

## Medidas de atitudes

José Pastore

### 1. A NATUREZA DAS ATITUDES

O estudo das atitudes vem ocupando um lugar cada vez mais central na psicologia social e sociológica em geral. O conceito e os procedimentos de medida, entretanto, ainda variam muito na literatura das ciências sociais. Thurstone, um dos primeiros especialistas a se dedicar ao estudo das atitudes, definiu atitude como "a soma total das inclinações e sentimentos do homem, preconceitos, noções, idéias preconcebidas, medos, ameaças, e convicções a respeito de um tópico específico" (Thurstone, 1923, p.216). Desse modo, uma atitude em relação ao negro significaria para aquele autor tudo o que o indivíduo sente, pensa e valoriza a respeito de relações entre brancos e pretos.

Thurstone fazia uma nítida distinção entre atitude e opinião. O conceito de opinião significava para ele a expressão verbal da atitude; portanto a opinião seria a expressão verbal dos sentimentos, preconceitos, valorizações e convicções. A opinião, assim entendida, configurava-se como uma simbolização de atitudes e, portanto, o veículo crucial para o estudo da atitude. Assim, já em 1922, Thurstone levantava o problema da dissociação entre opinião, atitude e comportamento manifesto. De fato, sabe-se modernamente que os três fenômenos não ocorrem juntos necessariamente: o comportamento efetivo pode se constituir em distorção da atitude do indivíduo. Muitas vezes a opinião e o comportamento efetivo são contrários a suas atitudes a respeito do objetivo em questão. Outras vezes, pessoas possuidoras de atitudes diferentes comportam-se identicamente em relação a um objeto. (Fisabein, 1967, D. 477). Apesar dessa dissociação, a metodologia moderna continua desenvolvendo esforços no sentido de medir e compreender as atitudes através das verbalizações individuais deixando claro, porém, que a medida e compreensão das atitudes não significa uma previsão do comportamento individual. Então, para que se estudam as atitudes?

A resposta à pergunta acima é de grande relevância teórica e pode ser resumida da seguinte maneira: O conhecimento da atitude de uma pessoa em

relação a um objeto pode ser utilizada em conjunção com variáveis estruturais e situacionais para prever e explicar o comportamento daquela pessoa em relação àquele objeto ou classe de objetos.

Em outras palavras, o conhecimento das atitudes – embora insuficiente para prever e explicar comportamentos – constitui um poderoso auxiliar para a previsão e explicação de reações humanas. O leitor familiarizado com os problemas teóricos das ciências sociais sabe que o comportamento humano é determinado por um conjunto de forças bastante complexo, dentre elas, as atitudes. Portanto, o estudo das atitudes assume importância na medida em que se utilizam seus resultados em combinação com outras variáveis para a compreensão do comportamento humano. Conhecendo-se os princípios que governam a mudança de atitudes, isto ajuda na manipulação das reações individuais em terapia e educação.

Os conceitos surgidos depois dos trabalhos de Thurstone constituíram importantes passos para especificar melhor definição nominal de atitude. Shaw e Wright apresentam um inventário bastante útil das diferentes maneiras de se definir atitudes de Thurstone a nossos dias (Shaw e Wright 1967, pp 1-14). Em um artigo geral como este convém ressaltar apenas algumas definições mais freqüentes e utilizadas no trabalho empírico de hoje.

English e English definem atitude como “uma predisposição (aprendida) para se comportar de um modo consistente em relação a uma determinada classe de objetos”. (in Shaw e Whight, 1967, p.3). Krech e colaboradores apresentam uma definição em 1962 que vem sendo largamente utilizada atualmente: “Atitude é um sistema de valorizações positivas ou negativas de sentimentos emocionais e tendências pró e contra em relação a um objeto social” (in Shaw e Wright, 1967, 1967, p.3). Anderson e Fishbein definiram atitude em 1965 como “a soma de crenças a respeito de um objeto e o aspecto valorativo destas crenças” (in Shaw e Wright, 1967, 1967, p.3).

As definições acima mencionadas sugerem que uma atitude é um fenômeno composto de componentes cognitivos, valorativos e afetivos, embora haja autores que preferem medir e examinar mais um dos componentes do que

outros. De qualquer forma, sendo a atitude um fenômeno complexo e de relativa estabilidade, não há possibilidade de mensuração através de indicadores simples. Em outros termos, o exame da complexidade do conceito de atitude sugere a utilização de indicadores combinados para medir a atitude de um indivíduo em relação ao preconceito racial, capitalismo, militarismo, populismo e outros objetos sociais.

## 2. REQUISITOS BÁSICOS PARA A MEDIDA DAS ATITUDES

Apresentamos nesta seção certos problemas básicos que o investigador deve resolver antes de começar a se preocupar com as técnicas de mensuração propriamente ditas. Destacaremos com maior ênfase as questões relativas a: a) Conceituação; b) Definição de um continuum.

### a) CONCEITUAÇÃO

É fato evidente que não se pode medir o que não se sabe o que é. Isso significa que a primeira tarefa do investigador interessado em medir determinada atitude é defini-la. Um exaustivo trabalho de **definição nominal** se impõe antes que o pesquisador comece a se preocupar com os indicadores que comporão a sua **definição operacional** do fenômeno (Rapoport, 1965, cap. I). Assim, se o investigador deseja estudar atitudes em relação à mudança social, a ele competirá definir precisamente o que entende por mudança social. É evidente que tal definição terá grande apoio na literatura sociológica existente e, ainda, poderá ser especificada para situações particulares.

