



## UMA PROPOSTA DE ANÁLISE DE UM CONSTRUTO PARA MEDIÇÃO DOS FATORES CRÍTICOS DA GESTÃO PELA QUALIDADE POR INTERMÉDIO DA TEORIA DA RESPOSTA AO ITEM

**João Welliandre Carneiro Alexandre**  
**Dalton Francisco de Andrade**  
**Alan Pereira de Vasconcelos**  
**Ana Maria Souza de Araujo**

Departamento de Estatística e Matemática Aplicada,  
Universidade Federal do Ceará,  
Campus do Pici, Bloco 910, CEP 60455-760, Fortaleza, CE,  
e-mails: [jwca@ufc.br](mailto:jwca@ufc.br), [dandrade@ufc.br](mailto:dandrade@ufc.br),  
[alan@ufc.br](mailto:alan@ufc.br), [anamaria@ufc.br](mailto:anamaria@ufc.br)

### **Resumo**

*Neste artigo propomos o uso de modelos da Teoria da Resposta ao Item (TRI) na análise de construtos elaborados para medir a Gestão pela Qualidade Total (GQT) como uma alternativa à Teoria Clássica de Medida (TCM). São apresentados o modelo geral para itens dicotomizados assim como as interpretações dos parâmetros do modelo. Os resultados mostram que a TRI pode ser uma poderosa ferramenta na análise das práticas da GQT e da maturidade organizacional, dentro da filosofia da qualidade.*

**Palavras-chave:** *gestão pela qualidade total, teoria da resposta ao item, análises dos construtos da GQT.*

### **1. Introdução**

**P**esquisas organizacionais para avaliação dos fatores críticos de sucesso na implantação da Gestão pela Qualidade Total (GQT) foram introduzidas por Saraph *et al.* (1989). Esse trabalho, considerado pioneiro nessa linha de pesquisa, sugere um construto para modelagem dos fatores críticos da GQT, como também

apresenta um conjunto de práticas, isto é, medidas operacionais da qualidade para medição dos construtos. Esse estudo foi corroborado por Badri *et al.* (1995), que, por intermédio de sua replicação, confirmaram a confiabilidade e a validade desse instrumento. Surgiram, então, diversos outros trabalhos nessa linha: Tamimi (1995), Tamimi *et al.* (1995), Tamimi & Gershon (1995), Ahire (1996), Huq & Stolen (1998) e Alexandre (1999).

Nas pesquisas empíricas, em particular nas organizacionais, a confiabilidade das conclusões está diretamente relacionada à validação do instrumento de medição – questionário. Para a validação, medidas como a consistência interna das escalas (confiabilidade), a correlação de cada item com cada escala (análise detalhada do item) e se o item mede aquilo que pretende medir (validade) são geralmente aplicadas. Duas referências importantes nessa área são Cronbach (1951) e Nunnally (1967).

Contudo, outros questionamentos surgem nas pesquisas que visam investigar os fatores críticos de sucesso da GQT nas organizações:

1. Como modelar o grau de dificuldade de uma organização implantar uma prática da GQT?
2. Como modelar a discriminação entre as organizações quanto a seu grau de maturidade na implantação de uma prática da GQT?
3. Como representar o nível de maturidade organizacional quanto à GQT, por intermédio de uma escala de medida?
4. Como medir, por intermédio de um modelo, a maturidade organizacional da implantação das práticas da GQT ao longo do tempo?
5. Como comparar o desempenho na implantação das práticas entre organizações de setores diferentes a partir de um modelo?

Respostas a essas questões permitem efetiva comparação entre as organizações quanto à aplicação da GQT e também análise dos itens (práticas da qualidade) que compõem o instrumento de medida, pela estimação do grau de maturidade na GQT das organizações e dos parâmetros dos itens, em uma mesma métrica. A obtenção das respostas está associada à utilização de modelos que priorizam o item e não o instrumento de medida como um todo. Todavia, a literatura é carente de estudos desses modelos dentro da GQT.

A proposta deste artigo é contribuir nessa linha de pesquisa, sugerindo um modelo de

análise dos construtos da GQT com base de Teoria da Resposta ao Item (TRI), a qual tem aplicação bastante difundida na área educacional, com ênfase na definição e na interpretação dos parâmetros do modelo sugerido, no estudo de sua adequabilidade e na discussão de uma aplicação.

## 2. Teoria clássica de medida

Uma abordagem adequada quando se deseja verificar o grau (nível) de maturidade de uma organização na implantação de práticas da GQT é a aplicação de um questionário a ser respondido pelas organizações. Das respostas às perguntas ou sentenças obtém-se um escore que indica o nível de maturidade das organizações na implantação dessas práticas. De acordo com a Teoria Clássica de Medida (TCM), o escore observado é composto de um escore verdadeiro (o nível real de maturidade) e um erro de medida. Assume-se que o erro é aleatório. A equação básica da TCM, conhecida como modelo clássico de medida (Hayes, 1992), descreve a relação entre os escores observados, os escores verdadeiros e o erro:

$$X = T + E$$

em que: X é o escore observado; T, o escore verdadeiro; e E, o erro de medida.

Pequenos valores para o erro de medida E significa que os escores observados X são representativos em relação aos escores verdadeiros T.

Os erros podem estar associados aos seguintes fatores:

1. Impossibilidade de incluir um número infinito de perguntas no questionário.
2. Impossibilidade de aplicar o questionário um número infinito de vezes.
3. Impossibilidade de submeter os questionários a um número infinito de organizações.

