



## **COEFICIENTES DE FIDEDIGNIDADE PARA MENSURAÇÕES QUALITATIVAS**

**Carlos Alberto Ferreira Bispo (EESC-USP)**

cafbispo@sc.usp.br

**Daniela Gibertoni (UFSCar)**

dagibertoni@ig.com.br

*O objetivo deste artigo é fazer uma revisão da literatura e apresentar as principais técnicas para se estimar a fidedignidade de mensurações qualitativas. São apresentados: a forma genérica de estimação da fidedignidade de uma amostra de dados; os métodos teste-reteste, das formas alternativas, da bipartição, de Kuder-Richardson; o coeficiente alpha; a discussão do valor ideal para o coeficiente de fidedignidade; a estimação do erro padrão da mensuração.*

*Palavras-chaves: Coeficiente de Fidedignidade, Mensuração Qualitativa, Dados Qualitativos.*

## 1. Introdução

O mundo atual está se transformando rapidamente, principalmente por meio da globalização proporcionada pela revolução das tecnologias da informação e da comunicação (BISPO; CAZARINI, 2001). As empresas estão sendo obrigadas a se adaptarem frequentemente às mudanças de diversos tipos, impostas pelo cenário tecnológico, sócio-econômico, mercadológico, político e organizacional, etc., tanto nacional como internacional. E como saber se as mudanças realizadas na empresa estão adequadas às necessidades impostas? Surge então a necessidade de uma avaliação constante nos diversos indicadores que demonstram a “saúde” organizacional e sua capacidade de adaptação às mudanças, assim como uma melhor investigação no cenário em que a organização está inserida. A Engenharia de Produção tem um papel fundamental neste processo por estar ligada a um dos pilares organizacionais: a produção. Além de estar fortemente ligada a outros importantes pilares: finanças, marketing, recursos humanos e a pesquisa e desenvolvimento de novos produtos e serviços.

A maioria dos indicadores organizacionais é obtida por meio de métodos quantitativos consagrados. Porém, nem sempre esses indicadores conseguem descrever plenamente comportamentos, efeitos, características e propriedades de elementos organizacionais como mercado, Economia, Sociedade, Recursos Humanos, etc. Há então a necessidade da utilização de métodos qualitativos para investigá-los melhor. Porém, esses métodos empregam técnicas que fazem uso da subjetividade, ou seja, da percepção, abstração e interpretação de propriedades, características, comportamentos ou efeitos daquilo que está sendo avaliado. O uso dessa subjetividade proporciona maior quantidade de erros de mensuração. Aí surge um dos principais problemas deste tipo de investigação científica: é possível calcular ou estimar esses erros e/ou seus efeitos? De fato, existem técnicas para se estimar a fidedignidade dos dados obtidos por meio de mensurações qualitativas, essas técnicas serão apresentadas nas próximas seções.

O objetivo deste artigo é fazer uma revisão da literatura e apresentar as principais técnicas para se estimar a fidedignidade de mensurações qualitativas. Como justificativa e relevância, afirma-se que a literatura consultada não reúne todos os itens abordados neste artigo.

## 2. Validade e fidedignidade para dar credibilidade aos dados obtidos em mensurações

Embora a validade de mensurações qualitativas não esteja ligada diretamente aos objetivos deste artigo, esse assunto é interligado à estimação da fidedignidade, sendo citado algumas vezes ao longo do texto. Por isso, será abordado, porém, de forma bem resumida.

Se a validade e a fidedignidade de uma mensuração qualitativa não são conhecidas, podem surgir dúvidas acerca dos resultados e conclusões obtidos por trabalhos que tenham como suporte esse tipo de mensuração (RICHARDSON, 1999), podendo comprometer decisões importantes. Por isso, é importante conhecer as duas técnicas e relatar seus resultados e conclusões juntamente com os resultados e conclusões dos trabalhos envolvendo mensurações qualitativas. As duas técnicas juntas, validade e fidedignidade, possibilitam dar maior credibilidade às mensurações qualitativas. A fidedignidade é um complemento da validade.

