção dos funcionários sobre a maturação das empresas de tecnologia no planejamento de Sistemas de Informação Estratégica. Foi utilizado um questionário com 30 itens com cinco categorias de resposta, indo de 1 (não atendido) até 5 (atendido) aplicado a uma amostra, cujo tamanho não foi informado, de funcionários de uma empresa de telecomunicações. O autor utiliza o ML1, no entanto, não mostra como os itens foram dicotomizados e não apresenta as estimativas dos parâmetros do Modelo utilizado. Além disso, o autor não cria uma escala para o traço latente analisado.

Roriz Junior (2008) seguiu a mesma metodologia utilizada por Castro Junior (2007) para avaliar o nível da maturidade das Indústrias para a implantação do comércio eletrônico B2B. Foi utilizado um questionário de 41 itens com sete categorias de resposta para quantificar a intensidade que os itens inibidores aplicado a uma pequena amostra de 20 funcionários. Cabe salientar que uma amostra de 20 indivíduos é muito pequena para uma análise de TRI, pois não permite uma estimação precisa e adequada dos parâmetros dos itens (Nunes; Primi, 2005). Da mesma forma que Castro Junior (2007), o autor utiliza o ML1, mas não mostra como os itens foram dicotomizados, não apresenta as estimativas dos parâmetros do Modelo utilizado e não cria uma escala para o traço latente analisado.

Mafra (2011) desenvolveu uma sistemática para avaliação do risco de crédito por meio da TRI. Foi utilizado um questionário com 26 itens aplicados a 289 empresas. A autora desenvolveu duas escalas de medida de avaliação de risco de crédito com base nos dados dicotomizados (MRISCd) e nos dados politômicos (MRISCP), ambas com média 50 e desvio padrão 10. Para os dados dicotomizados, fora utilizado o ML2, enquanto que para os dados politômicos, fora utilizado o MRG. A MRISCd identificou quatro níveis âncoras e 9 itens âncoras:

1. Nível 60 (3 itens): estão localizadas as empresas que possuem transparência e bons relacionamentos com fornecedores e clientes.
2. Nível 70 (3 itens): estão localizadas as empresas que acumulam todas as características do nível anterior e ainda possuem bons relacionamentos com os colaboradores, são organizadas e fazem separação de bens.
3. Nível 80 (2 itens): estão localizadas as empresas que acumulam todas as características dos níveis anteriores e ainda são mais

capacitadas e possuem melhores práticas de qualidade.

4. Nível 90 (1 item): estão localizadas as empresas que acumulam todas as características dos níveis anteriores e ainda possuem melhores condições de mercado.

A MRISCd identificou seis níveis âncoras, 8 itens âncoras e 15 categorias âncoras:

1. Nível 30 (1 item e 1 categoria): estão localizadas as empresas que possuem bons relacionamentos com os colaboradores.
2. Nível 40 (2 itens e 2 categorias): estão localizadas as empresas que acumulam todas as características do nível anterior e ainda possuem boa organização e boa transparência.
3. Nível 50 (2 itens e 2 categorias): estão localizadas as empresas que acumulam todas as características dos níveis anteriores e ainda possuem boa qualidade e boa separação de bens.
4. Nível 60 (3 itens e 5 categorias): estão localizadas as empresas que acumulam todas as características dos níveis anteriores e ainda possuem bom e ótimo relacionamento com fornecedores e clientes e ótima transparência.
5. Nível 70 (4 itens e 4 categorias): estão localizadas as empresas que acumulam todas as características dos níveis anteriores e ainda possuem ótimo relacionamento com os colaboradores e ótima qualidade, organização e separação de bens.
6. Nível 80 (1 item e 1 categoria): estão localizadas as empresas que acumulam todas as características dos níveis anteriores e ainda possuem ótima capacitação.

Mafra (2011) conclui que o MRISCp apresentou um ganho de informação relevante em comparação ao MRISCd, apresentado uma quantidade maior de informação para o construto risco de crédito, nas escalas de medida criadas. Quanto à avaliação das empresas para a análise do risco de crédito, conclui-se que empresas classificadas a partir do nível 50, que corresponde ao escore médio, passariam pela primeira fase da avaliação, pois atendem aos critérios mínimos relacionadas aos itens de caráter. No entanto, a autora recomenda uma avaliação final tendo em vista a pouca quantidade de itens presentes no modelo final.

Vargas (2007) e Vargas et al. (2008) elaboraram um instrumento composto por 33 itens com cinco categorias (não praticado, quase nada praticado, pouco praticado, bastante praticado e muito praticado) para avaliar os aspectos intangíveis das organizações. O questionário foi aplicado a uma

amostra de 203 indústrias de Santa Catarina. Com base no ML2, foi criada a Escala de Medida do Desempenho nos Intangíveis (EMDI), com média 500 e desvio padrão 50, com quatro níveis âncoras e 14 itens âncoras, assim definidos:

1. Nível 500 (1 item): empresas caracterizadas pela prática de preparar adequadamente uma pessoa que está começando na empresa para o exercício de um cargo e oferecer condições que facilitem sua aprendizagem e plena integração na organização.
2. Nível 550 (4 itens): empresas caracterizadas por: formalizar políticas de incentivos financeiros para cursos técnicos profissionalizantes, oficinas, seminários, palestras, etc.; proporcionar aos seus colaboradores treinamentos técnicos e desenvolvimento gerencial, quando surge uma oportunidade; possuir centro próprio de treinamento e realizar programas de qualidade de vida no trabalho; priorizar o recrutamento interno e considerar os interesses e os objetivos dos colaboradores na participação dos programas; prever critérios de avaliação na fase de planejamento; fazer análise de custo benefício de cada programa; realizar de modo formal e informal a avaliação dos resultados alcançados; e executar melhorias a partir desses resultados. Essas empresas também acumulam todas as características dos níveis anteriores.
3. Nível 600 (8 itens): empresas caracterizadas por: possuírem estruturas para prever em orçamento anual os programas de desenvolvimento gerencial; possuírem sistemas de divulgação interna das competências necessárias para cada cargo, bem como sistemas de avaliação de desempenho que fornece feedback aos colaboradores, recompensando-os por apresentarem melhoria nos seus desempenhos, tanto individual como por equipe; também fazem parcerias com instituições de ensino superior para ministrar educação continuada a seus colaboradores. Essas empresas também acumulam todas as características dos níveis anteriores.
4. Nível 650 (1 item): empresas caracterizadas por realizarem parcerias com instituições de ensino superior para desenvolver programas personalizados de acordo com suas competências essenciais e possuírem um sistema estruturado de administração de carreiras. Essas empresas também acumulam todas as características dos níveis anteriores.