Na TCM são avaliadas as propriedades psicométricas dos itens (questões), especificamente, a confiabilidade (consistência interna da escala), a análise do item e a validade do construto, conforme citado anteriormente. Contudo, uma característica dessa teoria consiste no

fato de a análise psicométrica do construto ter por ênfase o instrumento de medição como um todo. Para mais detalhes sobre essa teoria, veja Vianna (1978) e Gulliksen (1967), e para aplicação, Saraph *et al.* (1989) e Badri *et al.* (1995).

### 3. Teoria da resposta ao item aplicada à educação

Tradicionalmente, no processo de avaliação e seleção de indivíduos, são utilizados resultados obtidos em provas, expressos apenas por seus escores brutos ou padronizados. Isso significa que, por exemplo, quanto maior a nota do indivíduo em uma prova, melhor sua classificação. A característica desse procedimento consiste em que as análises e interpretações estão sempre associadas ao escore total e não a um item (questão) em particular. A Teoria da Resposta ao Item é um instrumento poderoso, que vem sendo progressivamente aplicado nos processos quantitativos na área de avaliação educacional. A TRI propõe modelos de variáveis latentes para representar a relação entre a probabilidade de um indivíduo apresentar determinada resposta a um item e seus traços latentes ou habilidades na área do conhecimento avaliada, os quais não podem ser observados diretamente.

Uma das grandes vantagens da TRI sobre a TCM é que aquela permite a comparação entre populações submetidas a provas diferentes, desde que essas provas tenham alguns itens comuns, ou, ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente diferentes. Isso porque uma das características da TRI é ter como elementos centrais os itens e não a prova como um todo. Como exemplo, vamos supor que desejamos comparar o nível de conhecimento entre alunos de duas séries distintas (3ª e 8ª séries do ensino fundamental, por exemplo). Na TCM essa comparação somente será possível caso seja aplicada a mesma prova para as duas turmas. Na TRI, são necessárias apenas

algumas questões em comum nas provas aplicadas às duas séries.

A TRI foi aplicada pela primeira vez no Brasil em 1995, na análise dos resultados do Sistema Nacional de Ensino Básico (SAEB). O Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP) tem sido planejado e implementado de forma a ser analisado pela TRI. Para outros detalhes sobre o modelo, interpretações e aplicações, veja Lord (1980), Andrade *et al.* (2000) e Andrade (2001).

A TRI tem despertado interesse de aplicação em diversas outras áreas: na área psicossocial, Granger *et al.* (1998) estudam o padrão de proficiência do profissional no cenário norte-americano; na área médica, DeRoos & Meares (1998) apresentam um estudo sobre as causas da depressão entre crianças americanas de origem africana e de cor branca; na área de marketing, Bayley (2001) investiga o nível de satisfação dos clientes de uma empresa pública australiana.

### 4. A TRI aplicada à GQT

O modelo apresentado a seguir é adequado para situações em que o construto é medido por meio de um questionário estruturado, isto é, objetivo, com respostas dicotomizadas, ou quando a análise é realizada de forma dicotomizada dentro de uma avaliação subjetiva – aberta. Como exemplo podem ser citadas: implantação da GQT satisfatória ou não satisfatória, implantação acima ou abaixo da média, implantação efetiva ou não efetiva etc.

#### 4.1 O modelo para itens dicotomizados

Este modelo, representado pela Figura 1, é denominado de Modelo Logístico de Três Parâmetros, definido por:

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

Com  $i = 1, 2, 3, \dots, p$  (representando as  $p$  práticas da GQT que medem o construto); e

$j = 1, 2, 3, \dots, n$  (representando as  $n$  organizações pesquisadas).

Classificando-se como “implantada ou não implantada” satisfatoriamente a prática da GQT como forma dicotomizada de representar as questões, interpreta-se como:

- $U_{ij}$ , variável dicotômica que assume o valor 1 (um) quando a  $j$ -ésima organização tem implantada satisfatoriamente a  $i$ -ésima prática da GQT, ou assume 0 (zero) quando a  $j$ -ésima organização não tem implantada satisfatoriamente a  $i$ -ésima prática da GQT.
- $\theta_j$  representa o grau de maturidade quanto à GQT da  $j$ -ésima organização. Define-se como maturidade o nível de implantação na organização das práticas da GQT.
- $P(U_{ij} = 1/\theta_j)$  é a probabilidade de a  $j$ -ésima organização com grau de maturidade  $\theta_j$  aplicar satisfatoriamente a  $i$ -ésima prática da GQT.
- $a_i$  é o parâmetro que representa o poder de discriminação das organizações na  $i$ -ésima prática, com valor proporcional à inclinação da curva no ponto  $b_i$ .
- $b_i$  é o parâmetro que representa a dificuldade de implantação satisfatória da  $i$ -

ésima prática, medido na mesma escala da maturidade.

- $c_i$  é o parâmetro que representa a probabilidade de uma organização com baixo grau de maturidade implantar a  $i$ -ésima prática da GQT.
- $D$  é um fator de escala, constante e igual a 1. Quando se deseja comparar esse modelo com os resultados obtidos por intermédio da função ogiva normal, utiliza-se  $D = 1,7$ . Entretanto, esse não é o objetivo deste trabalho e, nesse sentido, utilizaremos  $D = 1$ .

#### 4.2 Estimativa dos parâmetros e recursos computacionais utilizados

O problema de estimar os parâmetros dos itens  $a$ ,  $b$  e  $c$  e da maturidade  $\theta$  pode ser dividido, do ponto de vista teórico, em 3 situações: quando já se conhecem os parâmetros dos itens e se tem somente que estimar a maturidade; quando já se conhece a maturidade e se está interessado em estimar os parâmetros dos itens; e quando ambos os parâmetros dos itens e a maturidade devem ser estimados simultaneamente.

Na TRI o processo de estimação dos parâmetros dos itens é denominado *calibração*.