Richardson (1999), Bunchaft & Cavas (2002), Erthal (2003) e Silva & Ribeiro Filho (2006) afirmam que há um conjunto de itens a serem averiguados para assegurar a validade de um instrumento de mensuração qualitativo ou de um conjunto de dados qualitativos:

– validade de conteúdo – demonstra o grau em que uma mensuração é representativa de um

universo definido ou de um domínio de conteúdo, frequentemente baseia-se nos julgamentos de especialistas;

- validade aparente – refere-se ao julgamento de quão relevante a mensuração parece ser, é frequentemente concebida a partir do ponto de vista dos pesquisados e não, exclusivamente, do ponto de vista do pesquisador, normalmente contribui para aumentar a cooperação dos pesquisados durante a aplicação das mensurações;
- validade de constructo – um constructo é uma idéia científica desenvolvida ou hipotetizada para descrever ou explicar as propriedades, características ou comportamentos de um “objeto”, dando origem a um modelo acerca da natureza do constructo; é esse modelo que se pretende mensurar; há várias evidências para averiguar a validade de constructo:
  - a) evidência da homogeneidade – a homogeneidade refere-se ao quão uniforme está sendo mensurado um conceito ou um constructo;
  - b) evidência de mudanças – são evidências de que as mensurações mudam em função de alguma alteração no “objeto” entre um pré-teste e um pós-teste; mudanças ocorridas na direção predita contribuem para aumentar essa evidência;
  - c) evidência da comparação entre grupos-extremos – um grupo que tenha indícios de possuir os mais altos escores da escala é comparado com outro grupo que tenha indícios de possuir os menores escores dessa mesma escala; espera-se que as mensurações captem essa distorção;
  - d) evidência convergente – correlações entre mensurações teoricamente similares deveriam ser altas (convergentes) enquanto correlações entre medidas teoricamente dissimilares deveriam ser baixas (divergentes);
  - e) evidência discriminante – correlações entre mensurações de constructos totalmente distintos devem apresentar baixos valores;
- validade concorrente – ocorre quando a mensuração do “objeto” relaciona-se com alguma medida de um critério externo obtido ao mesmo tempo; se aquela mensuração tiver alta relação com os indicadores externos, há um reforço para sua validação;
- validade preditiva – ocorre quando por meio da mensuração realizada no “objeto” é possível se realizar alguma predição a respeito de um comportamento futuro do “objeto” que não esteja previsto originalmente naquelas mensurações.

Quanto à fidedignidade, alguns autores empregam o termo fidedignidade para mensurações qualitativas e o termo confiabilidade para mensurações quantitativas, como por exemplo, os autores da Psicometria. Essa ciência é a que mais estuda e, em consequência, mais possui publicações a respeito de mensurações qualitativas. Outro grupo de autores usa as duas palavras como sinônimas. Outro grupo de autores somente emprega o termo confiabilidade, principalmente os autores da área de Ciências Exatas. Neste artigo será adotado o termo fidedignidade, por ser mais empregado nos artigos consultados.

Entre os estudos relacionando a validade e a fidedignidade de um instrumento de mensuração ou de um conjunto de dados, tanto o instrumento quanto o conjunto de dados são considerados válidos para um dado propósito se de fato medem o que se pretende medir (SILVA & RIBEIRO FILHO, 2006). Por exemplo, em mensurações quantitativas, uma régua é um instrumento válido para se medir o comprimento de uma mesa e um hodômetro é um instrumento válido para medir a distância entre cidades. Embora ambos usem a mesma categoria de escala (o sistema métrico), a inversão destes instrumentos neste exemplo os tornaria inválidos. Em mensurações qualitativas, o questionário e a entrevista são instrumentos válidos para avaliar a satisfação dos clientes, assim como a observação direta é

um instrumento válido para a avaliação do comportamento dos funcionários, a inversão destes instrumentos neste outro exemplo os tornaria inválidos.

