

Selltiz, Wrightsman e Cook

Métodos de Pesquisa nas Relações Sociais

Volume 2
Medidas na Pesquisa Social

Organizadora da
4ª edição norte-americana:
Louise H. Kidder

Coordenadores da
2ª edição brasileira:
José Roberto Malufe
Bernardete A. Gatti

**Dados de Catalogação na Publicação (CIP) Internacional
(Câmara Brasileira do Livro, SP, Brasil)**

| | | |
|-------------------------|---|------------|
| M552 v.1-3 2ª ed. | Métodos de pesquisa nas relações sociais / Selltiz, Wrightsman e Cook ; organizadora da 4ª edição norte-americana Louise H. Kidder ; (tradutores Maria Martha Hubner d'Oliveira, Miriam Marinotti del Rey); — 2ª ed. brasileira / coordenadores José Roberto Malufe, Bernardete A. Gatti. — São Paulo : EPU, 1987. Edição revista por Louise H. Kidder, com 6 novos capítulos de sua autoria. Bibliografia. Conteúdo: v. 1. Delineamentos de pesquisa — v. 2. Medidas na pesquisa social — v. 3. Análise de resultados. 1. Ciências sociais — Pesquisa I. Selltiz, Claire. II. Wrightsman, Lawrence Samuel. III. Cook, Stuart Wellford, 1913 — IV. Kidder, Louise H. 87-0126 | CDD-300.72 |
|-------------------------|---|------------|

Índices para catálogo sistemático:

1. Ciências sociais : Pesquisa 300.72
2. Pesquisa social : Ciências sociais 300.72
3. Pesquisa social : Planejamento : Ciências sociais 300.72
4. Planejamento : Pesquisa social : Ciências sociais 300.72

E.P.U.



EDITORA PEDAGÓGICA
E UNIVERSITÁRIA LTDA.

e itens e respostas padronizadas são obtidos. Também não se verificou forte tendência para concordar ou discordar sistematicamente, ou em colocar itens em uma só classificação. Sua maior desvantagem é que se faz necessária a presença de um entrevistador para instruir o respondente durante a aplicação. Os cartões exigem também respondentes com boa capacidade de leitura. Finalmente, conjuntos de itens maiores que 60 ou 90 não são recomendados por causa de fadiga e do enfado.

Resumo

Fica patente que pesquisadores interessados em auto-relatos individuais têm inúmeras maneiras de produzi-los. Ao tomar decisões quanto a usar um questionário ou entrevista, de forma estruturada ou não, e quanto a complementar o material verbal com recursos visuais, é preciso considerar as vantagens e desvantagem de cada método à luz dos objetivos de seu estudo. Embora os métodos tenham sido discutidos independentemente, combinações entre eles são certamente apropriadas em determinadas situações. A entrevista estruturada padronizada foi enfatizada neste capítulo por sua utilidade quando um grande número de variáveis precisam ser estudadas num grande grupo de pessoas.

Há algumas questões metodológicas fundamentais que devem ser observadas.

Primeiramente, há necessidade de *amostras representativas* se generalizações serão feitas para outros grupos, além do entrevistado. Os *índices de repostas* devem ser altos, pois, caso contrário, a representatividade será destruída, uma vez que as pes-

soas que não participam são provavelmente diferentes daquelas que participam.

Em segundo lugar, há uma necessidade de *indicadores múltiplos*, ou a elaboração de mais de uma questão para as variáveis-chaves do estudo. Nenhum item é fidedigno ou multifacetado o suficiente para ser o único a ser incluído, especialmente se está se lidando com questões não factuais.

Está implícito que pesquisadores que fazem levantamentos, como quaisquer outros pesquisadores, deveriam testar a fidedignidade de suas medidas-chaves, refazendo as mesmas questões através da comparação de resultados de diferentes entrevistadores ou por outros meios. A *validade* (exceto de conteúdo ou medidas não estatísticas da mesma) é um problema mais difícil porque as respostas ou critérios "verdadeiros" não são freqüentemente disponíveis para fins de checagem. Tal critério deveria ser buscado e utilizado, quando possível.