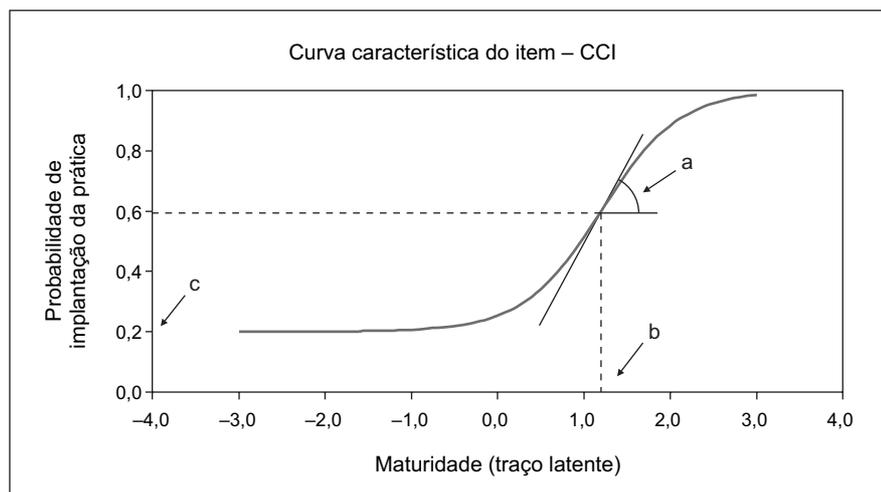


Figura 1 – Curva característica do modelo.

Em qualquer uma dessas situações, geralmente a estimativa é feita pelo Método da Máxima Verossimilhança (EMV), por meio da aplicação de algum processo iterativo, como o algoritmo de Newton-Raphson ou Scoring de Fisher. Alguns procedimentos bayesianos são aplicados com frequência. Para outros detalhes, veja Baker (1992).

O recurso computacional utilizado na análise desta pesquisa foi o programa BILOG (veja Mislevy & Bock, 1990). Esse programa executa a análise em três etapas (fases 1, 2 e 3), que se caracterizam pelo tipo de tarefas realizadas em cada uma delas. Na fase 1, fase de entrada e leitura de dados, os arquivos de saída fornecem algumas estatísticas descritivas e correlações de interesse, usadas na Teoria Clássica de Medida. A fase 2 é a fase de calibração dos itens. Nessa fase são estimados os parâmetros dos itens. A estimativa das habilidades (grau de maturidade) dos respondentes (organizações) é executada na fase 3. O BILOG fornece ainda gráficos contendo algumas informações de interesse, como: curvas características e curvas de informação de cada item e do teste.

Outro programa que pode ser utilizado é o BILOG-MG (Zimowski *et al.*, 1996). Os dois programas citados são específicos para análise via TRI de itens dicotômicos ou dicotomizados e ambos têm implementados os modelos unidimensionais logísticos de 1, 2 e 3 parâmetros. A diferença básica entre eles é que o BILOG-MG permite a análise de mais de um grupo de respondentes, enquanto o BILOG permite analisar apenas respondentes considerados provenientes de uma mesma população.

### 4.3 Interpretação dos parâmetros e adequação do modelo

#### *Interpretação do parâmetro a*

Na área educacional o parâmetro  $a$  está relacionado à discriminação dos indivíduos quanto à habilidade (conhecimento) no assunto. Sua interpretação dentro da qualidade está associada à discriminação das organizações, quanto à maturidade, na aplicação das práticas de qualidade.

O valor de  $a$  é proporcional à derivada da tangente da curva no ponto  $b$  (ponto de inflexão – ponto onde a curva muda de concavidade).

Valores negativos de  $a$  significam que, à medida que uma organização aumenta sua maturidade, diminui a probabilidade de ter a prática da GQT implantada. Essa afirmação não tem sentido, uma vez que organizações mais maduras, com uma cultura mais voltada para a qualidade, têm um ambiente mais adequado para implantar esse modelo. Assim, valores negativos de  $a$  não são esperados.

Valores positivos e altos do parâmetro  $a$  conduzem ao formato mais íngreme da curva e, como consequência, a diferença entre as probabilidades de implantação da prática entre organizações com diferentes níveis de maturidade é maior. Por outro lado, à medida que  $a \rightarrow 0$ , a curva perde a forma de “S” de inclinação, tornando-se mais achatada, o que significa que organizações com diferentes níveis de maturidade têm aproximadamente a mesma probabilidade de ter a prática implantada.

Esse parâmetro também permite investigar a qualidade dos itens. Quanto melhores forem os itens, mais apropriada é a escala elaborada. Hambleton & Swaminathan (1985) mostram que itens com  $a \geq 1$  apresentam bom poder de discriminação. Neste artigo, considerando que foram pesquisadas apenas 75 indústrias para estimativa dos parâmetros dos itens (veja Tabela 1), somente os itens com valores de  $a < 0,7$  foram considerados inadequados e, conseqüentemente, eliminados do estudo.

#### *Interpretação do parâmetro b*

Este parâmetro é medido na mesma escala da maturidade  $\theta$ . Na área educacional ele está associado à dificuldade de um indivíduo responder corretamente uma questão. No contexto da qualidade, ele está associado ao grau de dificuldade de uma organização em implantar a prática da GQT: *à medida que b cresce, aumenta o grau de dificuldade de implantação das práticas (o que significa que poucas organizações as implantam), e vice-versa.*

A correspondência entre o parâmetro  $b$  e a maturidade organizacional  $\theta$  consiste na premissa de que as organizações que ainda não despertaram efetivamente para o modelo de gestão com base na GQT, isto é, com pouca maturidade, terão grandes dificuldades na implantação dessas práticas. Essa afirmação torna-se possível porque os parâmetros de dificuldade e maturidade estão na mesma escala. Esta é, provavelmente, uma vantagem significativa dos modelos da TRI sobre a TCM.