Ao definir o objeto social é sempre aconselhável restringir o conceito a uma dimensão apenas (quando é possível). Um fenômeno como mudança social, por exemplo, apresenta muitas dimensões e, por conseguinte, deve ser restringido a aspectos particulares para fins de mensuração inicial. A construção de escalas multidimensionais em sociologia constitui tarefa bastante complexa e os resultados até agora obtidos podem ser considerados apenas como exploratórios (Torgerson, 1965, Cap. I).

O investigador familiarizado com as técnicas de construção de escalas conhece bem a importância de definições nominais claras e sucintas. Sem elas é impossível selecionar-se os itens que **de fato** tenham relevância para medir o fenômeno em estudo. O pesquisador não familiarizado, por outro lado, tende a negligenciar o trabalho de definição nominal, partindo imediatamente para o arrolamento de itens que, segundo ele, medem a atitude em foco. Como conseqüência, tais procedimentos geralmente terminam em fracasso pois nem o investigador, nem o informante, e “nem o computador” sabem o que os itens pretendem medir.

A título de ilustração, admitamos que o pesquisador está interessado em estudar atitudes em relação ao fenômeno do cosmopolitismo. A primeira tarefa a cumprir será então a de elaborar uma definição nominal de tal fenômeno. Como passo inicial pode-se dizer que cosmopolitismo é o grau em que o indivíduo está orientado para fora de seu sistema social. Após esta primeira aproximação verificamos que o conceito é bastante geral e, portanto, pode ser medido em várias dimensões ou situações específicas. Suponhamos que o investigador eleja, como uma das dimensões, cosmopolitismo em relação à política. Neste caso, poderíamos restringir o conceito para “o grau em que o indivíduo está orientado para fora do seu sub-sistema político”. Esta definição, ainda precária, constitui elemento fundamental para a escolha de itens que indiquem em que medida o ator conhece e valoriza as pessoas, posições e relações políticas fora de seu sistema.

#### b) DEFINIÇÃO DE UM CONTINUUM

A idéia de escala em si implica na existência de um continuum. Assim, quando utilizamos o metro ou a balança estamos admitindo um continuum de estaturas e pesos. Da mesma forma, quando utilizamos uma escala para medir atitudes em relação ao cosmopolitismo estamos admitindo um continuum de posições que varia do extremo mais desfavorável ao mais favorável, passando-se por inúmeros pontos intermediários.

Teoricamente, o pesquisador pode utilizar todos os pontos do continuum. Na verdade, um instrumento da medida é considerado **preciso** quando é capaz de detectar todas as diferenças entre os diversos pontos do continuum. Entretanto, nem sempre se consegue ou se precisa de instrumentos com tão alta precisão. Por exemplo, o metro utilizado pelo construtor de casas não necessita de precisão além de centímetros ou milímetros. Assim, a precisão requerida para os instrumentos de medidas de atitudes depende do tipo de atitude que está sendo medida, da ambição do pesquisador e das medidas previamente elaboradas.

De modo geral, os sociólogos e psicólogos sociais contentam-se em examinar apenas alguns pontos do continuum na escala; comumente, medem-se as posições extremas e alguns pontos intermediários. Geralmente as escalas baseiam-se apenas nas posições “fortemente favorável”, “favorável”, “neutro”, “desfavorável” e “fortemente desfavorável”.

Nestas posições atitudinais, quando colocadas em uma linha devem ser eqüidistantes. Somente neste caso podemos falar em escalas intervalares. Graficamente, as 5 posições acima descritas colocar-se-iam da seguinte maneira:

Os problemas para a determinação do continuum de uma escala estão intimamente relacionados com a questão do isomorfismo. Uma escala é isomórfica quando as observações realizadas e os números de escala a elas associados têm uma relação tão direta que pela simples manipulação dos números, o pesquisador obtém uma nova informação a respeito dos fenômenos compreendidos a observação. Na apresentação gráfica acima, a escala será isomórfica se pudermos dizer, de fato, que os indivíduos colocados ao extremo direito são 2

vezes mais favoráveis do que os que se localizam na posição 1. Portanto, trata-se de estabelecer uma correspondência entre essa linguagem numérica e a realidade observada.

O fato das escaladas serem baseadas em um continuum dividido convencionalmente não implica que esta divisão seja arbitrária. Ao contrário, o estudo de diversas formas de medida e o cortejo dos resultados com trabalhos de outros pesquisadores devem ser sempre o guia para o estabelecimento da “melhor” unidade de medida ou dos intervalos isomórficos seguidos pela observação. A “melhor” escala é sempre aquela baseada em pressupostos formais que se ajustam matematicamente a um sistema de relações regidas por leis conhecidas (Nunaly, 1967, p.28).

### **3. VALIDADE DE UMA ESCALA**

Validade é o grau segundo o qual uma escala mede o que se pretende medir. A validade é um primeiro critério para se decidir sobre a utilidade de uma escala. Somente a escala válida é útil.