Quanto à fidedignidade, se a mensuração não for fidedigna também torna o instrumento inválido. Por exemplo, em mensurações quantitativas, usar uma régua com a escala em polegadas para medir em centímetros o tamanho de uma mesa, torna a mensuração inválida, assim como o próprio instrumento de mensuração. Mesmo que a escala seja a mais adequada, mas se não estiver aferida, apresentará erros de mensuração que não a tornarão fidedigna, invalidando-a juntamente com o instrumento de mensuração. O mesmo raciocínio é empregado nas mensurações qualitativas, utilizar escalas inapropriadas ou imprecisas torna o instrumento de mensuração inválido ou as mensurações realizadas invalidadas, conforme o que estiver tentando ser validado.

### 3. Estimação da fidedignidade

Nas mensurações quantitativas, primeiro obtêm-se a precisão dos instrumentos de mensuração. Garantida a precisão do instrumento, é possível assegurar-se a precisão das mensurações realizadas por esse instrumento. Mesmo com toda essa precisão, há a necessidade do cálculo do erro de mensuração, que na maioria dos casos é muito pequeno, considerado insignificante e somente sendo apresentado quando necessário.

Devido às características inerentes às mensurações qualitativas, ao invés de se calcular sua precisão, é estimada a sua fidedignidade (SILVA & RIBEIRO FILHO, 2006). Assim como acontece com os instrumentos de mensuração quantitativa, inicialmente é estimada a fidedignidade do instrumento de mensuração qualitativa. Assegurada sua boa fidedignidade, é possível averiguar a fidedignidade das mensurações realizadas com esse instrumento.

Para Erthal (2003), três aspectos fundamentais devem ser considerados na fidedignidade de mensurações qualitativas: 1) *precisão* – repetir várias vezes a mesma mensuração no mesmo “objeto” e coletar praticamente os mesmos erros em todas elas; 2) *estabilidade* – repetir várias vezes a mesma mensuração no mesmo “objeto” e obter resultados praticamente idênticos; 3) *consistência interna ou homogeneidade* – indica que todos os itens da investigação qualitativa mensuram o mesmo aspecto do “objeto” previsto nos objetivos daquela investigação.

Assim, surgiram vários métodos para estimar os aspectos da fidedignidade. Cada um tem um significado diferente e pode ter uma interpretação particular dada a sua utilidade prática (SILVA & RIBEIRO FILHO, 2006). Por esta razão, quando se fala de um coeficiente de fidedignidade, é preciso indicar qual aspecto da fidedignidade se está sendo mensurado (ANASTASI, 1977).

#### 3.1 Forma genérica de estimação da fidedignidade

Spearman (1913, apud De Bem, 2004) propôs um modelo de estimação dos erros de mensuração que atualmente se constitui em uma das equações fundamentais da Psicometria:

$$X_{ij} = X_i^v + \epsilon_{ij} \quad (1)$$

Onde:

$X_{ij}$  representa a j-ésima medição observada pelo i-ésimo indivíduo;

$X_i^v$  representa o escore verdadeiro, nunca conhecido, somente estimado pelo indivíduo i;

$\epsilon_{ij}$  representa a parcela de erro proveniente da ação de diversas fontes de erro de mensuração.

Segundo Cronbach (1996) e Silva & Ribeiro Filho (2006), a fidedignidade ( $r$ ) de uma

mensuração pode ser concebida como a proporção de sua “verdade”, assim, sua estimativa de pode ser assim calculada:

$$r = \frac{\text{variância do escore verdadeiro}}{\text{variância do escore observado}} = \frac{\text{var}(T)}{\text{var}(X)} \quad (2)$$

Para Silva & Ribeiro Filho (2006), não é possível calcular a variância dos escores verdadeiros, porém, é possível estimar sua variabilidade. A covariância mede a variância “compartilhada” entre as duas mensurações, então, é um indicador da variabilidade dos escores verdadeiros de duas mensurações  $X_1$  e  $X_2$ , pois são as únicas coisas que ambas compartilham. Assim, a parte superior das razões acima é essencialmente uma estimativa de var (T) neste contexto; a parte inferior multiplica os desvios padrão das duas medidas, portanto, torna-se a variância da medida ou var (X):

$$r = \frac{\text{cov}(X_1 \cdot X_2)}{\sigma(X_1) \cdot \sigma(X_2)} \quad (3)$$