### **Interpretação do parâmetro $c$**

No contexto da educação (veja Andrade *et al.*, 2000), este parâmetro está associado à “probabilidade de um indivíduo com baixa habilidade no tema da prova acertar a questão”. Pode ser denominado acerto casual. O parâmetro  $c$  é uma probabilidade, portanto, assume valores entre 0 e 1.

Na interpretação deste parâmetro quanto à qualidade, cabem os seguintes questionamentos:

- Qual a probabilidade de uma organização com baixo grau de maturidade na filosofia da qualidade ter implantada a prática da GQT?
- Há probabilidade de implantação da prática da GQT associada ao acaso?

A construção de modelos teóricos da GQT evolui de forma inversa ao que normalmente ocorre: em princípio a implantação da filosofia da GQT era uma “colcha de retalhos”. A partir de experiências bem-sucedidas das organizações, diversos autores sintetizaram essa filosofia em um modelo teórico conceitual. Como ilustração podem ser citados Shiba *et al.* (1993), Merli (1993) e Galgano (1993). Essas organizações, contudo, já tinham conhecimento adquirido, isto é, conhecimento intrínseco, sobre gestão pela qualidade. As implantações dos elementos da GQT, portanto, não ocorreram de forma casual.

Uma premissa básica da filosofia da GQT é seu enfoque sistêmico: as definições da política e da estratégia organizacional devem ser desdobradas

em ações sincronizadas por toda a organização, desde a alta gerência até o nível operacional.

Diante deste contexto, o valor assumido pelo parâmetro  $c$  no âmbito da GQT é zero: a probabilidade de uma organização com baixa maturidade implantar a GQT é zero. Como consequência, o modelo adequado é o de dois parâmetros: discriminação e dificuldade.

Uma ilustração que visa referendar as interpretações anteriores e a adequabilidade do modelo, considerando a variação dos valores dos parâmetros, com a teoria da GQT é descrita na Figura 2.

- *Comparação das práticas com igual valor para  $b$ .* Comparando a prática 2 com a prática 4 (igualmente a prática 1 com a 3), percebe-se que a diferença entre as probabilidades de implantação desta prática entre organizações com graus de maturidade diferentes é maior para as práticas que têm maiores valores para o parâmetro  $a$ . Particularmente sejam  $\theta_1 = 1$  e  $\theta_2 = 2$ , a diferença para a prática 4 é 0,358 (= 0,679 – 0,321) e para a prática 2 é 0,222 (= 0,611 – 0,389). Isso significa que a prática 4 é mais apropriada para discriminar duas organizações do que a prática 2. Nesse sentido, o parâmetro  $a$  é chamado de parâmetro de discriminação.
- *Comparação de práticas com valores iguais de  $a$ .* Comparando, agora, as práticas 1 e 2 (igualmente as práticas 3 e 4), percebe-se que as práticas com maiores valores de  $b$  exigem maior maturidade para uma mesma probabilidade de implantação da prática. Como ilustração, a maturidade requerida para uma probabilidade de implantação 0,4 é igual a –1,3 para a prática 1 e igual a 1,2 para a prática 2. Isto é, a prática 2 é mais difícil de implantar do que a prática 1. Dessa forma,  $b$  é denominado parâmetro de dificuldade.

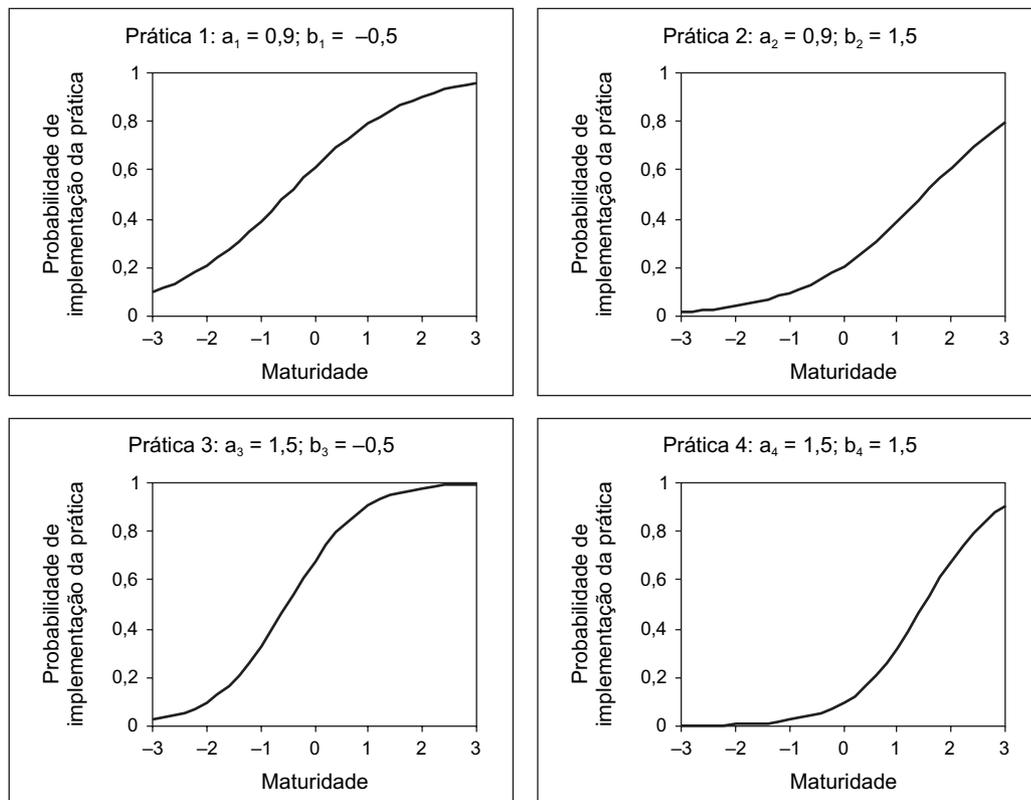


Figura 2 – Curva característica das práticas: uma interpretação dos parâmetros.