A validade de certos instrumentos é verificável facilmente; o metro mede o comprimento; a balança mede o peso; o termômetro, a temperatura. Estes instrumentos se ajustam aos conceitos de comprimento, peso e temperatura. Entretanto, há vários instrumentos de medida que não são tão facilmente verificáveis quando à validade e por isso o investigador deve realizar certos procedimentos empíricos para estar certo que o seu teste mede inteligência, ou que sua escala mede realmente atitudes em relação ao modernismo, por exemplo.

#### **a) TIPOS DE VALIDADE**

##### **1) Validade de previsão**

O problema de se examinar a validade de previsão se coloca quando se pretende testar a capacidade de escala prever o fenômeno em estudo. Uma escala tem validade de previsão quando aquilo que ela mede aproxima-se significativamente daquilo que realmente acontece. Suponhamos que um pesquisador construa uma escala para medir atitudes em relação à obediência a

regulamentamos dos calouros de uma universidade. Ele aplica o instrumento e avalia os resultados. Em seguida ele anota no decorrer de seus anos de universidade e compara estes resultados com o da escala. Esta terá validade de previsão se aquilo que foi inicialmente obtido aproxima-se do comportamento posterior.

Todo o elemento que é usado como ponto de comparação com uma escala é chamado critério. No exemplo acima, portanto, o comportamento e as verbalizações dos estudantes constituem o “critério” para a escala imaginada.

Como se observa, a verificação e a validade de previsão constituem um procedimento restritamente empírico. Para análise de aproximação ou coincidência entre a escala e o critério é comum o uso de testes não paramétricos (por exemplo,  $\chi^2$ ) e correlação.

Convém lembrar ao leitor que este tipo de validade é dificilmente alcançável em altos níveis nas escalas de atitudes quando se utiliza o comportamento efetivo como critério de comparação. Como vimos, atitude é apenas um dos vários determinantes do comportamento humano.

## 2) Validade de conteúdo

Como o próprio nome indica, uma escala tem validade de conteúdo se ela mede o conteúdo que esta expresso na variável em estudo. Este tipo de validade é especialmente importante para escalas de atitudes e testes educacionais. Por exemplo, se um determinado curso pretende ensinar história do Brasil, o educador estará interessado em um teste capaz de avaliar o aluno em história do Brasil e não em historia da Índia. Se o teste possuir tal característica, ele tem validade de conteúdo.

Refinar a validade de conteúdo em uma escala de atitudes é uma tarefa simultaneamente necessária e difícil. Por exemplo, uma escala que pretende medir autoritarismo terá validade de conteúdo na medida em que mede apenas autoritarismo.

Para o teste deste tipo de validade é extremamente importante o check constante dos itens com a definição nominal da variável em estudo. Se a definição

nominal confunde, por exemplo, autoritarismo com tradicionalismo e conservadorismo, não se pode exigir da escala uma validade de conteúdo que se restrinja o autoritarismo.

O estudo da validade de conteúdo exige técnicas mais sofisticadas como, por exemplo, análise fatorial. O uso de correlação e testes de contingência são aconselháveis somente no caso em que haja prévios estudos e escalas com esse tipo de validade.

### 3) Validade concorrente

Um outro aspecto da validade pode ser examinado quando já existe algum procedimento de medir a variável em estudo. Este procedimento pode ser uma outra escala já aferida e considerada válida ou a observação direta e indireta do informante. Em psicologia é comum examinar-se a validade decorrente de um novo teste para diagnóstico com as observações do psicólogo que entrevista o paciente.

No caso de atitude, uma escala tem validade concorrente quando aquilo que ela mede se aproxima ou coincide com o resultado de outras escalas válidas ou observações de comportamentos. O que se pretende é obter alguma segurança de que, comparando-se com outros critérios, a escala realmente mede o que pretende medir. Testes de correlação e análise de contingência são os procedimentos indicados para a realização dos testes de validade concorrente.

### 4) Validade de construto

Este é o tipo de validade especialmente importante para o trabalho de teorização. Suponhamos que o pesquisador tenha um conceito para qual pretende desenvolver um instrumento de medida mais seguro. O referido conceito faz parte de uma teoria e existem resultados técnicos envolvendo o conceito, embora com os outros tipos de mensuração. Por exemplo, suponhamos que o investigador esteja estudando alienação e conheça várias correlações entre alienação e outros fatores psico-sociais. Neste caso, ele formula uma escala para medir alienação e, aplicando em determinados informantes, analisa se aquelas correlações se

confirmam ou não. Quando as correlações não se confirmam, o teste deve ser alterado, ou o conceito modificado. Raros são os casos em que se deve modificar a teoria existente à partir de resultados obtidos com instrumentos de medida em elaboração.

Como se verifica, a validade de construto é uma característica bem mais complexa do que as anteriores apresentadas. Em geral, os procedimentos técnicos envolvem várias invocações e repetições de experimentos ou surveys. Praticamente, tais procedimentos e técnicas de correlação são aplicados somente após a verificação dos 3 tipos de validade anteriormente expostos.

## b) MÉTODOS DE DETERMINAÇÃO DA VALIDADE

A validade de uma escala pode ser determinada através de vários métodos. A literatura especializada em construção de escalas refere-se frequentemente a 3 métodos. (Havens 1965, p.58)

### 1) Grupos conhecidos

Quando se conhecem grupos que possuem opiniões contrárias a respeito do conceito em estudo, a escala é aplicada a esses grupos comparando-se os resultados. Por exemplo, se conhecemos indivíduos que tenham demonstrado repetidas vezes opiniões contrárias a respeito de liberdade sexual, poderíamos utilizar esses grupos como critérios para examinar a validade de uma escala de liberalismo sexual.