Vargas *et al.* (2008) concluem que a criação da EMDI é útil para identificar em que nível cada uma das empresas respondentes se encontra na escala e atribuir a interpretação associada à sua posição, verificar quais as empresas que estão no mesmo nível e comparar o desempenho das empresas entre si em relação à avaliação dos intangíveis e descobrir quais intangíveis a empresa necessita desenvolver para melhorar seu nível de desempenho.

Moreira Junior, Vargas e Andrade (2010) fizeram uma atualização nas estimativas dos parâmetros dos itens desenvolvidos por Vargas *et al.* (2008), utilizando uma amostra de 537 indústrias, sendo 203 catarinenses, já analisadas por Vargas (2007) e mais 334 gaúchas. Esse novo estudo identificou cinco níveis âncoras e 17 itens âncoras. Apesar de ter sido utilizada uma escala com média 0 e desvio padrão 1, em geral, os itens âncoras mantiveram a mesma ordem em termos de nível dentro da EMDI, porém, alguns novos itens foram acrescentados:

1. Nível -1 (1 item): Corresponde ao nível 500 da EMDI e possui a mesma interpretação.
2. Nível 0 (1 item): Empresas que priorizam o recrutamento interno (item que não era âncora no EMDI). Essas empresas também acumulam todas as características dos níveis anteriores.
3. Nível 1 (11 itens): Esse nível possui todos os itens dos níveis 550 e 600 da EMDI com exceção de 2 que não se caracterizaram e de outro que se alocou no nível 2. Além disso, mais dois itens que não haviam sido âncoras na EMDI, nessa análise foram âncoras nesse nível.
4. Nível 2 (2 itens): Esse nível alocou um item do nível 600 da EMDI e mais um item que não haviam sido âncoras na EMDI.
5. Nível 3 (1 item): Corresponde ao nível 650 da EMDI e possui a mesma interpretação.

Moreira Junior, Vargas e Andrade (2010) concluíram que o aumento da amostra não mudou significativamente a ordem com que os itens foram alocados na escala, no entanto, melhorou a precisão das estimativas, o que contribuiu para o aparecimento de novos itens âncoras e o enriquecimento da interpretação da escala.

Alves e Bornia (2011) criaram uma escala para medir o potencial empreendedor das pessoas, com base no ML2 da TRI. Foi elaborado um questionário com 103 questões dicotômicas, aplicado a 764 estudantes de graduação e pós-graduação de instituições de ensino superior. Nem todos os estudantes responderam exatamente o mesmo questionário (os quais possuíam itens em comum), por isso foi necessário recorrer aos métodos de equalização da TRI. Foram identificados 6 níveis âncoras e 42

itens âncoras e quase âncoras na escala com média 8 e desvio padrão 2:

1. Nível 5 (7 itens): Os indivíduos posicionados nesse nível ou acima gostam de realizar seu trabalho de forma assídua e correta, procuram obter aprendizado sobre os assuntos pertinentes ao ramo em que atuam e saber a amplitude dos riscos para tomar decisões, entendem que obstáculos existem para ser superados e não se abatem com os erros cometidos, buscando de forma permanente atingir seus objetivos, e procuram prestar suporte à sua rede de relações.
2. Nível 6 (17 itens): Os indivíduos posicionados nesse nível ou acima além das características dos níveis anteriores, têm expectativas altas para consigo, são hábeis em identificar oportunidades, enxergar as necessidades das pessoas e encontrar meios para atendê-las, buscam informações para enfrentar desafios e obter vantagens competitivas, sabem o que querem alcançar e, para isso, fazem planejamentos com os quais monitoram e alteram suas atividades para obterem êxito, têm boa habilidade de relacionamento interpessoal: procuram manter contato constante com sua rede de relações, são motivadores, líderes e administradores de conflitos.
3. Nível 7 (15 itens): Os indivíduos posicionados nesse nível ou acima, além das características dos níveis anteriores, consideram-se independentes para tomar decisões sobre a sua vida e correr riscos calculados, são pró-ativos e capazes de filtrar as oportunidades que são lucrativas, têm objetivos claramente definidos e são bastante habilidosos em detalhar metas e os passos para alcançá-las, e sabem maneiras de convencer as pessoas a mudarem de opinião e não só mantêm uma rede de relacionamentos, mas a utilizam para alcançar seus objetivos.

Alves e Bornia (2011) não interpretaram os níveis âncoras 4, 8 e 9, com 2, 1 e 1 itens respectivamente, argumentando que a pouca quantidade de itens não iria fornecer uma boa interpretação. Os resultados mostraram que a escala é mais adequada para avaliar indivíduos no intervalo de baixo a moderadamente alto potencial empreendedor. Dessa forma, os autores sugerem que deve-se criar, calibrar e adicionar itens com parâmetros de dificuldade mais elevados para interpretar para avaliar com boa precisão níveis mais elevados.

