

Propriedades Psicométricas de Instrumentos de Avaliação: Um debate necessário

Psychometric properties of measurement instruments: a necessary debate

Luiz Alberto Pilatti

Bruno Pedroso

Gustavo Luis Gutierrez

Resumo

O presente trabalho tem como objetivo discutir os procedimentos que garantem qualidade aos instrumentos de coleta de dados elaborados pelo pesquisador. Trata-se de um estudo de cunho bibliográfico, com características exploratórias. Constatou-se que os procedimentos que qualificam um instrumento, evitando possíveis erros no levantamento de dados para o teste da hipótese, são a confiabilidade, a consistência interna e a validade. Conclui-se que a falta destas qualidades num instrumento, além de um problema ético, pode comprometer a fidedignidade do instrumento e, por extensão, produzir resultados distorcidos da pesquisa.

Palavras-chave: instrumentos de avaliação, propriedades psicométricas, validação, confiabilidade, consistência interna.

Abstract

The paper objectifies to argue the procedures that guarantee quality to the measurement instruments elaborated by researchers. It's a bibliographical study,

with exploratory characteristics. It was verified that procedures that qualify an instrument, avoiding possible mistakes in the data collection for the hypothesis test, are reliability, internal consistency and validity. It concludes that omission of these properties in an instrument constitutes an ethical problem and can jeopardize the fidelity of the instrument and also to distort the research results.

Keywords: measurement instruments, psychometric properties, validity, reliability, internal consistency.

Introdução

Investigações conduzidas por pesquisadores, com alguma frequência, apresentam erros. Os erros podem levar o pesquisador a concluir como falsa uma hipótese verdadeira ou, em medida inversa, aceitar aquelas que deveriam ser refutadas caso os cuidados epistemológicos devidos fossem observados (TURATO, 2003).

De acordo com Ioannidis (2005), a maioria das pesquisas científicas apresenta erros nos resultados. Tais erros, via de regra, são produto de procedimentos falhos na coleta de dados e/ou análise equivocada dos dados. Entre diversas classificações possíveis do erro em ciência, pode-se situar aquele entre dois extremos, o inevitável e o metodologicamente evitável (DEMO, 1989).

Inobstante a existência dos mais variados tipos de instrumentos de avaliação presentes na literatura, a ausência de instrumentos previamente construídos para a avaliação de determinadas variáveis perfaz com que seja requerida a elaboração de novos instrumentos.

Ao passo que diversos pesquisadores optam pela construção de instrumentos para utilização exclusiva em suas respectivas pesquisas, é notório que tais instrumentos, em grande parte, não são construídos com a preocupação de que apresentem características psicométricas satisfatórias. Há na literatura, portanto, um número considerável de instrumentos cujas propriedades psicométricas são inexistentes ou permaneçam obscuras.

Frente ao intróito perspectivado, o objetivo do presente estudo é discutir os procedimentos que garantem qualidade aos instrumentos de coleta de dados elaborados pelo pesquisador.

Qualidade dos instrumentos de avaliação

Seguindo as ações metodológicas propostas por Quivy e Campenhoudt (1992), é sugerida a adoção do modelo para procedimentos científicos que considera três atos para o estudo

científico: a ruptura, a construção e a verificação, que podem ser subdivididas em sete etapas, conforme pode ser verificado no Quadro 1:

Ruptura	Etapa 1 – Pergunta de Partida Etapa 2 – A exploração Etapa 3 – A problemática
Construção	Etapa 4 – A construção do modelo de análise
Verificação	Etapa 5 – A observação Etapa 6 – A análise das informações Etapa 7 - As conclusões

Quadro 1: Etapas do procedimento metodológico

Fonte: Quivy e Campenhoudt (1992)

Frente ao modelo orientado por Quivy e Campenhoudt (1992), é pertinente ressaltar que há incipiência de erros de três magnitudes produzidos no decorrer da Etapa 4 – A construção do modelo de análise, que retrata os procedimentos metodológicos em que a pesquisa é submissa:

- Erros incipientes dos instrumentos de coleta de dados: escolha por um instrumento que não atende ao propósito da pesquisa e/ou um instrumento não fidedigno;
- Procedimentos falhos na coleta de dados: amostra pequena demais, estratificação incorreta da amostra, equipamento descalibrado ou com defeito;
- Análise incorreta das informações: armazenamento incorreto dos dados, não eliminar os dados que estão em desacordo com as condições de estudo previamente estabelecidas, não eliminar os *outliers*.