Relacionado às questões anteriores está a necessidade de uma atenção mais cuidadosa na arte e ciência da construção de *questionário e entrevista*. Há técnicas e experiências a serem utilizadas, algumas diretrizes gerais a serem seguidas, pré-testes exaustivos a serem feitos e, normalmente, várias revisões a serem conduzidas antes do roteiro de entrevista estar concluído. O pesquisador deve se conscientizar que uma vez concluídas as questões do levantamento, as variáveis dependentes não são mais abstrações ou constructos representados por frases sumárias do tipo "atitudes em relação à democracia". Mais do que isto, as variáveis dependentes são precisamente aquelas questões que foram formuladas ao medir as abstrações — e nada mais.

9. Construção de escalas

Na mensuração é freqüentemente necessário fazer distinções de *grau* e não de *qualidade*. Podemos, por exemplo, desejar saber se o sr. Verde sente-se mais inclinado do que o sr. Azul a concordar com que as crianças devam freqüentar escolas distantes das suas casas, para favorecer a integração racial. Ou podemos desejar saber se o estado de Mississippi gasta mais, *per capita*, em educação superior pública do que o estado de Ohio, levando-se em conta o poder econômico relativo dos dois estados.

Fazer tais distinções de grau pode ser considerado como uma função da análise, e não da coleta de dados; isto é, um procedimento realizado *depois* que os dados são coletados. No entanto, o desejo de fazer tais distinções influi na *forma* de coletar os dados. Pelo menos significa que as questões elaboradas devam ser tais que dêem informações em que possam se basear os julgamentos de grau. Freqüentemente, as distinções de grau são introduzidas nos próprios instrumentos de mensuração.

Em cada um dos exemplos precedentes estamos comparando características de duas parcelas de dados ao longo de uma escala comum. Estas escalas funcionam co-

mo esquemas de mensuração — sejam de atitudes, de indicadores de gastos e poder econômico do Estado, ou de outras características — que possibilitam fazer comparações entre casos, quer as observações ocorram ao mesmo tempo que ocorram em momentos diferentes.

Escalas versus dimensões

É necessário distinguir claramente escalas e dimensões, uma vez que ambos os termos são utilizados com freqüência considerável na pesquisa das ciências sociais. Ao falar de *escalas*, estamos nos referindo às técnicas empregadas na combinação de uma ou mais mensurações com o objetivo de estabelecer um único escore para cada indivíduo. A *dimensão* dos dados é a inferência, a partir dos dados das escalas, de que uma característica subjacente existe, e que descreve os casos a nível conceitual. Podemos definir conceitualmente, por exemplo, liberalismo — conservadorismo; poderíamos, então, elaborar alguns itens de questionários que acreditássemos iriam medir liberalismo — conservadorismo. Se nossos itens formarem uma escala satisfatória (de acordo com critérios que serão

discutidos adiante neste capítulo), poderíamos inferir a presença de uma dimensão teórica mais ampla, liberalismo-conservadorismo, para aquela particular amostra de respostas. As pesquisas mais antigas das ciências sociais procuravam elaborar escalas que preenchiem dimensões únicas; mais recentemente, tem-se percebido a natureza multidimensional da maioria das escalas.

Tradicionalmente, indivíduos têm sido colocados em escalas por várias razões. Em alguns casos o objetivo era o de reduzir a complexidade dos dados, de calcular um escore único que representasse muitas variáveis de tal forma que a análise pudesse ser simplificada. Em outros casos, pesquisadores procuravam testar a hipótese de que muitas variáveis mediam, de fato, uma única; subjacente a um conceito, de tal forma que se poderia dizer que elas variavam na mesma "dimensão" de um conceito. O objetivo poderia ser, finalmente, o de reduzir o erro nas mensurações e desse modo aumentar a fidedignidade das medidas finais que eram utilizadas na análise. Técnicas para a colocação de indivíduos em escalas são utilizadas em muitas ocasiões no sentido de atingir todos esses três objetivos. Deixe-nos apresentar é pretendido em cada uma dessas tarefas.