Silva & Ribeiro Filho (2006) e Richardson (1999) lembram que a fórmula genérica para estimar a fidedignidade de uma escala é mais aplicável a mensurações quantitativas onde as escalas são mais precisas, por exemplo, para medir comprimentos, pesos, etc. Em mensurações qualitativas onde as escalas não são tão precisas, por exemplo, para mensurar níveis de interesses, habilidades, comprometimento, etc., há a necessidade de se estimar a fidedignidade levando em consideração os fatores qualitativos. Por isso surgiram diversos métodos e coeficientes para estimar a fidedignidade de mensurações qualitativas.

### 3.2 Método teste-reteste

Anastasi (1977), Richardson (1999), Bunchaft & Cavas (2002), Erthal (2003) e Silva & Ribeiro Filho (2006) afirmam que o método teste-reteste avalia o *aspecto da estabilidade* da fidedignidade. Envolve a aplicação de um determinado teste (ou conjunto de mensurações) por duas ou mais vezes em um mesmo “objeto” com certo intervalo de tempo entre elas. Esse intervalo vai depender do tipo de “objeto” e o que está sendo mensurado neste “objeto”. Dependendo da estabilidade dos componentes do “objeto”, o intervalo pode ser de horas (p. ex.: greves, revoltas), dias (p. ex.: pesquisa de opinião), semanas (p. ex.: satisfação dos clientes) ou meses (p. ex.: sistema de ensino, sistema de saúde, sistema de transporte, etc.).

Segundo Erthal (2003) e Silva & Ribeiro Filho (2006), a correlação entre os dois conjuntos de dados qualitativos obtidos fornecerá o *coeficiente de estabilidade* do instrumento ou das mensurações realizadas, calculado pelo coeficiente produto-momento de Pearson. Quanto mais próxima de 1 for esse coeficiente, maior será a estabilidade do instrumento ou das mensurações ao longo do tempo. Em contrapartida, se a correlação for próxima de zero ou negativa, significa que o instrumento ou as mensurações realizadas não têm estabilidade, comprometendo sua fidedignidade e, em consequência, sua validade.

Para Richardson (1999), Bunchaft & Cavas (2002), Erthal (2003) e Silva & Ribeiro Filho (2006), considerando-se X e Y como dois conjuntos de mensurações qualitativos do mesmo “objeto”, a fórmula para o cálculo para o *coeficiente de estabilidade* do método é dado pela correlação produto-momento de Pearson:

$$r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}} \quad (4)$$

Onde:  $r_{xy}$  = *coeficiente estabilidade* dado pela correlação produto-momento de Pearson;  $N$  = número de mensurações de um dos dois conjuntos de dados  $X$  ou  $Y$  (iguais);  $\sum XY$  = somatório do produto de cada valor do conjunto  $X$  pelo seu respectivo valor em  $Y$ ;  $\sum X$  = somatório dos valores do conjunto de dados  $X$ ;  $\sum Y$  = somatório dos valores do conjunto de dados  $Y$ ;  $\sum X^2$  = somatório do quadrado dos valores do conjunto de dados  $X$ ;  $\sum Y^2$  = somatório do quadrado dos valores do conjunto de dados  $Y$ ;  $(\sum X)^2$  = somatório dos valores do conjunto de dados  $X$  ao quadrado;  $(\sum Y)^2$  = somatório dos valores do conjunto de dados  $Y$  ao quadrado.