É preciso prevenir o leitor, porém, que muitas variáveis intervenientes podem mascarar o teste, o que exige vários controles. Muitas vezes, as opiniões a respeito de liberdade sexual variam devido às diferenças de idade, sexo, status social, etc. Por essa razão, tais fatores devem ser controlados mantendo-os constantes nos 2 grupos (tanto quanto o possível).

### 2) Método do júri.

Neste caso, compara-se a escala em construção com as opiniões de um grupo de “especialistas” no assunto ou pessoas bem informadas sobre o conceito

em questão. Dá-se a esses indivíduos uma definição do conceito e solicita-se que classifiquem certos indivíduos em ordem, segundo aquela dimensão. Posteriormente, aplica-se a escala aos indivíduos classificados e faz-se a comparação.

O grande cuidado a tomar neste método é o de fornecer definições de significado idêntico para os juizes e pesquisadores.

### 3) Critérios independentes

Outro método bastante comum é o de se correlacionar a escala com instrumentos existentes, cuja validade é conhecida. Este método, embora bastante adequado, é raramente usado, uma vez que existindo um instrumento válido, não há grande interesse em criar outro. Do ponto de vista teórico, porém, há várias situações em que se justifica a elaboração de novos instrumentos de medida. Às vezes, a população não tem condições de responder aos estímulos do instrumento existente e, por isso, requer a elaboração de nova escala para medir o mesmo conceito. Outras vezes são as hipóteses da pesquisa que requerem modificações nos itens da escala para medir o mesmo conceito. Outras vezes são as hipóteses da pesquisa que requerem modificações nos ítems da escala o que, por seu turno, exige nova validação.

### C) Técnicas de validação

O exame da validade pode ser feito através de várias técnicas.

As técnicas estatísticas mais usadas são os coeficientes de correlação que medem a magnitude da relação existente entre a escala em construção e outros critérios. Como se observa, a existência de “critérios” aferidos constitui uma condição que limita bastante o teste de validade. Por isso, outras técnicas vem sendo utilizadas para contornar esse problema (Análise Fatorial).

Os dois coeficientes de correlação mais utilizados para validar uma escala são os coeficientes de correlação de Postes de Spearman e o Coeficiente de Pearson. O leitor interessado na fundamentação matemática desses coeficientes pode consultar os textos básicos de estatísticas. Nesta seção limitar-nos-emos a

apresentar dois exemplos de validação utilizando cada um dos coeficientes mencionados e, no final exporemos a técnica de análise de item e análise fatorial.

1) Coeficiente de Spearman.

Exemplificamos o uso deste coeficiente para testar validade de previsão. Neste caso o teste deve responder a pergunta do tipo: Esta escala é capaz de prever aquilo que ela diz medir?

A fim de ilustrar o uso dessa técnica, suponhamos que um investigador tenha aplicado duas escalas de 30 itens para medir atitudes em relação à obediência a regulamentos entre calouros de uma universidade e, posteriormente compara seus resultados com comportamentos efetivo para testar validade de seu instrumento. Para simplificar, vamos considerar apenas 10 universitários e o problema é o de se saber qual das duas escalas é o melhor instrumento de previsão. Consideremos ainda que cada item das escalas permitia 3 tipos de respostas com os seguintes valores: concordo (3), sem opinião (2) e discordo (1), Assim sendo, cada universitário poderia ter um escore geral na escala que variaria entre 30 e 90. Após a obtenção desse escore, os informantes seriam ordenados de 1 a 10. Suponhamos ainda que o comportamento obediente efetivo tenha sido obtido através do número de admoestações recebidas pelo estudante (durante 3 anos) como decorrência de infração dos regulamentos. A tabela I sumariza a situação descrita.

TABELA I  
DADOS PARA TESTE DE VALIDADE

ESTUDANTE	ESCALA I	ESCALA II	CRITÉRIOS	ORDEM DO CRITÉRIO	ORDEM NA ESCALA I	ORDEM NA ESCALA II
A	40	64	11	6	3	5
B	50	72	12	7	6	7
C	30	38	15	10	1	1
D	48	52	13	8	5	3
E	75	48	14	9	9	2
F	85	62	10	5	10	4
G	60	68	9	4	8	6
H	53	74	8	3	7	8
I	41	76	7	2	4	9
J	33	80	6	1	2	10

Para o cômputo do coeficiente de Spearman, o (Rho), utiliza-se a fórmula:

$$\rho = 1 - \frac{6(\sum D_i^2)}{N(N^2 - 1)}$$

Onde  $D_i$  = Diferença entre critérios e cada teste e N= número de indivíduos estudados. O cálculo de  $\rho$  para o exemplo acima é obtido através das seguintes operações:

CÁLCULO DE COEFICIENTE  $\rho$

ORDEM DO CRITÉRIO C (1)	ORDEM DA ESCALA I (2)	ORDEM DA ESCALA II (3)	$D_1 = (1)-(2)$ (4)	$D_1^2$ (5)	$D_2 = (1) - (3)$ (6)	$D_2^2$ (7)
6	3	5	3	9	1	1
7	6	7	1	1	0	0
10	1	1	9	81	9	81
8	5	3	3	9	5	25
9	9	2	0	0	7	49
5	10	4	-5	25	1	1
4	8	6	-4	16	-2	4
3	7	8	-4	16	-5	25
2	4	9	-2	4	-7	49
1	2	10	-1	1	-9	81

$$D^2 = 162$$

$$D^2 = 316$$

$$Ic = 1 - \frac{6(162)}{10(100-1)} = 1 - \frac{972}{990} = 1 - 0,97 = 0,03$$

$$Iic = 1 - \frac{6(316)}{10(100-1)} = 1 - \frac{1896}{990} = 1 - 1,90 = -0,90$$

No primeiro caso, o  $Ic = 0,03$  indica em si uma baixa correlação entre os resultados escalares e o critério. No segundo caso, o  $Iic = -0,90$  indica uma forte correlação entre a escala e o critério. O sinal negativo sugere que quando “mais obediente” o aluno se mostrou na escala, menos admoestações ele recebeu no decorrer do tempo. Portanto, a escala tem validade de previsão.

## 2) Coeficiente de Pearson

Ilustraremos o uso deste coeficiente para testar a validade concorrente.

Para tanto, suponhamos que Escala II do exemplo anterior tenha sido considerada como um instrumento aferido e válido e que o investigador necessita elaborar uma nova escala de obediência e que é, então, validada comprando-a com a anterior. O coeficiente de Pearson, embora mais trabalhoso, é de fácil computação através da seguinte fórmula:

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} = \frac{10 \times 346,21 - (634)(522)}{\sqrt{[10 \times 41852 - (634)^2][10 \times 29678 - (522)^2]}}$$

Suponhamos que os resultados obtidos na nova escala sejam os da TABELA abaixo e o nosso problema é verificar a validade de concorrência entre essa escala e a anterior considerada como critério:

#### DADOS PARA TESTE DE VALIDADE ATRAVÉS DO COEFICIENTE DE PEARSON

ESTUDA NTE	ESCALA II $x$	ESCALA III $y$	$x^2$	$y^2$	$xy$
A	64	48	4096	2304	3072
B	72	56	5184	3136	4032
C	38	20	1444	900	760
D	52	49	2704	2401	2548
E	48	52	2304	2704	2496
F	62	46	3844	2116	2852
G	68	49	4624	2401	3332
H	74	62	5476	3844	4548
I	76	64	5776	4096	4864
J	80	72	6400	5776	6080

$$\sum x = 634 \quad \sum y = 522 \quad \sum x^2 = 41852 \quad \sum y^2 = 29678 \quad \sum xy = 34624$$

O coeficiente  $r = 0,76$  indica que a nova escala mede aquilo que deve medir com grande acuidade e, portanto, pode ser considerada uma escala válida.

O leitor mais exigente, entretanto pode perguntar até que ponto o valor encontrado acima não é uma obra do acaso ou de flutuações amostrais e, em consequência, passa a exigir testes de significância para os coeficientes de correlação. De fato, esse é o procedimento aconselhável para garantia do investigador. Testes de significância para  $r$  são apresentados em textos estatísticos referentes a técnica de correlação (Blalock, 1960, p. 333) Gatti, Feres, 1975, p. 136).

### 3) Análise de Item

Na literatura de metodologia, são comuns as críticas ao uso de critérios para validação de uma escala (Nunnally, 1967, p.246- 250). Um dos problemas mais sérios é de se considerar determinados instrumentos ou observações como **critério válido**. Em segundo lugar, devemos destacar que o uso de critérios tem pouca generalidade. Assim, por exemplo, uma escala pode ter validade de previsão para medir obediência a regulamentos universitários, mas se mostrar completamente inadequada para a previsão de obediência a regulamentos da indústria. E finalmente, devemos ressaltar que é sempre muito difícil encontrar itens que apresentem alta correlação com o critério e baixa correlação entre si. Geralmente, os dois tipos de correlação ocorrem juntamente e, conseqüentemente, poucos itens satisfazem aqueles requisitos. Por isso, o investigador mais exigente vem usando a técnica de correlação com critérios apenas como uma primeira aproximação superficial ao problema da validação. O passo seguinte é análise do item.

A análise de itens consiste em examinar a validade de cada item em função das características internas da própria escala que está sendo construída. Os procedimentos mais utilizados para análise de itens são análise fatorial e correlações de itens com escore total.

### 4) Análise fatorial

A técnica de análise fatorial tem por função examinar os itens quanto ao seu conteúdo de modo que, uma vez realizada, o investigador tem condições de eliminar itens espúrios ao conceito central e pertencentes a outras dimensões não constantes do conceito. Portanto, a análise fatorial é usada como um procedimento de depuração de escala. Muitas vezes o investigador submete 100 itens a uma análise fatorial e verifica que apenas 20 se ajustam à dimensão do conceito; a escala, então, é constituída com apenas 20 itens.

A computação requerida pela técnica da análise fatorial é bastante laboriosa e não pode ser tratada nessa secção (ver Fruchter, 1954).