#### 4.4 O parâmetro de maturidade organizacional $\theta$

O parâmetro  $\theta$  representa maturidade organizacional na implantação das práticas da GQT (traço latente, isto é, variável que não pode ser medida diretamente). Teoricamente este parâmetro pode assumir valores de  $-\infty$  e  $+\infty$ . É preciso, portanto, estabelecer uma origem e uma unidade de medida para definição da escala de maturidade. Estes valores são determinados de forma a representarem, respectivamente, a média ( $\mu$ ) e o desvio-padrão ( $\sigma$ ) das maturidades das organizações inseridas no estudo. Uma escala bastante utilizada na TRI é aquela com  $\mu = 0$  e  $\sigma = 1$ , representada por  $(0,1)$ , a qual foi utilizada nas aplicações deste artigo.

Em termos práticos, não há a menor diferença em estabelecer qualquer outro tipo de escala. O importante são as relações existentes entre seus pontos. Por exemplo, na escala  $(0,1)$  uma organização com maturidade 1,20 está 1,20

desvio-padrão acima da maturidade média. Caso fosse utilizada a escala  $(200, 40)$  e o valor da maturidade fosse 248, a interpretação seria a mesma. *O grande desafio consiste na criação de uma interpretação prática para a escala de maturidade: o que significa, em termos de maturidade, uma organização com valor de  $\theta = 2$  na escala  $(0,1)$ ?*

Um importante comentário a ser feito é em relação à formulação dos itens para construção dessa escala, que pretende medir o grau de maturidade na GQT das organizações em todos os setores de atividades. Itens específicos não são adequados, isto é, uma prática da GQT específica a um tipo de indústria não é apropriada para essa escala. Por exemplo, a prática “*a indústria tem um sistema para reduzir custos de estocagem*” não é apropriada para a construção da escala em questão, pois esse item não é aplicável no setor de serviços – serviços não são estocados.

A escala deve refletir uma maturidade acumulada de todas as organizações na implantação das práticas da GQT. Esse item seria apropriado para medir a maturidade entre as organizações do setor manufatureiro.

## 5. Uma aplicação da TRI na GQT no cenário industrial cearense

O modelo foi aplicado aos resultados de uma pesquisa realizada em Fortaleza, CE, nos meses de agosto de 1998 a janeiro de 1999, cujo objetivo era investigar as práticas da GQT no setor manufatureiro de portes médio e grande do Estado do Ceará (para outros detalhes veja Alexandre, 1999). A escala de Likert com categorias variando de 1 (discordo totalmente) a 5 (concordo totalmente) foi utilizada nas questões que mediam os construtos da GQT segundo um modelo de referência (Tabela 1). O entrevistado marcava a dimensão nas sentenças conforme a indústria se enquadrava na escala. Nessa tabela são apresentadas as estimativas dos parâmetros dos itens, com seus respectivos erros-padrão. A coleta de dados foi realizada por meio de entrevista pessoal com o profissional da área da qualidade da indústria e, em sua ausência, com o gerente geral ou diretor.

Para efeito da aplicação deste modelo (dicotomizado), as categorias da escala de Likert foram recodificadas em: aplica satisfatoriamente a prática da GQT (categorias  $\geq 4$ ) e não aplica satisfatoriamente a prática da GQT (categorias  $\leq 3$ ). É importante salientar que essa recodificação foi feita após a realização da pesquisa e, portanto, a escala original não foi respondida pelas indústrias. O motivo principal dessa recodificação foi a aplicação do modelo aqui proposto.

Nesta análise, conforme mencionado anteriormente, assume-se que valores de  $a < 0,7$  indicam que a prática definida para medir o construto não foi bem elaborada. Assim, dentre as 43 práticas inseridas no questionário, temos, como resultado final, que somente 5 delas têm

valores menores que 0,7 e, portanto, foram retiradas do questionário na primeira fase de execução do programa: práticas 3.1, 3.2, 4.3, 6.1, 6.2 (esses valores correspondem à primeira fase da análise e foram incluídos na Tabela 1 apenas para evidenciar a retirada das práticas citadas anteriormente. Logo, no resultado final, elas não são consideradas).

Assumindo que a distribuição do grau de maturidade pode ser representada pela distribuição normal padronizada, valores negativos de  $b$  significam que mais de 50% das indústrias com grau de maturidade maior do que  $b$  aplicam aquela prática; já valores positivos de  $b$  significam que menos de 50% das indústrias aplicam a prática. Isso indica a dificuldade de implantação.

A Tabela 2 apresenta as estimativas do parâmetro de maturidade  $\theta$ , com seus respectivos erros-padrão, para as 75 indústrias inseridas na pesquisa. Mesmo considerando que neste artigo não é fornecida uma interpretação prática para a escala da maturidade, uma análise dos valores extremos revela que: as indústrias de números 32 ( $\theta = -2,34$ ) e 53 ( $\theta = -2,92$ ) são cearenses de porte médio e não têm programa de gestão da qualidade; as políticas das indústrias são, respectivamente, implantar primeiro a GQT e depois a ISO 9000 e não implantar nem a ISO 9000 nem a GQT. As indústrias de números 10 ( $\theta = 1,76$ ) e 70 ( $\theta = 2,65$ ) são não-cearenses de portes, respectivamente, médio e grande. Ambas têm programas formais de GQT, experiência acima de dois anos e seu principal fator para adoção do modelo é a busca da melhoria dos processos. Os resultados preliminares ilustram a adequabilidade do modelo: quanto maior o valor para  $\theta$ , maior a maturidade organizacional na filosofia da GQT.