A aplicação deste método pode levar um ou mais de quatro possíveis problemas:

- a) a contaminação do teste para o reteste, ou seja, os participantes da pesquisa ficam sujeitos ao efeito-memória, onde na segunda aplicação eles meramente repetem as respostas dadas na primeira aplicação sem maiores esforços;
- b) variações no “objeto” ou em seu ambiente no intervalo entre as duas aplicações do teste;
- c) as pessoas envolvidas na mensuração, após a primeira aplicação do teste, se interessam mais pelo assunto e, durante o intervalo de tempo entre as duas aplicações, aperfeiçoam seus conhecimentos sobre o “objeto”, apresentando na segunda mensuração resultados distintos daqueles da primeira aplicação;
- d) pode ocorrer desinteresse das pessoas durante a segunda aplicação.

Portanto, o intervalo de tempo entre as aplicações é uma variável importante neste método.

### 3.3 Método das formas alternativas

Segundo Richardson (1999), Bunchaft & Cavas (2002), Erthal (2003) e Silva & Ribeiro Filho (2006), o método das formas alternativas, também chamado de formas paralelas ou formas equivalentes, não busca qualquer um dos três aspectos fundamentais da fidedignidade, ou seja, a *precisão*, a *estabilidade* e a *consistência interna* ou *homogeneidade*. O método consiste em utilizar dois instrumentos equivalentes para mensurar o mesmo “objeto”. Busca-se a equivalência entre dois instrumentos por meio da correlação entre dois conjuntos de dados mensurados por eles. Assim, obtém-se o *coeficiente de equivalência* calculado também pela correlação produto-momento de Pearson (ERTHAL, 2003). Quanto maior for a correlação, maior será a equivalência dos instrumentos utilizados. A principal dificuldade deste método é conseguir dois instrumentos distintos e equivalentes. Essa equivalência deve abranger: o conteúdo, o grau de dificuldade, as instruções, o tempo, o número de itens, etc. Como existe a necessidade de duas avaliações no mesmo “objeto”, o método está sujeito aos mesmos problemas do método teste-reteste. Porém, como são dois instrumentos distintos, podem ser aplicados sequencialmente ou simultaneamente, minimizando aqueles problemas.

Segundo Erthal (2003), após a indicação de alta equivalência entre dois instrumentos de mensuração qualitativa, estes podem ser utilizados de forma semelhante ao método teste-reteste, nesse caso, tornando-se uma alternativa para a obtenção do *coeficiente de estabilidade*. Também podem ser utilizados para que um sirva de referência para outro durante o aperfeiçoamento de um deles.

### 3.4 Método da bipartição

Richardson (1999), Bunchaft & Cavas (2002), Erthal (2003) e Silva & Ribeiro Filho (2006) afirmam que o método da bipartição ou repartição em metades busca o aspecto da *homogeneidade* ou *consistência interna* da fidedignidade. Nesse método, um conjunto de dados qualitativos é dividido em duas ou mais partes iguais e comparáveis. Normalmente, é

dividida em duas partes, daí a origem de seu nome. Dependendo do tipo de mensuração, pode-se realizar a divisão de diversas formas: dividir ao meio, separar os itens/questões pares dos ímpares, separação aleatória, ou outra forma que demonstre ser viável para a estimação do coeficiente. Porém, qualquer que seja a divisão, não se estima a homogeneidade da mensuração inteira e sim de suas partes. Para se obter a estimativa do *coeficiente de homogeneidade* ou *consistência interna* da mensuração por inteiro é necessário aplicar-se a fórmula desenvolvida por Spearman e Brown em 1910, que permite a correção da subestimação do coeficiente de correlação obtido nessas divisões:

$$r_{xx''} = \frac{n r_{xx'}}{1 + (n - 1) r_{xx'}} \quad (5)$$

Onde:  $r_{xx''}$  = coeficiente de correlação de Spearman-Brown para a mensuração completa;  $n$  = número de divisões da mensuração completa;  $r_{xx'}$  = coeficiente de correlação entre as divisões da mensuração completa.