Vasconcelos, Lezana e Andrade (2013) propuseram uma escala para medir o sucesso empresarial

de Micro e Pequenas Empresas MPE, utilizando o MRG em uma escala com média 70 e desvio padrão 10. O questionário, composto por 30 itens, utilizou uma escala likert de 7 pontos e foi aplicado a 368 MPEs de Santa Catarina. Foram identificados seis níveis âncoras, descritos resumidamente:

1. Nível 50 – Empresas nesse nível são desorganizadas, não fazem planejamento e o proprietário não tem conhecimento sobre o mercado e o produto e tem pouca ou nenhuma experiência sobre empreendedorismo.
2. Nível 60 – Nesse nível, a empresa apresenta melhorias na organização e tem conhecimento sobre o mercado que atua, mas precisa melhorar em vários aspectos.
3. Nível 70 – A empresa apresenta melhorias na organização e tem conhecimento sobre o mercado que atua, os produtos/serviços são adequados às necessidades do mercado.
4. Nível 80 – A empresa tem boa organização, o proprietário tem conhecimento sobre o mercado que atua, os produtos/serviços são adequados às necessidades do mercado, e são feitos planejamentos.
5. Nível 90 – A empresa tem boa organização, o proprietário tem conhecimento sobre o mercado que atua e experiência em empreendedorismo, os produtos/serviços são adequados às necessidades do mercado, são feitos planejamentos e os relacionamentos com os sócios são bons.
6. Nível 100 – A empresa tem um bom ou excelente desempenho em todos os aspectos.

Almeida (2009) elaborou uma medida de desempenho ambiental dos Estabelecimentos de Saúde (MDAES), quanto ao gerenciamento dos resíduos dos serviços de saúde (RSS). O questionário foi composto por 63 itens e aplicado em 728 Estabelecimentos de Saúde (ES) em todos os Estados do Brasil. A escala com média 500 e desvio padrão 50 foi criada com base no ML2, onde foram identificados 5 níveis âncoras e 37 itens âncoras e quase âncoras:

1. Nível 450 (3 itens): Os ES orientam os funcionários envolvidos com o manejo de resíduos, quanto aos procedimentos adequados de limpeza no caso de um vazamento de resíduos, oferecendo os EPIs (Equipamentos de Proteção Individual) necessários para a realização dos procedimentos, além de possuírem um ambiente físico adequado à guarda dos resíduos e procedimentos periódicos de manutenção de seus reservatórios de água.

2. Nível 500 (14 itens): Os ES orientam seus funcionários, envolvidos com o manejo de resíduos, quanto ao procedimento de segregação na fonte, utilizando horários adequados para o transporte dos mesmos dentro do estabelecimento, armazenando seus resíduos em local apropriado, oferecendo segurança física do ambiente e que este seja frequentado somente por pessoas autorizadas. O ES realiza higienização adequada após cada coleta e realiza manutenções periódicas na estrutura física do local. O ES também realiza a manutenção periódica de todos os sistemas (hidráulico, elétrico, ar condicionado) possuem programas de treinamentos periódicos para os funcionários, sobre o manejo dos RSS e riscos relacionados a estes.
3. Nível 550 (12 itens): ES possui procedimentos documentados, inclusive dos treinamentos realizados com os funcionários envolvidos com o manejo dos RSS, identifica todos os acondicionadores de resíduos, de forma clara e apresenta locais apropriados para a sua guarda temporária, além de possuir um armazenamento externo com boxes distintos para cada tipo de resíduos. O ES também possui programas de prevenção de riscos ambientais (biossegurança, PPRA - Programa de Prevenção de Riscos Ambientais, PCMSO - Programa de Controle Médico Ocupacional, PMOC - Plano de Manutenção de Operacionalização e Controle, etc.) e um sistema de gerenciamento dos seus resíduos, com uma equipe responsável pelas questões de saúde e segurança ocupacional.
4. Nível 600 (4 itens): O desempenho ambiental apresentado pelos ES é bastante próximo do que a legislação determina. Eles possuem mapas de riscos, elaborados e divulgados a todos os funcionários e plano de emergência no caso de acidentes com substâncias perigosas. Além disso, os ES possuem um PGRSS (Plano de Gerenciamento de Resíduos de Serviços de Saúde) implementado e um programa de qualidade em implantação.
5. Nível 650 (4 itens): Os ES que possuem um sistema de gerenciamento dos RSS bastante avançado, a preocupação desses estabelecimentos vai além do gerenciamento interno, se preocupando com a forma de retorno dos efluentes líquidos para o ambiente e o reaproveitamento dos resíduos, além da reciclagem dos mesmos. Os ES também

apresentam procedimentos que vão além do que a legislação determina, com um PGRSS que contempla procedimentos vinculados à sustentabilidade.

Bornia *et al.* (2009) avaliou a satisfação dos congressistas em relação aos Congressos Brasileiros de Custos (CBB) dos anos de 2006 a 2008, por meio de um questionário composto por 12 itens. Foi utilizado o ML2 da TRI e construída uma escala de satisfação com dois níveis âncoras, em que os indivíduos classificados no nível 60 estão altamente satisfeitos com todos os itens e os indivíduos classificados no nível 50 estão satisfeitos com os 7 itens e avaliam de forma positiva os demais itens.

Bortolotti (2003) aplicou um questionário de avaliação de ensino da UNIOESTE – Cascavel – PR, com 35 itens com seis categorias de satisfação (totalmente insatisfeito, insatisfeito, pouco insatisfeito, pouco satisfeito, satisfeito, totalmente satisfeito) a cerca de 500 alunos dos cursos de Tecnologia do Centro Federal de Educação Tecnológica de Medianeira, PR (CEFET – MD – PR). Os dados foram avaliados segundo o modelo GGUM, o qual foi utilizado para a criação da escala de satisfação, no entanto, essa escala não foi caracterizada com níveis e itens âncoras (Bortolotti; Andrade, 2007; Bortolotti; Souza Junior; Andrade, 2009). Os mesmos dados foram utilizados em análises posteriores com outros modelos da TRI. O ML2 foi utilizado (Bortolotti; Moreira Junior; Souza Junior, 2010) para criar uma escala de satisfação com média 100 e desvio padrão 10, onde foram identificados três níveis âncoras e 20 itens âncoras:

1. Nível 100 (4 itens) – Caracteriza-se por identificar alunos que possuem satisfação com relação ao curso, ou seja, quanto ao conhecimento e seriedade acadêmica por parte dos docentes e quanto à qualidade do mesmo.
2. Nível 110 (12 itens) – Caracteriza-se por identificar alunos que consideram satisfeitos quanto ao seu desempenho, em relação às notas obtidas, à seriedade acadêmica, à qualificação para a elaboração de trabalhos, quanto à iniciativa dos alunos para a complementação de sua formação acadêmica e também ao nível de formação obtido. Esses alunos também estão satisfeitos com relação ao desempenho dos docentes no ensino, quanto à metodologia utilizada por estes nas aulas, ao seu comprometimento na qualificação do curso, à orientação e ao atendimento aos alunos, aos procedimentos para qualificar os conteúdos desenvolvidos e oferecer oportunidades para treinamento e inserção no mercado de trabalho.