É importante ressaltar que os valores dos erros-padrão, tanto para as estimativas dos itens quanto para as estimativas do grau de maturidade, são altos em razão de haver, respectivamente, pequeno número de indústrias pesquisadas e pequeno número de itens aplicados.

**Tabela 1 – Distribuição de itens (práticas da GQT), estimativas dos parâmetros dos itens e respectivos erros-padrão entre parênteses.**

Elementos da GQT/descrição das práticas (itens)	Estimativas	
	a	b
<b>1. Comprometimento da alta administração</b>		
1.1 Avaliação periódica da qualidade executada pela alta administração	1,71 (0,63)	-0,93 (0,21)
1.2 Discussão da importância da qualidade pela alta administração	1,89 (0,76)	-1,34 (0,28)
1.3 Alocação de verbas e recursos para a qualidade definidos em orçamento	1,48 (0,37)	-0,68 (0,22)
1.4 Definição (identificação) e documentação das metas da qualidade	2,64 (1,05)	-0,38 (0,16)
1.5 As metas da qualidade inseridas no planejamento estratégico da empresa	1,44 (0,44)	-1,06 (0,25)
1.6 Comunicação ativa da alta administração do compromisso pela qualidade	1,38 (0,36)	-0,41 (0,23)
<b>2. Foco no consumidor</b>		
2.1 Comparação da satisfação do cliente com indicadores internos/concorrentes	1,04 (0,27)	-0,95 (0,34)
2.2 Fornecimento das reclamações dos consumidores a todos os departamentos	0,73 (0,20)	0,25 (0,33)
2.3 Utilização das reclamações do cliente como base para a melhoria da qualidade	1,24 (0,35)	-1,71 (0,45)
2.4 Manutenção do serviço de atendimento ao consumidor	1,09 (0,30)	-0,52 (0,27)
2.5 Realização periódica de pesquisas com o consumidor	1,03 (0,27)	0,52 (0,26)
<b>3. Parceria com o fornecedor</b>		
3.1 Qualidade e preço como critérios para a seleção do fornecedor	0,51 (0,18)	-4,17 (1,57)
3.2 Realização de contratos de longo prazo com fornecedor	0,41 (0,12)	2,32 (0,92)
3.3 Fornecimento de assistência técnica ao fornecedor	2,00 (0,53)	0,64 (0,19)
3.4 Participação do fornecedor no processo de desenvolvimento e fabricação	0,81 (0,20)	0,46 (0,34)
<b>4. Envolvimento dos funcionários (relação com os funcionários)</b>		
4.1 Realização periódica de reuniões em cada área para discussão sobre qualidade	1,29 (0,32)	0,11 (0,22)
4.2 Reuniões periódicas de equipes interfuncionais para discussão sobre qualidade	1,80 (0,45)	-0,37 (0,18)
4.3 Avaliação de todas as sugestões dos funcionários	0,57 (0,17)	-1,53 (0,57)
4.4 Implantação das sugestões dos funcionários	0,74 (0,21)	-1,37 (0,47)
4.5 Prêmios/recompensas não financeiras pelas melhores sugestões	0,90 (0,26)	1,29 (0,41)
<b>5. Treinamento</b>		
5.1 Alocação de recursos necessários para o treinamento em qualidade	1,01 (0,26)	0,74 (0,30)
5.2 Envolvimento de todos os escalões no treinamento em qualidade	1,16 (0,31)	-0,70 (0,25)
5.3 Treinamento de muitos funcionários em técnicas de solução de problemas	1,98 (0,60)	0,64 (0,19)
5.4 Treinamento dos funcionários em ferramentas/técnicas estatísticas	1,66 (0,53)	1,05 (0,23)

**Tabela 1 – Distribuição de itens (práticas da GQT), estimativas dos parâmetros dos itens e respectivos erros-padrão entre parênteses (continuação).**

Elementos da GQT/descrição das práticas (itens)	Estimativas	
	a	b
<b>6. Mensuração da qualidade</b>		
6.1 Execução de inspeções por amostragem durante o processo de produção	0,67 (0,19)	-2,64 (0,79)
6.2 Avaliação da qualidade não executada somente com inspeção final	0,53 (0,16)	-2,66 (0,90)
6.3 Medição periódica dos desperdícios e falhas de produto não-conforme	0,84 (0,25)	-2,27 (0,63)
6.4 Manutenção de registros das avaliações da qualidade	1,16 (0,33)	-1,03 (0,29)
6.5 Fornecimento dos resultados das avaliações a todos os funcionários	1,18 (0,34)	1,02 (0,29)
6.6 Resultados das avaliações como suporte para a melhoria da qualidade	1,40 (0,41)	-1,07 (0,29)
<b>7. Melhoria contínua</b>		
7.1 Manutenção de estrutura organizacional específica para apoiar a qualidade	1,18 (0,32)	-0,01 (0,23)
7.2 Há programa formal para a redução de tempo e custo nos processos internos	1,31 (0,37)	0,70 (0,25)
7.3 Execução de avaliações nos processos-chave de produção	1,39 (0,41)	-0,95 (0,27)
7.4 Há um programa formal para a redução do tempo de entrega de produto	0,96 (0,27)	0,27 (0,28)
7.5 Há um programa formal para a redução do tempo de fabricação	1,15 (0,30)	-0,19 (0,25)
<b>8. Benchmarking</b>		
8.1 Visita a outras organizações reconhecidamente líderes	1,13 (0,30)	0,12 (0,25)
8.2 Efetivo procedimento de <i>benchmarking</i> dos competidores mais fortes	0,83 (0,24)	1,46 (0,48)
8.3 Efetivo procedimento de <i>benchmarking</i> com líderes não competidores	1,16 (0,32)	1,50 (0,40)
8.4 Manutenção efetiva do <i>benchmarking</i> como uma política da empresa	0,83 (0,23)	0,30 (0,31)
<b>9. Empowerment (delegação de poderes) aos funcionários</b>		
9.1 Delegação de poderes aos funcionários para a solução de problemas	0,86 (0,24)	-1,24 (0,38)
9.2 Fornecimento de apoio aos funcionários para a solução de problemas	1,23 (0,35)	-1,63 (0,42)
9.3 Inspeção da qualidade executada pelos funcionários (não por um inspetor)	0,94 (0,24)	-0,28 (0,28)
9.4 Divulgação das experiências de sucesso na solução de problemas	0,93 (0,25)	0,49 (0,30)