Suponha que há muitas variáveis, todas com o objetivo de relatar a percepção individual de força e fraqueza no que se refere às relações com o governo e com a definição de programas públicos. Chamemos estes itens de medida de "percepção de força política" (Campbell *et al.*, 1954; Campbell *et al.*, 1960). Seria um tanto enfadonho testar hipóteses sobre força política relacionando cada item do questionário individualmente, a outras variáveis que são supostamente relacionadas com força política em termos teóricos. A escala permite-nos simplificar a análise, combinando muitas variáveis em um único valor.

Os procedimentos — discutidos em maior detalhe adiante neste capítulo — pa-

ra converter os valores de cada variável em um único valor para uma escala envolvem usar um conjunto de regras para "recodificar" cada caso na análise. Neste sentido, uma "nova" variável é criada, utilizando-se regras explícitas para transformar os valores iniciais em valores de escala.

O quanto variáveis individuais podem ser colocadas em uma única escala, sem que haja uma perda de grandes quantidades de informação, depende do critério de *dimensionalidade*. Nem todas as escalas que discutimos fazem uso explícito da noção de dimensionabilidade, mas este é, sem dúvida, um conceito importante.

Se todas as variáveis medem as mesmas características gerais de uma atitude, ou de algum outro atributo de um indivíduo, então poderíamos poder demonstrar que todas as variáveis são altamente inter-relacionadas. Supomos que as variáveis medem diferentes quantidades de uma mesma coisa e consideramos que elas são *unidimensionais*. Se, contudo, estas relações forem baixas, precisamos pesquisar a possibilidade de existência de mais de uma dimensão. Assim, os procedimentos de colocar indivíduos em escalas são também testes da hipótese que as variáveis individuais que foram combinadas para formar um escore de uma única escala podem ser também agrupadas ao longo de uma única *dimensão*.

É importante enfatizar que a existência ou não existência de uma única dimensão reflete uma hipótese de pesquisa que é como qualquer outra hipótese. Se ela é ou não verdadeira é uma questão de teste e não de especulação. Além disso, o que é verdadeiro numa época, para uma amostra, pode não ser verdadeiro numa época posterior. Podemos novamente considerar, como um exemplo, o caso da eficácia política.

Diversas pesquisas realizadas durante os anos 50 e o início dos anos 60, com populações diferentes e utilizando quatro itens

que pretendiam medir a eficácia política, isto é, a percepção de força política, concluíram que os quatro itens utilizados formavam uma única dimensão (Campbell *et al.*, 1954; Campbell *et al.*, 1960). Assim, supôs-se que a eficácia proporcionava uma atitude unidimensional, altamente geral, e bastante imutável, a qual tinha se mostrado útil para explicar diferentes aspectos da participação política.

Posteriormente, uma análise de itens mais completa, realizada com os itens de eficácia que tinham sido utilizados, e baseada em dados coletados durante os anos 60, revelou, contudo, que durante essa turbulenta década de quatro itens tinham se separado, de maneira que eles representavam agora duas dimensões diferentes em vez de uma única dimensão. Os itens que pertenciam a uma das duas dimensões pareciam estar medindo atitudes sobre o potencial de alguém para afetar a vida política, ao passo que os itens pertencentes a outra dimensão pareciam estar medindo atitudes sobre a capacidade que tinham os políticos para provocar mudanças nos programas públicos (Balch, 1974).

Em resumo, foi demonstrado empiricamente que o conjunto de itens não mais media o que eles haviam medido, porque a maneira pelas qual estavam inter-relacionados havia mudado. E a dimensionalidade da escala original havia sofrido uma mudança concomitante.

Muitas técnicas de construção de escalas aumentam a fidedignidade das medidas reduzindo a magnitude dos erros casuais que estão associados com respostas a itens de questionários e de entrevistas pessoais. O erro pode também ser reduzido quando a escala é montada a partir de agregados de tais tipos de respostas, na análise de dados agrupados, tais como materiais de censo, relatos estatísticos de vários tipos e outras medidas (uma vez que a maneira pela qual a construção de escalas reduz os erros casuais de mensuração é muito complexa para se detalhar neste livro, o leitor de-

ve consultar outras fontes sobre mensuração de atitudes, tais como Edwards, 1957; Scott, 1968; Torgerson, 1958; e Upshaw, 1968).