Além disso, esses alunos também possuem as características dos itens anteriores.

3. Nível 120 (4 itens) – Caracteriza-se por identificar alunos que estão satisfeitos quanto à qualidade do ensino, às medidas adotadas para aprimorar a metodologia das aulas nas disciplinas do curso, à capacidade manifestada pelos alunos para a leitura de textos científicos durante o curso de graduação e quanto às medidas adotadas para superar as dificuldades dos alunos com deficiências nas disciplinas. Além disso, esses alunos também possuem as características dos itens anteriores.

Os mesmos dados foram utilizados para criar outra escala de satisfação com a utilização do MRC da TRI (Bortolotti *et al.*, 2012; Moreira Junior *et al.*, 2010), com média 100 e desvio padrão 10, em que foram identificados seis níveis âncoras e 27 categorias âncoras:

1. Nível 80 (4 categorias âncoras) – Caracteriza-se por identificar respondentes que estão pouco insatisfeitos com a organização na exposição de conteúdos, com as formas de avaliação utilizadas para avaliar os níveis de aprendizagem, com a seriedade demonstrada pelos docentes e dos alunos, com a qualidade do curso e com o nível de formação dos alunos que concluem o curso.
2. Nível 90 (8 categorias âncoras) – Além das características anteriores, este nível identifica os respondentes que estão pouco insatisfeitos com a capacidade e qualificação manifestada pelos alunos para leitura e escrita de textos científicos, com a oportunidade de iniciação dos alunos na pesquisa, com as medidas adotadas para aprimorar a metodologia das aulas e avaliação dos alunos nas disciplinas, com os procedimentos adotados pelo curso para a qualificação dos conteúdos desenvolvidos nas disciplinas, com a dinâmica das aulas para manter a atenção dos alunos, com os mecanismos de atendimento e orientação acadêmica dos alunos no dia a dia dos cursos, com o comprometimento efetivo dos docentes com a qualificação do curso, do conhecimento da situação dos alunos que concluíram o curso, com o nível de formação dos alunos que ingressam no curso e com a oportunidade de treinamento e inserção no mercado de trabalho oferecido pelo curso.
3. Nível 100 (2 categorias âncoras) – Caracteriza-se por identificar respondentes que estão satisfeitos com relação ao nível de for-

A UTILIZAÇÃO DA TEORIA DA RESPOSTA AO ITEM COMO FERRAMENTA PARA
A CONSTRUÇÃO DE ESCALAS NA ÁREA DE GESTÃO

mação atingido pelos alunos que concluem o curso e com a qualidade do curso e com a dinâmica das aulas para manter a atenção dos alunos, além das características dos níveis anteriores .

4. Nível 110 (8 categorias âncoras) – Caracteriza-se por identificar respondentes que estão satisfeitos quanto à organização de conteúdos pelos docentes, com a qualidade dos planos de ensino, bem como com a estrutura curricular do curso, com a criatividade demonstrada, oportunidade de iniciação dos alunos na pesquisa, com as condições para a dedicação ao curso, com o atendimento da orientação acadêmica, com o comprometimento dos docentes e conhecimento deles quanto à qualificação do curso, além das características dos níveis anteriores .
5. Nível 120 (2 categorias âncoras) – Caracteriza-se por identificar respondentes que estão satisfeitos com as medidas para superar as dificuldades dos alunos com deficiências nas disciplinas, com as medidas adotadas para aprimorar a metodologia das aulas, bem como a avaliação e com a capacidade dos alunos para a leitura e escrita de traba-

lhos científicos, além das características dos níveis anteriores .

6. Nível 130 (3 categorias âncoras) – Caracteriza-se por identificar respondentes que estão muito satisfeitos quanto à qualidade do curso que realizam, bem como com as condições dos alunos quanto à dedicação ao curso e com o comprometimento dos docentes com a qualificação do curso e dinâmica das aulas para manter a atenção dos alunos, além das características dos níveis anteriores.

Observa-se que o MRG apresentou o dobro de níveis âncoras em relação ao ML2. Também percebe-se que no MRG foram identificadas categorias âncoras, enquanto que no ML2 foram identificados itens âncoras. A vantagem que se observa na utilização do MRG é que é possível caracterizar não só os indivíduos satisfeitos, mas também os insatisfeitos, enquanto que o ML2 só permite caracterizar aqueles que estão satisfeitos.

Essa seção apresentou uma revisão sobre alguns resultados obtidos por meio da aplicação da Teoria da Resposta ao Item como método para a construção de escalas na área de Gestão no Brasil. O Quadro 1 apresenta uma síntese dos estudos revisados, destacando as suas características.