A qualidade de um instrumento de avaliação é determinada por variáveis denominadas propriedades psicométricas, dentre as quais se destacam a confiabilidade e a validade. A confiabilidade é a capacidade de um instrumento medir fielmente um fenômeno. A validade é a capacidade de um instrumento medir com precisão o fenômeno a ser estudado, como pode ser observado na figura 1:

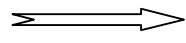
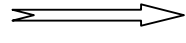
Confiabilidade		Medir fielmente
Validade		Medir com precisão

Figura 1: Distinção entre confiabilidade de validade de instrumentos de avaliação

Na seção Metodologia do relatório de pesquisa, em que são apresentados os instrumentos de avaliação, devem-se incluir informações sobre a confiabilidade e validade de cada instrumento escolhido, em relação a cada uma das variáveis estudadas. Estas informações estão contidas em estudos que tratam do tema.

Quando existe um número significativo de estudos utilizando determinado instrumento, usualmente o pesquisador retém somente aqueles que tratam de populações similares à população alvo. Quando um teste é pouco utilizado ou recente, os documentos sobre suas qualidades métricas são escassos. Mais improvável ainda é encontrar informações psicométricas aplicadas a uma população similar à população alvo.

Neste caso, no projeto de pesquisa faz-se necessária a indicação dos procedimentos para se apreciar a qualidade do instrumento. Para tanto, podem-se utilizar resultados preliminares e/ou incluir no projeto um estudo prévio para pré-testar os instrumentos de medida, cuja qualidade é desconhecida.

No momento das definições metodológicas, o pesquisador, com frequência, tem de optar entre a utilização de um instrumento de avaliação existente e de qualidade conhecida, mas que necessariamente não responde aos objetivos específicos do estudo, ou construir um novo instrumento perfeitamente adaptado à proposta da pesquisa, mas cuja qualidade deve ser demonstrada.

Quando a opção for pela elaboração de um novo instrumento de medida ou pela utilização cuja validade não foi avaliada, deve ser precisado como será realizado o pré-teste. O estudo prévio, tal qual o exame de pesquisas que utilizam determinado instrumento, serve para estabelecer a confiabilidade e a validade dos instrumentos a serem utilizados.

A escolha pela utilização de um instrumento existente ou pela construção de um novo instrumento muitas vezes constitui um entrave para o pesquisador. Frente a esse embate, é pertinente analisar alguns fatores que permeiam a decisão:

	UTILIZAÇÃO DE UM INSTRUMENTO EXISTENTE	CONSTRUÇÃO DE UM NOVO INSTRUMENTO
Vantagens	<ul style="list-style-type: none"> - As qualidades métricas do instrumento são conhecidas; - Permite comparar os resultados obtidos junto a outras populações, em outros momentos; - Reduz o cronograma do projeto; - Permite realizar a pesquisa com um orçamento menor. 	<ul style="list-style-type: none"> - Permite medir exatamente as variáveis estudadas.
Desvantagens	<ul style="list-style-type: none"> - Pode não medir exatamente o que se quer; - Demanda frequentemente modificações, o que complica a apreciação da qualidade; - As qualidades métricas podem ter sido estabelecidas sobre populações muito diferentes das do estudo. 	<ul style="list-style-type: none"> - Demanda tempo e recursos; - Atrasa o início do estudo; - Precisa da preparação de um estudo prévio para pré-testar o instrumento.

Quadro 2: Comparativo de vantagens e desvantagens associadas à escolha de um novo instrumento

Fonte: Contandriopoulos et al. (1999)

A opção pela construção de um novo instrumento de avaliação acarreta ao pesquisador a responsabilidade de garantir que este, factualmente, apresente características psicométricas satisfatórias, que lhe atribuam a fidedignidade dos resultados do estudo, ou, em contrapartida, consentir com o risco de se obter um resultado equivocado.

Não obstante ao risco descrito, a maior parte dos instrumentos de avaliação concebidos em trabalhos acadêmicos de mestrado e doutorado não são pré-testados previamente à sua aplicação final, perfazendo com que suas característica psicométricas permaneçam desconhecidas.

Confiabilidade de um Instrumento de Avaliação

De acordo com Contandriopoulos *et al.* (1999, p. 80), “A confiabilidade de um instrumento de medida é a sua capacidade em reproduzir um resultado de forma consistente no tempo e no espaço, ou com observadores diferentes quando for utilizado corretamente.”.