Podemos também distinguir as escalas como base no fato de utilizarem uma só variável, ou diversas variáveis, para se calcular um escore.

Em algumas escalas muito simples, por exemplo, os indivíduos podem classificar a si mesmos de acordo com alguma característica, ou juízes (na maioria das vezes codificadores especialmente treinados) podem avaliar indivíduos, grupos, ou instituições de acordo com características que cada um possui, colocando-os em uma escala. Para a questão "Você se considera muito liberal, liberal, tão liberal quanto conservador, conservador ou muito conservador?", indivíduos classificam-se a si próprios pelas respostas que dão, com base em suas autodesignações de liberalismo — conservadorismo. Esta resposta permite, então, que cada pessoa seja comparada, legitimamente, com outra que também tenha respondido à mesma questão.

Ou considere a tarefa dos codificadores na tentativa de avaliar uma reportagem jornalística sobre a participação do governo norte-americano na resolução das hostilidades árabes israelenses, como sendo, "favorável, neutra ou desfavorável" (Holsti, 1969, p. 107). Os codificadores tentaram avaliar os conteúdos e formas de apresentação da reportagem com base num conjunto de regras que eles haviam sido treinados a usar. A unidade de análise pode ser diferente (uma reportagem, em vez de um indivíduo), e a variável pode ser diferente (uma caracterização aproximada de uma matéria de noticiário em vez de uma atitude de uma pessoa), mas ambos — a matéria de noticiário e o indivíduo — são colocados em escalas, de tal forma que se podem fazer comparações entre casos.

Um segundo tipo muito geral de escala é construído combinando-se os escores em um determinado número de itens para for-

mar um único escore em uma escala, como foi mencionado anteriormente. Indivíduos respondem a uma série de itens num questionário, ou se coletam medidas de uma série de características sócio-econômicas. Em cada caso, assume-se que cada variável mede uma dimensão mais geral, comum, de tal forma que escores de uma escala única representem uma mesma dimensão quando as variáveis são combinadas, mediante as técnicas de construção da escala, num escore único.

Por exemplo, no caso de atitudes sobre integração racial, um homem branco pode ser caracterizado de acordo com sua concordância ou discordância com cada uma das seguintes afirmações: "Preferiria ter o mesmo número de pretos e brancos em minhas aulas. Acredito que a qualidade de minha formação educacional seria melhor se negros fossem admitidos em minha escola; valores de propriedades não caem quando pessoas negras mudam-se para a vizinhança", e assim por diante. Um escore que representasse uma atitude mais geral sobre o sentimento integracionista poderia ser, então, construído, somando-se o número total de respostas de *concordância* e subtraindo-se, então, o número total de respostas de *discordância*. Presumir-se-ia que o escore único representaria uma atitude mais geral sobre integração racial, mesmo que cada item individual refira-se a um aspecto diferente do problema.

Escalas de avaliação

Tabela 9.1. Respeito pelos negros no conjunto residencial (coloque um sinal na posição adequada da linha, ou então assinale X ou Y).

| Tem elevada opinião dos negros do conjunto, sem restrições | De modo geral respeita os negros que moram no conjunto | É ambivalente: em parte respeita, em parte pensa que são inferiores | De modo geral, pensa que são inferiores | Tem opinião formada sobre sua inferioridade | | | | |
|--|--|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

X: É indiferente aos negros como um grupo; não pensa neles.

Y: Não pensa nos negros como grupo; considera-os como indivíduos.

Têm sido empregados vários tipos de escalas de avaliação, mas uma característica é comum a todos os tipos: o avaliador coloca a pessoa ou o objeto medido em determinado ponto de um contínuo, ou em uma categoria, numa série ordenada de categorias; dá-se um valor numérico ao ponto ou à categoria. As escalas diferem quanto à sutileza de distinções que permitem e quanto aos procedimentos exigidos para a distribuição de pessoas ou objetos pelas posições. Tais diferenças ficarão evidentes em nossa discussão de alguns dos tipos mais comuns de escalas de avaliação.