Quadro 1 – Síntese dos estudos

AUTORES	ESCALA	MÉTRICA (Média; Desvio Padrão)	MRI	NÍVEIS E ITENS E/OU CATEGORIAS ÂNCORAS
Alexandre et al., (2003a)	Nível de Práticas da GQT	(0; 1)	ML2	Nível -1: 3 itens Nível 0,5: 17 itens Nível 2: 12 itens
Oliveira (2010)	Escala do Grau de Maturidade (EGM)	(100; 25)	MRG	Nível 50: 2 itens Nível 75: 6 itens Nível 100: 2 itens Nível 125: 10 itens
Milnitz e Tubino (2012)	Escala para o Nível da Prática de Melhoria Contínua	(100; 10)	MRN	N1 (104): 3 itens N2 (106): 4 itens N3 (108): 7 itens N4 (110): 5 itens N5 (112): 1 item
Castro Junior (2007)	Percepção dos funcionários sobre a maturação das empresas de tecnologia	Não informado	ML1	Não informado
Roriz Junior (2008)	Nível da maturidade das Indústrias para a implantação do comércio eletrônico B2B.	Não informado	ML1	Não informado
Mafra (2011)	Escalas de Medida de Avaliação de Risco de Crédito (MRISCd)	(50; 10)	ML2	Nível 60: 3 itens Nível 70: 3 itens Nível 80: 2 itens Nível 90: 1 item
Mafra (2011)	Escalas de Medida de Avaliação de Risco de Crédito (MRISCp)	(50; 10)	MRG	Nível 30: 1 item e 1 categoria Nível 40: 2 itens e 2 categorias Nível 50: 2 itens e 2 categorias Nível 60: 3 itens e 5 categorias Nível 70: 4 itens e 4 categorias Nível 80: 1 item e 1 categoria

Vargas (2007) e Vargas et al. (2008)	Escala de Medida do Desempenho nos Intangíveis (EMDI)	(500; 50)	ML2	Nível 500: 1 item Nível 550: 4 itens Nível 600: 8 item Nível 650: 1 item
Moreira Junior, Vargas e Andrade (2010)	Escala de Medida do Desempenho nos Intangíveis (EMDI)	(0; 1)	ML2	Nível -1: 1 item Nível 0: 1 item Nível 1: 11 itens Nível 2: 2 itens Nível 3: 1 item
Alves e Bornia (2011)	Escala para o Potencial Empreendedor	(8; 2)	ML2	Nível 4: 2 itens Nível 5: 7 itens Nível 6: 17 itens Nível 7: 15 itens Nível 8: 1 item Nível 9: 1 item
Vasconcelos, Lezana e Andrade (2013)	Escala de sucesso empresarial de MPE	(70; 10)	MRC	Nível 50 Nível 60 Nível 70 Nível 80 Nível 90 Nível 100
Almeida (2009)	Medida de Desempenho Ambiental dos Estabelecimentos de Saúde (MDAES)	(500; 50)	ML2	Nível 450: 3 itens Nível 500: 14 itens Nível 550: 12 itens Nível 600: 4 itens Nível 650: 4 itens
Bornia et al. (2009)	Satisfação dos congressistas	(50; 10)	ML2	Nível 50: 7 itens Nível 60: 5 itens
Bortolotti, Moreira Junior e Souza Junior (2010)	Nível de satisfação de alunos	(100; 10)	ML2	Nível 100: 4 itens Nível 110: 12 itens Nível 120: 4 itens
Bortolotti et al. (2012) e Moreira Junior et al. (2010)	Nível de satisfação de alunos	(100; 10)	MRC	Nível 80: 4 categorias Nível 90: 8 categorias Nível 100: 2 categorias Nível 110: 8 categorias Nível 120: 2 categorias Nível 130: 3 categorias

Fonte: elaborado pelo autor.

4 CONSIDERAÇÕES FINAIS

Esse artigo apresentou uma discussão sobre como a Teoria da Resposta ao Item pode contribuir como método para a construção e interpretação de escalas por meio de aplicações em trabalhos realizados no Brasil na área de Gestão. A interpretação dessa escala se beneficia pelo fato da TRI posicionar tantos os itens quanto os respondentes na mesma escala, permitindo verificar a relação de dominância entre eles.

Os estudos revisados apontam que têm sido utilizados apenas os modelos unidimensionais da TRI, ou seja, aqueles que avaliam um único traço latente, com destaque para o ML1, o ML2, o MRC e o MRN. Esses modelos também se caracterizam por serem acumulativos, isto é, há uma relação de dominância entre itens e sujeitos, de tal forma que se um indivíduo (ou empresa) domina certo item, conseqüentemente, ele domina (ou possui) todos os itens que estão posicionados abaixo desse item

na escala, mas não domina (ou não possui) aqueles itens que estão posicionados acima dele. Os modelos dicotômicos, representados pelo ML1 e pelo ML2, possuem regras claras, descritas na literatura, para a elaboração da escala e para a definição dos níveis e dos itens âncoras. No caso dos modelos politômicos, representados nesse estudo pelo MRC e pelo MRN, não há regras claras para a construção das categorias âncoras, mas há algumas propostas, como foi mencionado nesse artigo. Nesse caso, verifica-se se a categoria é âncora, e não se o item é âncora, embora alguns autores tenham interpretado os itens como sendo âncora. Na interpretação da categoria âncora, encontra-se uma dificuldade nas escalas acumulativas, por exemplo, ao mencionar em um nível da escala que as empresas (ou sujeitos) ali posicionadas possuem determinada característica e todas as características mencionadas nos níveis anteriores, essa afirmação torna-se inconsistente se, a tal característica era ausente em algum nível anterior. No caso dos mo-

delos politômicos, deve-se utilizar com cuidado a afirmação “possui todas as características mencionadas nos níveis anteriores”, pois aquela empresa que não possuía (ou possuía pouco) determinada característica num nível anterior, pode ter mudado de comportamento e passou a possuir aquela característica em um nível superior, ou seja, ela não acumulou o fato de não possuir (ou possuir pouco) aquela característica.

Alguns estudos não mostraram os níveis e itens âncoras das escalas construídas. Provavelmente a escala não foi elaborada nesses estudos, embora pudesse ter sido. Alguns estudos também mostraram-se inconsistentes por utilizar uma amostra muito pequena para a calibração dos itens, fornecendo resultados que não possuem credibilidade suficiente para validar as escalas propostas por eles. A maioria dos trabalhos utilizou escalas diferentes do padrão dos softwares, ou seja, da escala com média zero e desvio padrão um, evitando, assim, a construção de níveis negativos, que pode dar uma impressão estranha ao resultado do trabalho, podendo gerar interpretações equivocadas.