Convém salientar a esta altura, entretanto, algumas vantagens e desvantagens da análise fatorial para testes de validação. Uma das principais vantagens da análise fatorial que tal procedimento estatístico proporciona um modelo matemático que pode ser usado para descrever áreas homogêneas do mundo empírico. Utilizando-se das inter-correlações dos diversos itens, a análise fatorial “diz” ao investigador quais deles pertencem e quais não pertencem à escala em construção. Outra grande vantagem da utilização da análise fatorial para fins de validação é que tal técnica pode analisar um grande número de itens e encontrar semelhanças em conteúdo no ambiente interno da escala em construção. Na verdade, a análise fatorial tem se mostrado como uma técnica de validação bem mais refinada do que as anteriormente apresentadas na seção acima.

Uma das principais desvantagens da técnica refere-se à morosidade de computação. A análise fatorial encontra as dimensões da escala (fatores) através de procedimento laborioso de inter-correlação entre todos os itens e rotação de fatores no espaço trigonométrico. Felizmente, esta desvantagem tem sido atenuada graças aos recursos oferecidos pelo computador e pelos programas pré-fabricados (Cooley e Lohnes, 1962, cap.8). Outra desvantagem diz respeito àquilo que muitos investigadores chamam de subjetivismo-nominalista. Esse problema aparece no momento em que, encontradas certas dimensões, estatisticamente homogêneas, o investigador é requisitado em rotulá-las.

É frequentemente a discordância na rotulação de fatores constituídos dos mesmos itens.

De qualquer forma, a análise fatorial pode ser considerada como um procedimento auxiliar de validação e não como um teste definitivo.

##### 5) Correlação de item com escore total

Esse tipo de procedimento consiste em correlacionar cada item da escala com o escore total da mesma. O item que apresenta correlação próxima de zero deve ser examinado cuidadosamente, pois é muito provável que seu conteúdo nada tenha a ver com aquilo que a escala pretende medir ou, talvez, trata-se de

um item de baixo poder discriminador e, portanto, pouco preciso. Inversamente, os itens que apresentam correlações altas são geralmente os melhores itens e um conjunto deles forma, com freqüência, um instrumento válido.

O coeficiente de correlação mais utilizado para esse procedimento é o bi-serial, que relaciona uma variável contínua (escore total) com uma dicotomizada (por exemplo, “concordo” ou “discordo”).

A fórmula da computação desse coeficiente é:

$$r_{pqi} = \frac{\bar{x}_p - \bar{x}_q}{S_t} \sqrt{pq}$$

Onde:  $\bar{x}_p$  e  $\bar{x}_q$  são os escores médios na variável contínua dos indivíduos entre as suas categorias;

p e q são proporções de observações nas duas categorias da variável dicotômica;

$S_t$  é o desvio padrão de todos os escores na variável contínua.

Uma das desvantagens do coeficiente de correlação bi-serial é que, necessariamente, os itens devem admitir apenas duas alternativas de resposta: concordo-discordo, certo-errado, aceito-rejeito, etc.

Com itens que admitem mais de duas alternativas de resposta recomenda-se o uso do coeficiente de Pearson, já mencionado na secção anterior (\*).

(\*) Quando se correlaciona um item com o escore total através do Coeficiente de Pearson, devemos lembrar que o item é parte do total, o que ocasiona certa distorção na correlação. Esta espuriedade pode ser eliminada usando-se a seguinte fórmula:

$$r_1(y-1) = \frac{r_{y1}\alpha_y - \alpha_1}{\sqrt{\alpha_1^2 + \alpha_y^2 - 2\alpha_1\alpha_y r_{y1}}}$$

onde:  $r_{y1}$  = correlação do item 1 com o escore total y

$\alpha_y$  = desvio padrão do escore total

$\alpha_1$  = desvio padrão do item 1

$r_1(y-1)$  = correlação do item 1 com a soma de todos os itens exclusive o item 1. (Nunnaly, 1967, p.262).

Existem muitas outras técnicas estatísticas para a análise de item, como é o caso do coeficiente phi, tetracórico, etc. (Nunnaly, 1967, p.261)

#### 4. Fidedignidade da escala

Um instrumento de medida é considerado fidedigno quando mede os objetos de modo consistente. O metro que se dilata ou se contrai devido a variações de pressão e temperatura não pode representar resultado consistente e, por isso, não é fidedigno.

Existe uma relação estreita entre fidedignidade e validade de uma escala. Uma escala valida deve ser necessariamente uma escala fidedigna. Entretanto, uma escala fidedigna não é necessariamente valida. Por exemplo, uma escala de atitudes em relação a modernismo pode apresentar resultados consistentes quando aplicada várias vezes a um grupo de indivíduos; entretanto, essa escala pode não ser válida porque, na realidade, ela mede outros fenômenos que não são modernismo. Assim, do ponto de vista prático, podemos inferir fidedignidade de validade, mas não o oposto.

Do ponto de vista estatístico, sabe-se que a fidedignidade exerce um poder limitante sobre a validade: “A validade de uma medida nunca pode exercer a raiz quadrada da fidedignidade” (Chrombach, 1960, p. 132). Assim, se a fidedignidade de uma escala é 0,64, a sua validade não pode ser maior que 0,80.

Existem vários fatores que podem limitar a fidedignidade de uma escala. Muitos deles são devido a variações dentro da própria escala, ou seja, à amostra de itens selecionada para compor a escala. Um indivíduo sempre tem uma determinada probabilidade de “concordar” com cada item; se entrevistarmos infinitos indivíduos e itens, tais probabilidade poderiam ser empiricamente determinadas (na verdade, é possível determina-los no âmbito teórico). Porém, limitando o número dos itens em cada escala, introduzimos certa variação nos escores de item para item. (Nunnaly, 1967, p. 207).