Em se tratando de aplicações em seres humanos, trata-se da capacidade como um instrumento de avaliação capaz de mostrar que indivíduos estáveis apresentem resultados

semelhantes em administrações repetidas do mesmo instrumento, sendo o mesmo reaplicado após um determinado período de tempo (BELASCO; SESSO, 2006).

Em termos práticos, a apreciação da confiabilidade reside na repetição da medida e na comparação dos resultados obtidos. Objetos de observação muito estáveis devem produzir medições similares em diferentes observações. Fenômenos transitórios podem produzir resultados com variação entre uma medida e outra, sem que isto se deva ao instrumento de medida. Com efeito, um instrumento distorcido de forma sistemática (sem ser ao acaso) pode ser perfeitamente confiável, apesar da falta de validade.

Para Contandriopoulos *et al.* (1999, p. 80), existem três abordagens principais para avaliar a confiabilidade de um instrumento:

1. A comparação dos resultados obtidos pela utilização de um mesmo instrumento, em diferentes momentos, para avaliar sua estabilidade;
2. Apreciação da equivalência dos resultados obtidos, quando um mesmo fenômeno é medido por vários observadores ao mesmo tempo;
3. Quando um instrumento é composto por vários itens ou indicadores, sua confiabilidade pode ser apreciada medindo a homogeneidade de seus componentes.

A escolha dos testes estatísticos utilizados para medir a confiabilidade de um instrumento depende do método de validação previsto e do tipo de dados produzido pela variável medida pelo instrumento (contínua, ordinal ou nominal). Sendo uma característica independente da questão de pesquisa, apesar de poder variar em função das populações estudadas.

O estimador de confiabilidade mais utilizado, segundo se trate de índice ou não, de variáveis contínuas ou de variáveis categóricas, são:

- O coeficiente de correlação de Pearson;
- O coeficiente alfa de Cronbach;
- O coeficiente de correlação intraclasse;
- O coeficiente Kappa de Cohen.

Na literatura existem normas de aceitação de valores de confiabilidade. Contandriopoulos *et al.* (1999) consideram que a aceitabilidade de um valor de confiabilidade depende da questão de pesquisa, devendo todas as indicações sobre a confiabilidade dos instrumentos utilizados estar contidas no projeto. A ausência de confiabilidade produzir consequências graves sobre as validades das conclusões estatísticas da pesquisa.

Consistência interna de um Instrumento de Avaliação

Desenvolvido por Lee Cronbach em 1951, o coeficiente alfa de Cronbach é uma ferramenta estatística que avalia a confiabilidade através da consistência interna de um questionário. Para a utilização do coeficiente alfa de Cronbach, é requisito que todos os itens do instrumento utilizem a mesma escala de medição (FREITAS; RODRIGUES, 2005).

A análise de consistência interna é pautada na correlação dos itens de um mesmo constructo entre si e na correlação de cada item com o escore total deste constructo. É desejável que os itens se apresentem moderadamente correlacionados entre si, e, também, que cada item apresente correlação com o escore total do constructo.

O alfa de Cronbach é obtido a partir da variância dos componentes individuais e da variância da soma dos componentes de cada avaliado, buscando investigar as possíveis relações entre os itens (CRONBACH, 1951). Dessa forma, as variáveis utilizadas no cálculo do coeficiente de Cronbach são: o número de questões do instrumento (K), a variância de cada questão (S_i^2) e a variância total do instrumento (S_t^2). O coeficiente alfa de Cronbach é calculado através da seguinte equação:

$$\alpha = \left(\frac{K}{K-1} \right) * \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right)$$

Figura 2: Cálculo do coeficiente alfa de Cronbach

Fonte: Cronbach e Shavelson (2004)

É importante salientar que, mesmo sendo amplamente utilizado em diversas áreas do conhecimento, ainda não existe um consenso quanto à avaliação do coeficiente alfa de Cronbach. Algumas literaturas consideram satisfatório um instrumento de pesquisa que obtenha $\alpha = 0,70$. Entretanto, cabe ao pesquisador decidir qual o valor mínimo de consistência interna aceitável para o seu respectivo instrumento (CRONBACH; MEEHL, 1955).