Os resultados demonstraram a potencialidade da Teoria da Resposta ao Item como um método válido para a construção e interpretação de escalas, em diversas áreas de Gestão. Verifica-se que ainda existe pouca aplicação da TRI na área de Gestão no país, em comparação com outras metodologias empregadas, todavia, essa metodologia deve ganhar espaço, em relação às metodologias tradicionais, principalmente pelas vantagens apresentadas, tais como a elaboração de uma escala em que respondentes e itens são posicionados simultaneamente, e a possibilidade de acompanhamento longitudinal, ao longo do tempo, do traço latente avaliado. Na área de Gestão, o acompanhamento longitudinal de empresas (ou indivíduos) e sua evolução na escala ainda é um desafio, uma vez que o estudo mais antigo encontrado foi de 2003 (vide Quadro 1), ou seja, é recente, não permitindo ainda uma análise temporal apropriada. Outro desafio está relacionado com a invariância dos parâmetros dos itens, ou seja, uma vez calibrados os itens e construída a escala, com o passar do tempo, deve-se verificar se os parâmetros dos itens não irão se alterar ao longo do tempo, devido à obsolescência do item, à mudanças nos conceitos de gestão e qualidade ou ao surgimento de novas tecnologias ou novas formas de avaliar a qualidade. A manutenção e atualização da escala devem ser levadas em conta para o seu funcionamento adequado.

REFERÊNCIAS

ALEXANDRE, J. W. C.; ANDRADE, D. F.; VASCONCELOS, A. P.; ARAUJO, M. A. S. Aplicação da Teoria da Resposta ao Item na Gestão da Qualidade: Proposta de um Modelo Probabilístico. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 20, 2001, Salvador. **Anais...** Salvador: ABEPRO, 2001.

ALEXANDRE, J. W. C.; ANDRADE, D. F.; VASCONCELOS, A. P.; ARAUJO, A. M. S. Uma proposta de análise de um construto para a medição dos fatores críticos da gestão pela qualidade através da teoria da resposta ao item. **Gestão & Produção**, v.9, n.2, p.129-141, 2002a.

ALEXANDRE, J. W. C.; ANDRADE, D. F.; VASCONCELOS, A. P.; ARAUJO, A. M. S.; BATISTA, M. J. Teoria da Resposta ao Item: Aplicação do Modelo de Escala Gradual na Gestão pela Qualidade. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 22, 2002b, Curitiba. **Anais...** Curitiba: ABEPRO, 2002b.

ALEXANDRE, J. W. C.; ANDRADE, D. F.; VASCONCELOS, A. P.; ARAUJO, A. M. S.; BATISTA, M. J. Uma Proposta de Análise da Maturidade Organizacional na GQT via Teoria da Resposta ao Item. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 35, 2003, Natal. **Anais...** Rio de Janeiro: SOBRAPO, 2003a. v. 1. p. 1-20.

ALEXANDRE, J. W. C.; ANDRADE, D. F.; VASCONCELOS, A. P.; ARAUJO, A. M. S.; BATISTA, M. J. Análise do Número de Categorias da Escala de Likert Aplicada à Gestão pela Qualidade Total Através da Teoria da Resposta ao Item. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 23, 2003, Ouro Preto. **Anais...** Santa Bárbara d'Oeste - SP: ABEPRO, 2003b. v. 1. p. 1-20.

ALMEIDA, V. L. **Avaliação do desempenho ambiental de estabelecimentos de saúde, por meio da Teoria da Resposta ao Item, como incremento da criação do conhecimento organizacional.** 2009. 189 f. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2009.

ALVES, L. R. R.; BORNIA, A. C. Desenvolvimento de uma escala para medir o potencial empreendedor utilizando a Teoria da Resposta ao Item (TRI). **Gestão e Produção**, São Carlos, v. 18, n. 4, 2011.

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da resposta ao item: conceitos e aplicações.** São Paulo: ABE - Associação Brasileira de Estatística, 2000.

ANDRICH, D. A rating formulation for ordered response categories. **Psychometrika**, n. 43, p. 561-573, 1978.

BAKER, F. B. *The Basics of Item Response Theory.* 2 ed. USA: ERIC Clearinghouse on Assessment and Evaluation, 2001.