Uma outra fonte de erro decorre das tentativas casuais que um indivíduo faz ao responder certos itens, embora tal problema não seja tão grave nas escalas de atitudes ( não há itens certos e outros errados). Além disso, muitos erros surgem devido à má interpretação dos itens, interferências psicológicas no entrevistado, erros de resposta, cansaço, interferência na aplicação da escala, etc. Devemos mencionar ainda erros decorrentes do subjetivismo do investigador ao codificar respostas dos itens que não são previamente codificados.

Do ponto de vista estatístico, faz-se uma nítida distinção entre erro sistemático e erro casual ( Ferguson, 1959, p.275). Instrumentos que superestimam ou subestimam determinado fenômeno de modo consistente produzem um erro sistemático. É o caso de um relógio que atrasa um minuto por dia. Uma escala de atitudes pode produzir erro sistemático após certo tempo de aplicação devido ao cansaço dos informantes.

Um erro casual, ao contrário, ocorre sem exibir uma tendência sistemática. Assume-se que o erro casual tende a desaparecer na medida em que a mensuração é feita inúmeras vezes e com um grande número de indivíduos.

Toda vez que se fala em erro, assume-se a existência de um valor verdadeiro; o erro seria a diferença entre o valor obtido pela mensuração e o valor verdadeiro. A obtenção desse valor verdadeiro nas ciências comportamentais é uma tarefa complexa. No estudo do comportamento humano não se pode, sequer, fazer grande número de observações do mesmo fenômeno. Além disso, o atributo medido pode mudar devido à intervenção do próprio investigador.

Por essa razão, a metodologia das ciências sociais é forçada a desenvolver técnicas para determinação de erro que independem de mensurações repetidas, adaptando-se o conceito de valor verdadeiro como a média de valores obtidos em mensurações infinitas (teoricamente). Convém mencionar que as técnicas apresentadas a seguir dizem respeito apenas aos erros casuais.

#### a) O Coeficiente de Fidedignidade

O estudo de fidedignidade resume-se, do ponto de vista estatístico, no exame da proporção da variância do atributo em estudo que é devida a erro de

mensuração. Consideremos uma escala composta de  $n$  itens. Cada item é sujeito a erro o que pode ser simbolizado como:

$$X_i = T_i + e_i$$

Onde  $X_i$  é o atributo medido pelo item,  $T_i$  é o valor verdadeiro e  $e_i$  é o erro. Na escala toda, portanto, nós podemos representar os atributos medidos da seguinte forma :

$$\sum_{i=1}^n X_i = \sum_{i=1}^n T_i + \sum_{i=1}^n e_i$$

Analogamente, a variância dos valores medidos será a soma da variância dos valores verdadeiros mais a variância dos erros de medida:

$$\sigma_x^2 = \sigma_1^2 + \sigma_e^2$$

O coeficiente de fidedignidade consiste em examinar a proporção

$$\rho_{xx} = \frac{\sigma_1^2}{\sigma_x^2}$$

Ou, para amostras:

$$\rho_{xx} = \frac{S_1^2}{S_x^2}$$

Quanto menos acurada é a mensuração, maior será a  $\sigma_x^2$  e menor o coeficiente de fidedignidade. Assim, se  $\sigma_x^2 = 400$  e a  $\sigma_1^2 = 360$

$$\rho_{xx} = \frac{360}{400} = 0,90$$

Esse resultado indica que 90% da variação da mensuração são devidas às variações do valor verdadeiro e apenas 10% são decorrentes de erros de medida.

Existem muitas variações para o cálculo coeficiente de fidedignidade; mas isso depende do tipo de método utilizado para estabelecer comparações entre os dados. Apresentamos a seguir os 3 métodos mais utilizados acompanhados de alguns comentários e instrumental estatístico.

## b) Métodos para Determinação da Fidedignidade

### 1) Teste-reteste

A mesma escala é aplicada duas vezes nos mesmos indivíduos em tempos diferentes, correlacionando-se os resultados através do coeficiente de fidedignidade acima descrito ou coeficiente de Pearson. Geralmente, considera-se a escala fidedigna quando o coeficiente atinge ao nível 0,85 ou mais.

Este método apresenta a desvantagem do chamado “efeito educacional”, isto é, a primeira aplicação da escala “ensina” muita coisa ao informante e acaba afetando os resultados e a segunda aplicação. Para atenuar este efeito procede-se, geralmente, a divisão casual da população em duas metades e a utilização de um grupo de controle.

Um outro problema apresentado por este método refere-se ao fato de que, entre o tempo 1 e o tempo 2 da aplicação, podem ocorrer modificações reais no tópico em estudo. Por exemplo, uma escala para medir atitudes políticas aplicadas nos mesmos indivíduos em 2 dias diferentes pode apresentar resultados inconsistentes devido a modificações na situação política.

### 2) Formas múltiplas

Neste caso constroem-se duas formas da mesma escala (com indicadores diferentes, mas de mesmo conteúdo) e se aplicam nos mesmos indivíduos a um só tempo. A correlação entre as duas escalas deve atingir a níveis igualmente altos (0,85 ou mais).