## 6. Conclusões e sugestões

O modelo proposto da TRI revelou-se uma ferramenta poderosa de aplicação na GQT como alternativa à TCM. As interpretações dos parâmetros e os resultados obtidos da aplicação são amplamente satisfatórios.

A construção da escala de maturidade na implantação das práticas da GQT é o grande desafio para futuras pesquisas, isto é, a elaboração de uma escala com níveis predefinidos e interpretados no contexto da GQT. Na educação, um exemplo prático de uma escala de habilidade é a Escala Nacional de Habilidade construída

pelo Instituto Nacional de Estudos e Pesquisas Educacionais do Ministério da Educação (INEP/MEC) a partir dos resultados do Sistema

Nacional de Avaliação do Ensino Básico (SAEB). Para mais detalhes veja o site: <http://www.inep.gov.br>.

**Tabela 2 – Estimativas dos parâmetros de maturidade das 75 indústrias pesquisadas, com os respectivos erros-padrão.**

<b>Indústria</b>	1	2	3	4	5	6	7	8	9
<b>Maturidade</b>	0,79	0,43	-0,04	0,52	1,51	-0,45	-0,35	-0,39	-0,45
<b>Erro-padrão</b>	0,44	0,15	0,44	0,25	0,37	0,10	0,26	0,21	0,11
<b>Indústria</b>	<b>10</b>	11	12	13	14	15	16	17	18
<b>Maturidade</b>	<b>1,76</b>	-1,34	1,11	1,42	-0,47	-0,90	0,61	-0,36	1,05
<b>Erro-padrão</b>	0,47	0,17	0,39	0,31	0,15	0,45	0,35	0,26	0,42
<b>Indústria</b>	19	20	21	22	23	24	25	26	27
<b>Maturidade</b>	0,41	0,46	-0,36	0,86	-0,48	1,56	-2,32	-1,11	-0,45
<b>Erro-padrão</b>	0,18	0,15	0,26	0,45	0,17	0,40	0,38	0,39	0,11
<b>Indústria</b>	28	29	30	31	<b>32</b>	33	34	35	36
<b>Maturidade</b>	-1,50	-2,11	-0,44	0,44	<b>-2,34</b>	0,45	-1,37	0,41	1,35
<b>Erro-padrão</b>	0,35	0,38	0,10	0,13	0,39	0,13	0,21	0,17	0,26
<b>Indústria</b>	37	38	39	40	41	42	43	44	45
<b>Maturidade</b>	-0,48	1,42	0,07	0,71	-0,86	-1,27	-0,51	-0,41	-0,75
<b>Erro-padrão</b>	0,17	0,31	0,44	0,41	0,44	0,24	0,23	0,17	0,42
<b>Indústria</b>	46	47	48	49	50	51	52	<b>53</b>	54
<b>Maturidade</b>	-1,32	1,63	-1,34	-0,40	-0,49	-1,26	-1,33	<b>-2,92</b>	-1,19
<b>Erro-padrão</b>	0,17	0,43	0,17	0,20	0,20	0,26	0,17	0,53	0,34
<b>Indústria</b>	55	56	57	58	59	60	61	62	63
<b>Maturidade</b>	0,47	0,52	-0,38	1,57	-0,45	-0,46	0,41	0,39	-0,45
<b>Erro-padrão</b>	0,16	0,25	0,24	0,40	0,10	0,13	0,18	0,23	0,10
<b>Indústria</b>	64	65	66	67	68	69	<b>70</b>	71	72
<b>Maturidade</b>	1,08	0,39	-0,49	-2,03	1,30	1,53	<b>2,65</b>	0,90	1,08
<b>Erro-padrão</b>	0,41	0,22	0,20	0,41	0,27	0,39	0,55	0,45	0,41
<b>Indústria</b>	73	74	75						
<b>Maturidade</b>	1,28	-0,43	-0,65						
<b>Erro-padrão</b>	0,28	0,12	0,38						

Como ilustração e para melhor compreensão desse desafio considere, na área educacional, a situação hipotética em que se deseja representar a habilidade – o equivalente à maturidade na GQT de estudantes em matemática. O valor de habilidade  $\theta_1$  pode significar, por exemplo, que o estudante tem conhecimento em adição e subtração; o valor de habilidade  $\theta_2$  ( $\theta_2 > \theta_1$ ) pode significar que, além de adição e subtração, o estudante tenha conhecimento de divisão e multiplicação. A escala da habilidade é, portanto, acumulativa. Na GQT, o valor de maturidade  $\theta_1$  pode significar, por exemplo, que as indústrias têm implantado somente algumas ferramentas estatísticas básicas da qualidade em seu sistema, enquanto para o valor de maturidade  $\theta_2$ , além dessas ferramentas, as indústrias têm implantado as ferramentas do planejamento da qualidade e o *benchmarking*. A escala representa,

assim, uma maturidade acumulativa da GQT das indústrias.