Freitas e Rodrigues (2005) propõem a seguinte escala para análise do coeficiente alfa de Cronbach:

Valor de α	Consistência
$\alpha \leq 0,30$	Muito baixa
$0,30 < \alpha \leq 0,60$	Baixa
$0,60 < \alpha \leq 0,75$	Moderada
$0,75 < \alpha \leq 0,90$	Alta
$\alpha > 0,90$	Muito alta

Tabela 1: Escala de consistência interna do coeficiente alfa de Cronbach

Fonte: Freitas e Rodrigues (2005)

De acordo com Cronbach e Shavelson (2004), os fatores que podem influenciar na consistência interna de um questionário são:

- Número de itens: um questionário com um extenso número de itens pode ocasionar respostas impulsivas e relapsas, além de tender a aumentar o número de itens sem resposta;
- Tempo de aplicação do questionário: limitar a um intervalo de tempo pré-estabelecido também pode ocasionar os mesmos problemas descritos no item anterior. É comumente observado que, devido à falta de tempo para o preenchimento do questionário, as últimas questões não sejam respondidas;
- Amostra de avaliados: aplicar um questionário a uma amostra muito semelhante deve reduzir a confiabilidade de um questionário, pois, quanto mais homogênea a amostra, a variância tende a se tornar nula.

Validade da Medida

Validade é um conceito com desenvolvimento teórico menos evidente que o conceito de confiabilidade. O conceito é muito mais abstrato, sendo, em medida significativa, dependente do contexto de utilização do instrumento. Com efeito, são pouco usuais publicações que demonstrem a validade dos instrumentos. As avaliações pertinentes são as realizadas em contextos de estudos similares ao estudo proposto. Segundo Belasco e Sesso (2006), a validação consiste em analisar se um instrumento de avaliação é capaz de medir ou avaliar aquilo a que se propõe a medir/avaliar.

Para Contandriopoulos *et al.* (1999), validade é a capacidade de um instrumento medir o fenômeno em exame, isto é, a adequação existente entre as variáveis escolhidas e o conceito teórico a ser medido. O desenvolvimento teórico inferior em relação ao conceito confiabilidade deve-se à impossibilidade desta adequação ser estimada diretamente por uma fórmula matemática que produza um coeficiente único e geral de apreciação da validade.

Existem três tipos de validade:

1. A validade de conteúdo;
2. A validade prática ou de critério;
3. A validade de construção.

A validade de conteúdo é determinada com o julgamento da proporção em que os itens selecionados para medir uma construção teórica representam todas as facetas importantes do conceito medido. A validade de conteúdo inclui também a validade aparente do instrumento, isto é, a coerência aparente entre o que se quer medir e o instrumento de medida escolhido.

Os procedimentos normalmente empregados para avaliar a validade de conteúdo são o parecer de juízes, especialistas são requisitados para emitirem parecer sobre a adequação aparente entre o instrumento proposto e a construção a ser medida, ou o julgamento de validade de conteúdo do instrumento por parte de alguns membros da população alvo.

Para Contandriopoulos *et al.* (1999, p. 83), um dos aspectos mais importantes da validade de conteúdo, no caso da construção de questionários, é a formulação das questões. A utilização de jargões complexos, e por extensão de difícil entendimento, afetam diretamente a validade de conteúdo do instrumento.

A validade prática, ou validade de critério, tem relação com a capacidade do instrumento em medir elementos relacionados com um critério de interesse, frequentemente um comportamento, ou seja, coloca em relação a medida obtida por um critério empírico do resultado. Quando o critério situa-se temporalmente no futuro, tem-se a validade preditiva, quando é contemporânea, tem-se a validade concorrente.

De acordo com Barros (2002), a validade concorrente é aplicada quando já se existe um instrumento conhecido e validado para se mensurar a variável em exame. Para tanto, deve-se aplicar simultaneamente o instrumento já validade e o que se deseja validar, esperando que os dados obtidos em ambos os instrumentos apresentem uma correlação significativa entre si.

Já a validade preditiva utiliza-se da predição de um comportamento futuro através de uma equação matemático. Assim, o instrumento apresenta validade preditiva se os dados obtidos através da aplicação deste estão significativamente correlacionados com uma característica, comportamento ou desempenho a ser apresentado no futuro (BARROS, 2002).