### 3.5 – Método de Kuder-Richardson

Para Anastasi (1977), Richardson (1999), Bunchaft & Cavas (2002) e Silva & Ribeiro Filho (2006), esse método é uma evolução do método anterior, portanto, também busca o aspecto da *homogeneidade* ou *consistência interna* da fidedignidade. O método utiliza o coeficiente proposto por Kuder e Richardson (1937), chamado de KR20. O K e o R são as iniciais de seus nomes e o número 20 refere-se à vigésima equação desenvolvida por eles. No método, para cada questão ou item avaliado há uma forma de pontuação, que pode ser “certo ou errado”, “0 ou 1”, ou outro tipo de pontuação adequado às mensurações.

$$r_{tt} = \frac{n}{n - 1} \cdot \frac{\sigma_t^2 - n p \bar{q}}{\sigma_t^2} \quad (6)$$

Onde:  $r_{tt}$  = fórmula 20 do coeficiente de correlação de Kuder-Richardson;  $n$  = número de itens da investigação;  $p$  = proporção de respostas favoráveis ao item;  $q$  = proporção de respostas desfavoráveis ao item;  $\sigma_t^2$  = variância total considerando todos os escores da escala.

Considerando que todos os itens tenham o mesmo grau de dificuldade, Kuder e Richardson (1937) desenvolveram a fórmula 21 – KR21, que simplifica os cálculos:

$$r_{tt} = \frac{n}{n - 1} \cdot \frac{\sigma_t^2 - n p \bar{q}}{\sigma_t^2} \quad (7)$$

Cronbach (1951) apresenta uma versão mais moderna da fórmula KR20:

$$r_{tt(KR20)} = \frac{n}{n - 1} \left( 1 - \frac{\sum_i p_i q_i}{\sigma_t^2} \right); (i = 1, 2, \dots, n). \quad (8)$$

Onde:  $r_{tt(KR20)}$  = fórmula 20 do coeficiente de correlação de Kuder-Richardson;  $n$  = número de itens da investigação;  $p_i$  = respostas favoráveis ao item;  $q_i$  = respostas desfavoráveis ao item;  $\sigma_t^2$  = variância total considerando todos os escores da escala.

### 3.6 – Coeficiente alpha

Cronbach (1951) afirma que o coeficiente alpha ( $\alpha$ ) é um caso especial da fórmula KR 20 e é a média de todos os possíveis coeficientes de correlação obtidos por meio das bipartições da mensuração completa. Portanto, também busca o aspecto da *homogeneidade* ou *consistência interna* da fidedignidade:

$$\alpha = \left( \frac{n}{n-1} \right) \left( 1 - \frac{\sum V_i}{V_t} \right) \quad (9)$$

Onde:  $\alpha$  é o coeficiente alpha;  $n$  é o número de itens da escala;  $V_i$  é a variância de cada item;  $V_t$  é a variância do total de escores, ou seja, da escala total.

Segundo Pereira (2004), existem vários estudos e vários coeficientes alpha, uma de suas fórmulas é:

$$\alpha = \frac{k \text{ cov/var}}{1 + (k-1) \text{ cov/var}} \quad (10)$$

Onde:  $\alpha$  é o coeficiente alpha;  $k$  é o número de itens da escala;  $\text{cov}$  é a média das covariâncias;  $\text{var}$  é a média das variâncias.

### 3.7 – Qual o valor ideal para o coeficiente de fidedignidade?

Richardson (1999) afirma que ele deve ser proporcional à importância da decisão a ser tomada e das suas conseqüências. Na prática, ou seja, na maioria das pesquisas observadas na literatura, esse escore é considerado bom quanto está acima 0,7, ou seja, quando se obtém uma fidedignidade acima de 70% em cada um de seus aspectos.

### 4 – Estimação do erro padrão da mensuração

Segundo Anastasi (1977), Richardson (1999), Bunchaft & Cavas (2002), Erthal (2003) e Silva & Ribeiro Filho (2006), outro conceito importante para avaliar a validade e a fidedignidade de um instrumento ou de um conjunto de mensurações é a estimativa do erro padrão de mensuração (ou de medida) – EPM. Após a estimação de um dos coeficientes de fidedignidade, estima-se a amplitude na qual o escore verdadeiro pode estar situado, com um nível específico de confiança. Sua fórmula é:

$$\sigma_{\text{EPM}} = \sigma \sqrt{1 - r_{xx'}} \quad (11)$$

Onde:  $\sigma_{\text{EPM}}$  é o erro padrão de mensuração – EPM;  $\sigma$  é o desvio padrão dos escores das mensurações; e  $r_{xx'}$  é o coeficiente de fidedignidade das mensurações.