- BATISTA, M. J.; ALEXANDRE, J. W. C. Teoria da resposta ao item: proposta de análise na gestão pela qualidade total. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 16, 2004, Caxambu - MG. **Resumos...** São Paulo: ABE - Associação Brasileira de Estatística, 2004.
- BATISTA, M. J.; VASCONCELOS, A. P.; ALEXANDRE, J. W. C.; ANDRADE, D. F.; ARAUJO, A. M. S. Teoria Clássica de Medida e Teoria da Resposta ao Item: uma abordagem comparativa na gestão da qualidade. In: REUNIÃO REGIONAL DA ASSOCIAÇÃO BRASILEIRA DE ESTATÍSTICA, 34, 2002, Fortaleza - CE. **Resumos.** São Paulo: ABE - Associação Brasileira de Estatística, 2002, p. 14.
- BEATON, A. E.; ALLEN, N. L. Interpreting Scales through Scale Anchoring. **Journal of Educational Statistics**, n. 17, p. 191-204, 1992.
- BIRNBAUM, A. **Some Latent Trait Models and Their Use in Inferring an Examinee's Ability.** In: LORD, F. M.; NOVICK, M. R. *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wesley, 1968.
- BOCK, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. **Psychometrika**, v. 37, p. 29-51, 1972.
- BORNIA, A. C.; ANDRADE, D. F.; POSSAMAI, O.; MAFRA, P. M. R.; ALMEIDA, V. L. Satisfação do congressista em relação ao Congresso Brasileiro de Custos por meio da teoria da resposta ao item. In: CONGRESSO BRASILEIRO DE CUSTOS, 16, 2009, Fortaleza. **Anais...** Fortaleza, 2009.
- BORTOLOTTI, S. L. V. Aplicação de um modelo de desdobramento graduado generalizado da teoria da resposta ao item – TRI. 2003. 107 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2003.
- BORTOLOTTI, S. L. V.; ANDRADE, D. F. **Aplicação de um modelo de desdobramento graduado generalizado – GGUM da teoria da resposta ao item.** Estudos em Avaliação Educacional, São Paulo, v. 18, n. 37, p. 157-188, 2007.
- BORTOLOTTI, S. L. V.; MOREIRA JUNIOR, F. J.; SOUZA JUNIOR, A. F.; ANDRADE, D. F. Proposta de Avaliação da Satisfação por meio do Modelo Logístico de Dois Parâmetros da Teoria da Resposta ao Item. In: CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 6, 2010, Rio de Janeiro. **Anais...** Niterói, RJ: UFF - Universidade Federal Fluminense, 2010.
- BORTOLOTTI, S. L. V.; SOUZA JUNIOR, A. F.; ANDRADE, D. F. Uma Metodologia para Avaliação da Satisfação através da Teoria da Resposta Ao Item - TRI. In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA, 6, 2009, Resende - RJ. **Anais...** Resende - RJ: AEDB - Associação Educacional Dom Bosco, 2009.
- 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 & Produção**, São Carlos, v. 19, n. 2, p. 287-302, 2012.
- CASTRO JUNIOR, J. S. **Método de Avaliação de Maturidade para a Implantação de Sistemas de Informação Estratégica em Empresas de tecnologia da Informação e Comunicação.** 2007, 103 f. Dissertação (Mestrado em Engenharia Elétrica). Programa de Pós-Graduação em Engenharia Elétrica. Departamento de Engenharia Elétrica, Faculdade de Tecnologia, Universidade de Brasília. Brasília, 2007.
- COSTA, M. B. F.; CHAVES NETO, A. Aplicação da Teoria da Resposta ao Item (TRI) na avaliação do grau de satisfação do consumidor em um serviço específico. In: REUNIÃO REGIONAL DA ASSOCIAÇÃO BRASILEIRA DE ESTATÍSTICA, 34, 2002, Fortaleza - CE. **Resumos.** São Paulo: ABE - Associação Brasileira de Estatística, 2002, p. 31-31.
- CUNHA, R. C. L. V.; SENA JUNIOR, M. R.; MATOS, G. S. Medindo Satisfação do Cliente através da Teoria da Resposta ao item. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 15, 2002, Águas de Lindóia - SP. **Resumos...** São Paulo: ABE - Associação Brasileira de Estatística, 2002.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm (with discussion). **Journal of the Royal Statistical Society, Series B**, 39, 1-38, 1977.
- EMBRETSON, S. E.; REISE, S.P. **Item Response Theory for Psychologists.** New Jersey, USA: Lawrence Erlbaum Associates, 2000.
- FONTANIVE, N. S.; ELLIOT, L. G.; KLEIN, R. Os desafios da apresentação dos resultados da avaliação de sistemas escolares a diferentes públicos. **REICE - Revista Eletrônica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 5, n. 2e, 2007.
- KLEIN, R.; FONTANIVE, N. S.; ELLIOT, L. G. O Exame Nacional do Ensino Médio – Tecnologia e Principais Resultados Em 2005. **REICE - Revista Eletrônica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 5, n. 2e, p. 116-131, 2007.
- KOLEN, M. J.; BRENNAN, R. L. **Test Equating - Methods and Practices.** New York: Springer, 1995.
- ISSAC, E; KELLER, H. B. **Analysis of Numerical Methods.** New York: Wiley & Sons, 1966.
- LORD, F. M. **Applications of Item Response Theory to Practical Testing Problems.** Hillsdale, USA: Lawrence Erlbaum, 1980.
- MAFRA, P. M. R. **Proposta de uma sistemática para a modelagem de risco de crédito sob a perspectiva da teoria da criação do conhecimento: uma abordagem**