Este procedimento, embora mais rápido e econômico, apresenta a grande dificuldade de se confiar que, de fato os itens das duas escalas são de mesmo

conteúdo e igualmente válidos. Duas escalas podem medir coisas diferentes e ainda assim estarem altamente correlacionadas.

Se o investigador desejar testar, por exemplo, três formas da mesma escala para determinar a fidedignidade, ele deve se lembrar que, nesse caso, as intercorrelações deve ser idênticas de modo que

$$r_{12} = r_{13} = r_{23}$$

### c) Método das metades

Este procedimento é um dos mais eficientes e consiste em dividir, casualmente, a escala em duas metades sendo que cada uma delas é considerada uma forma do mesmo instrumento como no caso anterior. As duas formas são administradas aos mesmos indivíduos e a um só tempo. Os valores obtidos são correlacionados através do coeficiente de Spearman-Brown:

$$r_{SB} = \frac{2r}{1+r}$$

Onde  $r$  = coeficiente de Pearson entre as duas metades. Assim, se  $r = 0,80$ , então:

$$r_{SB} = \frac{(2)(0,80)}{1+0,80} = 0,89$$

O que indica um nível alto de fidedignidade.

Um dos problemas deste método é que para cada tipo de divisão da escala em metade obter-se à diferentes resultados para  $r_{SB}$ . No sentido de eliminar este problema Kuder e Richardson desenvolveram um coeficiente, kR-20, que produz um valor único para qualquer escala dividida dos mais variados modos (Nunnaly, 1967, p. 196).

$$KR - 20 = \frac{K}{K - 1} \left( 1 - \frac{\sum pq}{s_x^2} \right)$$

(Nunnally, 1967, p.196)

Onde k = número de itens de escala

$s_x^2$  = variância dos valores na escala

pq = produto da proporção de “concordo”-“discordo” em cada item.

**Exemplo:** Consideremos que uma escala de 10 itens tenha sido aplicada a 10 indivíduos e desejamos saber o grau de fidedignidade. Cada item apresenta 2 alternativas de respostas, isto é, concordo discordo.

ITEM	CONCORDO ( $x_i$ )	DISCORDO ( $x_i$ )	Pxi	Pyi	pq	$X_i^2$
A	5	5	0,5	0,5	0,25	25
B	6	4	0,6	0,4	0,24	36
C	7	3	0,7	0,3	0,21	49
D	5	5	0,5	0,5	0,25	25
E	8	2	0,8	0,2	0,16	64
F	6	4	0,6	0,4	0,24	36
G	8	2	0,8	0,2	0,16	64
H	9	1	0,9	0,1	0,09	81
I	4	6	0,4	0,6	0,24	16
J	6	4	0,6	0,4	0,24	36

K=10

64

pq=2,08

 $x_i^2 = 432$ 

$$KR - 20 = \frac{K}{K - 1} \left( 1 - \frac{\sum pq}{s_x^2} \right)$$

$$S_x^2 = \frac{\Sigma x^2 - (\Sigma x)^2/n}{n-1} = \frac{432 - 4096/10}{9} = \frac{432 - 409,6}{9} = \frac{22,4}{9} = 2,15$$

$$KR - 20 = (10/9)(1 - 2,08/2,15) = (1,1)(1 - 0,97) = (1,1)(0,03) = 0,033$$

Esse resultado indica um baixo nível de fidedignidade para a escala em questão.

Para as escalas compostas de itens com três alternativas, por exemplo, “concordo”, “discordo”, com pesos de 3,2 e 1 respectivamente, sendo  $p_1$ ,  $p_2$  e  $p_3$  as proporções de indivíduos que responderam cada uma das categorias.

(Nunnaly, 1967, p.196).

## **Bibliografia**

1. Blalock, Humbert M.- 1960- SOCIAL STATISTICS, New York, McGraw Hill- Book Co.
2. Cooley, William W. e Lohnes, Paul R. – 1962 – Multivariate Procedures for the behavioral Sciences, New York, John Wiley e Sons.
3. Cronbach, Lee J. – 1960 – Essentials of Psychological Testing, New York, Harper & Brothers.
4. Fergurson, George A. – 1959 – Statistical Analysis in Psychology and Education, New York, MacGraw Hill Book Co.
5. Fishbein, Martin – 1967 – “Attitude ande the Prediction of Behaavior” in Martin Fishbein, Readings in Attitude Theory and Measurement, New York: John Wiley & Sons.
6. Fruchter, Benjamim – 1954 – Introduction to Factor Analysis – New York, D. Van Nostrand Co.
7. Havens, A. Eugene e col, - 1965 – Medicion em Sociologia: Conceptos y Métodos, Editora da Universidade Nacional de Colômbia, Bogotá.
8. Nunnaly, Jum – 1967 – Psychometric Theory, New York – McGraw Hill Book Co.
9. Repaport Anatal – 1953 – Operational Philosophy: Integrating Knowledge and Action, New York, John Wiley & Sons.
10. Shaw, Marvin E. e Wright J.M. – 1967 – Scales for the Meassurement of Attitudes, New York – MCGraw Hill Book Co.
11. Thurstone, Louis L. – 1928 – “Attitudes Can Be Meassured”, in American Journal of Sociology, Vol. 33, p. 529-554,
12. Torgerson, Warren S. – 1965 – Theory and Methods of Scaling – New York, John Wiley & Sons.