A validade de construção ou de constructo refere-se à relação entre os conceitos teóricos e sua operacionalização. Este tipo de validação tem relação epistêmica que deve existir entre um conceito e sua medida. Segundo Barros (2002), é necessário aplicar o instrumento a ser validado a

dois grupos isolados, que apresente uma significativa distinção com relação à variável examinada. É requestado que haja uma significativa diferença entre os resultados de ambos os grupos.

É cabível ressaltar que (CONTANDRIOPOULOS *et al.*, 1999, p. 86),

“Os indicadores de validade são raramente unívocos. As evidências de validade se acumulam no decorrer das pesquisas nas quais se utilizou o mesmo instrumento e, é somente depois de um longo período, que as publicações começam a mencionar esta validade”.

Considerações Finais

O processo de testagem para as propriedades psicométricas de um instrumento de avaliação deve ser metuculoso e judicioso. A tríade confiabilidade-consistência interna-validação deve ser fazer presente, tendo em vista que a exclusão de qualquer destas características pode comprometer a fidedignidade do instrumento.

O erro produzido por instrumentos inadequados, ou seja, sem qualidades psicométricas satisfatórias, é um erro evitável. É, na perspectiva de Demo, um ataque contra a opção de determinado método nos planos teóricos e práticos, mas, sobretudo um erro de forma. Em termos práticos, a qualidade dos instrumentos de avaliação, por si só, não garante uma boa pesquisa, mas numa fase extremamente importante, a construção, garante a possibilidade do pesquisador ter um trabalho publicável.

Referências

BARROS, M. V. G. Construção e validação de instrumentos: o que é um bom teste? **Página pessoal do Prof. Mauro Barros**, 05 mar. 2002. Biblioteca. Disponível em:

<www.maurovgb.hpg.com.br/docs/biblioteca/pdf/texto2.pdf>. Acesso em 26 mai. 2010.

BELASCO, A. G. S.; SESSO R. C. C. Qualidade de vida: princípios, focos de estudo e intervenções. In: DINIZ, D. P.; SCHOR N. (Org.). **Guia de qualidade de vida**. São Paulo: Manole, 2006. p. 1- 10.

CONTANDRIOPOULOS, A.-P. et al. **Saber preparar uma pesquisa**. 3. ed. São Paulo: Hucitec; Rio de Janeiro: Abrasco, 1999.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. **Psychometrika**, Greensboro, NC, v. 16, n. 3, p. 297-334, set. 1951.

CRONBACH, L. J.; Meehl, P. E. Construct validity in psychological tests. **Psychological Bulletin**, Washington, v. 52, n. 4, p. 281-302, jul. 1955.

CRONBACH, L. J.; Shavelson, R. J. My current thoughts on Coefficient alpha and successor procedures. **Educational and Psychological Measurement**, Thousand Oaks, v. 64, n. 3, p. 391-418, jun. 2004.

DEMO, P. **Educar pela pesquisa**. 2. ed. Campinas: Autores Associados, 1997.

FREITAS, A. L. P., RODRIGUES, S. G. A avaliação da confiabilidade de questionário: uma análise utilizando o coeficiente alfa de Cronbach. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 12., 2005, Bauru. **Anais...** Bauru: UNESP, 2005.

IOANNIDIS, J. P. A. Why most published research findings are false. **PLoS Medicine**, Cambridge-UK, v. 2, n. 8, p. 696-701, aug. 2005

QUIVY, R.; CAMPENHOUDT, L. Manual de investigação em ciências sociais. Lisboa: Gradiva, 1992.

TURATO, E. R. **Tratado da metodologia da pesquisa clínico-qualitativa**: construção teórico-epistemológica, discussão comparada e aplicação nas áreas da saúde e humanas. Petrópolis, RJ: Vozes, 2003.

Luiz Alberto Pilatti. Universidade Tecnológica Federal do Paraná, Campus Ponta Grossa. Professor do Programa de Pós-Graduação em Ensino de Ciência e Tecnologia. Bolsista de Produtividade em Pesquisa do CNPq. Doutor em Educação Física. lapilatti@utfpr.edu.br.

Bruno Pedroso. Faculdades Integradas de Itararé. Professor do Curso de Licenciatura em Educação Física. Doutorando em Educação Física. brunops3@brturbo.com.br.

Gustavo Luis Gutierrez. Universidade Estadual de Campinas. Professor do Programa de Pós-Graduação em Educação Física. Doutor em Administração. Fundação Getúlio Vargas. glgutierrez@terra.com.br.