É importante afirmar que a TCM e a TRI não são teorias conflitantes. A TRI foi construída para preencher as lacunas existentes na TCM.

Futuras pesquisas, portanto, são necessárias para corroborar a aplicação da TRI na GQT: melhores estimativas dos parâmetros, determinação da métrica e da interpretação do parâmetro de maturidade  $\theta$  e a utilização de outros modelos para representar a escala original de Likert (categorizada ordinal) são algumas sugestões.

Este artigo não tem a pretensão de esgotar o assunto. Ao contrário, o objetivo, aqui, consiste em introduzir uma fonte alternativa de análise de construto da GQT e, certamente, as perguntas que nortearam este artigo estão em aberto.

### Referências Bibliográficas

- AHIRE, S. L. TQM age versus quality: an empirical investigation. *Production and Inventory Management Journal*, first quarter, v. 37, n. 1, p. 18-23, 1996.
- ALEXANDRE, J. W. C. *Uma investigação das práticas da gestão da qualidade total no setor manufatureiro do estado do Ceará*. 1999. Tese (Doutorado) – Departamento de Engenharia de Produção, Escola Politécnica, USP, São Paulo.
- ANDRADE, D. F. Desempenhos de grupos de alunos por intermédio da teoria da resposta ao item. *Estudos em Avaliação Educacional*, n. 23, p. 31-69, jan.-jun. 2001.
- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. Teoria da resposta ao item: conceitos e aplicações. In: SINAPE. 14., 2000, São Paulo: ABE-Associação Brasileira de Estatística.
- BADRI, M. A.; DONALD, D.; DONNA, D. A study of measuring the critical factors of quality management. *International Journal of Quality & Reliability Management*, v. 12, n. 2, p. 36-53, 1995.
- BAKER, F. B. *Item response theory: parameter estimation techniques*. New York: Marcel Dekker, Inc., 1992.
- BAYLEY, S. Measuring customer satisfaction. *Evaluation Journal of Australasia*, v. 1 (new series), n. 1, mar. 2001.
- CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, v. 16, p. 297-334, 1951.
- DeROOS, Y.; MEARES, P. A. Application of Rasch analysis: exploring differences in depression between african-american and white children. *Journal of Social Service Research*, v. 23, n. 3/4, p. 93-107, 1998.
- GALGANO, A. *Calidad total, clave estratégica para la competitividad de la empresa*. Ediciones Diaz de Santos S.A., 1993.
- GRANGER, C. V.; DEUTSCH, A.; LINN, R. T. Rasch analysis of the functional independence measure (FIM™) mastery test. *Arch Phys med Rehabil*, v. 79, p. 52-57, jan. 1998.

- GULLIKSEN, H. *Theory of mental tests*. New York: John Wiley & Sons, Inc., 1967.
- HAMBLETON, R. K.; SWAMINATHAN, H. *Item response theory: principles and applications*. Boston: Kluwer-Nijhoff Publishing, 1985.
- HAYES, B. E. *Measuring customer satisfaction: development and use of questionnaires*. Milwaukee, Wisconsin: ASQC Quality Press, 1992.
- HUQ, Z.; STOLEN, J. D. Total quality management contrasts in manufacturing and service industries. *International Journal of Quality & Reliability Management*, v. 15, n. 2, p. 138-161, 1998.
- LORD, F. M. *Applications of item response theory to practical testing problems*. Hillsdale: Lawrence Erlbaum Associates, Inc., 1980.
- MERLI, G. *Eurochallenge: the TQM approach to capturing global markets*. Oxford-UK: IFS Ltd, 1993.
- MISLEVY, R. J.; BOCK, R. D. *BILOG 3, item analysis and test scoring with binary logistic models*. Chicago, Scientific Software, Inc., 1990.
- NUNNALLY, J. *Psychometric theory*. New York-NY: McGraw-Hill, 1967.
- TAMIMI, N. An empirical investigation of critical TQM factors using exploratory factor analysis. *International Journal and Product Research*, v. 33, n. 11, p. 3041-3051, 1995.
- TAMIMI, N.; GERSHON, M. A tool for assessing industry TQM practice versus the Deming philosophy. *Production and Inventory Management Journal*, first quarter, v. 36, n. 1, p. 27-32, 1995.
- TAMIMI, N.; GERSHON, M.; CURRALL, S. C. Assessing the psychometric properties of Deming's 14 principles. *Quality Management Journal*, v. 2, n. 3, p. 38-52, spring, 1995.
- SARAPH, J. V.; BENSON, P. G.; SCHROEDER, R. G. An instrument for measuring the critical factors of quality management. *Decision Sciences*, v. 20, n. 4, p. 810-829, 1989.
- SHIBA, S.; GRAHAM, A.; WALDEN, D. *A new american TQM: four practical revolutions in management*. Portland: Productivity Press, 1993.
- VIANNA, H. V. *Testes em educação*. São Paulo: IBRASA, 1978.
- ZIMOWSKI, M. F.; MURAKI, E.; MISLEV, R. J.; BOCK, R. D. *BILOG-MG, multiple-group IRT analysis and test maintenance for binary items*. Chicago: Scientific Software, Inc., 1996.

## A CONSTRUCT ANALYSIS PROPOSAL FOR MEASURING THE TOTAL QUALITY MANAGEMENT CRITICAL FACTORS THROUGH THE ITEM RESPONSE THEORY

### **Abstract**

*In this article we propose the use of models of item response theory (IRT) in the constructs analysis of the total quality management (TQM) as an alternative to the classical measure theory (CMT). Interpretations of the model parameters are done. The results show that the IRT can be a powerful tool in the analysis of TQM implementations as also in the study of organizational maturity in quality philosophy.*

**Key words:** *total quality management, item response theory, TQM constructs analysis.*