Segundo Richardson (1999) e Silva & Ribeiro Filho (2006), o EPM funciona de forma similar ao desvio padrão, assim:

- 68,26% dos escores verdadeiros estariam dentro  $\pm 1\sigma_{\text{EPM}}$  da média;
- 95,44% dos escores verdadeiros estariam dentro  $\pm 2\sigma_{\text{EPM}}$  da média;
- 99,74% dos escores verdadeiros estariam dentro  $\pm 3\sigma_{\text{EPM}}$  da média.

### 5 – Considerações finais

Foram apresentados mais de um coeficiente para dois dos três aspectos da fidedignidade, ou



seja, a *estabilidade* e a *consistência interna*, assim como, um coeficiente para estimar a *equivalência* entre dois instrumentos de mensuração. Porém, não foi encontrado na literatura, apesar de ser previsto sua existência, um coeficiente para o *aspecto da precisão*. Para a validação de um instrumento de mensuração ou de um conjunto de mensurações, seria importante empregar métodos que abrangessem os três aspectos da fidedignidade. Como um deles não foi localizado na literatura, está aberta a possibilidade de pesquisas nesta área.

## Referências

- ANASTASI, A.** *Testes psicológicos*. 2. ed. São Paulo, Editora Pedagógica e Universitária, 1977. Tradução: Dante Moreira Lima. 743 p. ISBN: 85-12650-30-3.
- BISPO, C. A. F. & CAZARINI, E. W.** A evolução, a situação atual e a tendência de alguns itens que influenciam a qualidade dos processos e dos produtos e também o nível de exigência dos consumidores. In: *Actas, IV Encontro para a Qualidade nas Tecnologias de Informação e Comunicações – QUATIC*, 2001. Lisboa - Portugal: Instituto Superior Técnico, 2001, p. 133-138.
- BUNCHAFT, G. & CAVAS, C. S. T.** *Sob Medida: um guia sobre a elaboração de medidas do comportamento e suas aplicações*. São Paulo, Vetor, 2002. 161 p. ISBN: 85-87516-79-5.
- CRONBACH, L. J.** Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, v. 16, n. 3, p. 297-334.
- CRONBACH, L. J.** *Fundamentos da testagem psicológica*. 5. ed. Porto Alegre, Artes Médicas, 1996. Tradução: Carlos Alberto Silva Neto e Maria Adriana Veríssimo Veronese. 559 p. ISBN: 85-7307-180-X.
- DE BEM, A. B.** *Confiabilidade e validade estatísticas da avaliação docente pelo discente: proposta metodológica e estudo de caso*. Florianópolis, Universidade Federal de Santa Catarina, 2004. 296 p. Tese (Doutorado em Engenharia de Produção).
- ERTHAL, T. C.** *Manual de Psicometria*. 7. ed. Rio de Janeiro: Zahar, 2003. 144 p. ISBN: 85-7110-341-0
- KUDER, G. F.; RICHARDSON, M. W.** The theory of the estimation of test reliability. *Psychometrika*, 1937, v. 2, n. 3, September, p. 151-160.
- PEREIRA, J. C. R.** *Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais*. 3. ed. São Paulo: Editora da Universidade de São Paulo – EDUSP, 2004. 157 p. ISBN 85-314-0523-8.
- RICHARDSON, R. J.** *Pesquisa Social: métodos e técnicas*. 3. ed. São Paulo, Atlas, 1999. 334 p. ISBN: 85-224-2111-0.
- SILVA, J. Ap. & RIBEIRO FILHO, N. P.** *Avaliação e mensuração da dor: pesquisa, teoria e prática*. Ribeirão Preto, FUNPEC Editora, 2006. 467 p. ISBN: 85-87528-96-3.