- da teoria da resposta ao item. 2010. Tese (Doutorado em Engenharia e Gestão do Conhecimento)-Universidade Federal de Santa Catarina, Florianópolis, 2010.
- MASTERS, G. N. A Rasch model for partial credit scoring. **Psychometrika**, n. 47, p. 149-174, 1982.
- MILNITZ, D. ; TUBINO, D. F. . Teoria de Resposta ao Item: medida do nível de prática da melhoria contínua por meio do Modelo de Resposta Nominal. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO DA REGIÃO NORDESTE, 7, 2012, Mossoró. **Anais...** Mossoró, 2012.
- MORAIS, J. H. M. Estudo de validação: padrões e dimensões da organização. In: CONGRESSO BRASILEIRO DE AVALIAÇÃO PSICOLÓGICA, 4, 2009, Campinas. **Resumos - Painéis**. Campinas: IBAP - Instituto Brasileiro de Avaliação Psicológica, 2009.
- MOREIRA JUNIOR, F. J. Aplicações da Teoria da Resposta ao Item (TRI) no Brasil. **Revista Brasileira de Biometria**, São Paulo, v.28, n.4, p. 137-170, out./dez. 2010.
- MOREIRA JUNIOR, F. J. **Sistemática para a Implantação de Testes Adaptativos Informatizados baseados na Teoria da Resposta ao Item**. 2011. 334 f. Tese (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2011.
- MOREIRA JUNIOR, F. J.; BORTOLOTTI, S. L. V.; ANDRADE, D. F.; SOUZA JUNIOR, A. F. Avaliação da Satisfação através da utilização da Teoria da Resposta ao Item (TRI). In: X SEPROSUL - Semana de la Ingeniería de la Producción Sudamericana, 2010, **Anais...** Santiago, Chile, 2010.
- MOREIRA JUNIOR, F. J.; VARGAS, V. C. C.; ANDRADE, D. F. Avaliação dos Intangíveis em Indústrias Gaúchas e Catarinenses. In: SEMANA DE LA ENGENIERÍA DE LA PRODUCCIÓN SUDAMERICANA, 10, 2010, **Anais...** Santiago, Chile, 2010.
- MURAKI, E. A generalized partial credit model: Application of an EM algorithm. **Applied Psychological Measurement**, n. 16, p. 159-176, 1992.
- NUNES, C. H. S. S.; PRIMI, R. Impacto do tamanho da amostra na calibração de itens e estimativa de escores por teoria de resposta ao item. **Aval. psicol.** v.4 n.2 Porto Alegre nov. 2005.
- OLIVEIRA, K. M. M. **Um Estudo da Evolução da GQT por meio do Modelo de Resposta Gradual da Teoria da Resposta ao Item**. 2010. Dissertação (Mestrado em Gestão Logística e Pesquisa Operacional)-Universidade Federal do Ceará, Fortaleza, 2010.
- PARTCHEV, I. **Irtoys: Simple interface to the estimation and plotting of IRT models**. R package version 0.1.6, 2013.
- PEREIRA, L. M. **Um método para medir o grau da maturidade das empresas para implantar o comércio eletrônico B2B através da Teoria da Resposta ao Item**. 2007. Dissertação (Mestrado em Engenharia Elétrica e de Computação). Programa de Pós- Graduação em Engenharia Elétrica e de Computação, Escola de Engenharia Elétrica, Universidade Federal de Goiás, Goiânia, 2007.
- RAO, C. R. **Linear Statistical Inference and Its Applications**. New York: Wiley & Sons, 1973.
- RASCH, G. **Probabilistic Models for Some Intelligence and Attainment Tests**. Copenhagen: Danish Institute for Educational Research, 1960.
- RECKASE, M. D. **Multidimensional Item Response Theory. Statistical for Social and Behavioral Sciences**. New York: Springer Science+Business Media, LLC, 2009.
- RIZOPOULOS, D. **lrm: Latent Trait Models under IRT**. R package version 0.9-9, 2013.
- ROBERTS, J. S.; DONOGHUE, J. R.; LAUGHLIN, J. E. A General model for unfolding unidimensional polytomous responses using item response theory. **Applied Psychological Measurement**, v. 1, n. 24, p. 3-32, 2000.
- RORIZ JUNIOR, G. S. **Da Maturidade da Gestão Industrial para o Comércio B2B em Empresas Farmacêuticas de Goiás**. 2008. 134 f. Dissertação (Mestrado em Tecnologia Farmacêutica) - Programa de Pós-Graduação em Gestão, Pesquisa e Desenvolvimento em Tecnologia Farmacêutica, Universidade Católica de Goiás, Universidade Estadual de Goiás, Centro Universitário de Anápolis, Goiânia, 2008.
- SAMEJIMA, F. A. Estimation of latent ability using a response pattern of graded scores. **Psychometric Monograph**, n. 17, 1969.
- SANTOS, P. S.; MENEZES, I. G.; BASTOS, A. V. B.; ALVES FILHO, A. P.; MENDONÇA, R. F. Comprometimento organizacional: um estudo empírico sobre a dimensionalidade do construto. In: CONGRESSO BRASILEIRO DE AVALIAÇÃO PSICOLÓGICA, 4, 2009, Campinas. **Resumos - Painéis**. Campinas: IBAP - Instituto Brasileiro de Avaliação Psicológica, 2009.
- SOUZA, S. Z. 40 Anos de Contribuição à Avaliação Educacional. **Estudos em Avaliação Educacional**, v. 16, n. 31, jan./jun. 2005.
- TOIT, M. **IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT**. Scientific Software International, 2003.
- 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**, n. 23, p. 71-92, 2001.

VARGAS, V. C. C. **Medida Padronizada para Avaliação de Intangíveis Organizacionais por Meio da Teoria da Resposta ao Item**. 2007. 207 f. Tese (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2007.

VARGAS, V. C. C.; SELIG, P. M.; ANDRADE, D. F.; RIBEIRO, J. L. D. Avaliação dos intangíveis: uma aplicação em capital humano. **Gestão & Produção**, São Carlos, v. 15, n. 3, p. 619-634, set./dez. 2008.

VASCONCELOS, A. M.; LEZANA, A. G. R.; ANDRADE, D. F. Escala para medição do sucesso de micro e pequenas empresas de Santa Catarina com o apoio da teoria da resposta ao item. **Revista Brasileira de Gestão e Desenvolvimento Regional**, Taubaté, v. 9, n. 1, p. 91-119, jan./mar. 2013.

VASCONCELOS, A. P.; BATISTA, M. J.; ALEXANDRE, J. W. C.; ANDRADE, D. F.; ARAUJO, A. M. S. Teoria da Reposta ao Item Aplicada à Gestão pela Qualidade Total: proposta do modelo de escala gradual. In: REUNIÃO REGIONAL DA ASSOCIAÇÃO BRASILEIRA DE ESTATÍSTICA, 34, 2002, Fortaleza - CE. **Resumos**. São Paulo: ABE - Associação Brasileira de Estatística, 2002.

VENDRAMINI, C. M. M.; SILVA, M. C.; CANEL, M. Análise de Itens de uma Prova de Raciocínio Estatístico. **Psicologia em Estudo**, Maringá, v. 9, n. 3, p. 487-498, set./dez. 2004.

WEEKS, J. P. **plink: IRT Separate Calibration Linking Method**. R package version 1.3-1, 2013.

WEISS, D. J.; GUYER, R. Manual for CATSim: Comprehensive simulation of computerized adaptive testing. St. Paul MN: Assessment Systems Corporation, 2010.

WRIGHT, B. D. **Sample-free test calibration and person measurement. Proceedings of the 1967 Invitational Conference on Testing Problems**. Princeton, N. J.: ETS - Educational Testing Service, 1968.