Escalas gráficas

A escala gráfica de avaliação é talvez a mais amplamente usada. Nesse tipo, o juiz (o sujeito, o entrevistador, o observador, o codificador etc.) indica sua avaliação através da colocação de um sinal no ponto adequado, numa linha que vai de um extremo a outro do atributo em questão. Os pontos de escala, com descrições breves, podem ser indicados ao longo da linha; sua função não é apresentar categorias distintas, mas tão-somente servir de guia para o juiz, ao localizar sua avaliação.

A Tabela 9.1. é uma escala assinalada por entrevistadores de pessoas que viviam em conjuntos residenciais inter-raciais (Deutsch e Collins, 1951).

Neste exemplo, os itens X e Y foram introduzidos para evitar que os entrevistado-

res fossem obrigados a fazer uma avaliação que eles considerassem inadequada, e para permitir que explicassem por que não poderia ser feita determinada avaliação.

Uma das principais vantagens das escalas gráficas de avaliação é sua relativa facilidade de emprego.

No entanto, para o emprego eficiente de escalas gráficas de avaliação, a experiência mostrou que é preciso tomar algumas precauções no seu planejamento e no seu emprego, por exemplo, devem ser evitadas afirmações finais tão extremas que provavelmente não serão usadas; as afirmações descritivas devem ser colocadas de forma a corresponder tão diretamente quanto possível aos pontos numéricos na escala. Para uma discussão mais minuciosa de práticas que devem ser seguidas na construção e no emprego de escalas gráficas de avaliação, o leitor deve consultar Guilford (1954).

Escalas com itens

As escalas de avaliação com itens têm sido denominadas, também, "escalas de categoria específica" (Krech e Grutchfield, 1948) e "escalas numéricas" (Guiford, 1954). Nesse tipo de escala, o avaliador escolhe uma categoria, em certo número de categorias ordenadas de acordo com sua posição na escala. O número de posições ou categorias na escala tem variado, de acordo com o problema de pesquisa e os tipos exigidos de julgamentos. As escalas com cinco ou sete categorias, são as mais frequentemente empregadas, mas muitos pesquisadores têm usado até nove ou onze pontos.

As escalas de avaliação com itens, tal como as escalas gráficas, podem ser usadas com dados obtidos através de diferentes métodos. Sherif, Sherif e Nebergall (1965), por exemplo, realizaram um estudo sobre atitudes pró-republicanas e pró-democratas, durante a campanha presidencial de 1960 nos Estados Unidos. Dentre

outras coisas, solicitava-se aos sujeitos que respondessem a nove afirmações sobre como a eleição de candidatos republicanos e democratas serviria aos maiores interesses do país. As afirmações variavam de declarações inequivocamente favoráveis aos candidatos republicanos:

As eleições de candidatos republicanos para a presidência e vice-presidência em novembro é absolutamente essencial para os interesses do país em todos os sentidos.

a uma posição neutra com relação aos candidatos:

Do ponto de vista dos interesses do país é difícil decidir se é preferível votar em candidatos do Partido Republicano ou Democrata para presidência e vice-presidência, nas eleições de novembro.

e a declarações inequivocamente favoráveis dos candidatos democratas:

A eleição de candidatos democratas à presidência e vice-presidência em novembro é absolutamente essencial para os interesses do país, em todos os sentidos.

Em uma parte do estudo, os sujeitos eram, então, instruídos a selecionar a afirmação que era mais aceitável por eles e aquela menos aceitável. (Foram dadas várias outras tarefas para os sujeitos nesse estudo. Por exemplo, sujeitos foram instruídos a assinalar todas as outras afirmações que eram aceitáveis para eles e todas aquelas não aceitáveis. Construiu-se, a partir desses dados, uma variedade de escores em diferentes escalas.) Contudo, mais importante para os propósitos do momento é a classificação simples das afirmações que mais se aproximam da posição dos sujeitos na eleição. Esta afirmação específica onde o sujeito se localiza, em uma escala referente às atitudes frente aos partidos Democrata e Republicano. Pessoas que se localizam no extremo pró-republicano do *continuum* são consideradas favoráveis aos candidatos republicanos, enquanto as pessoas que se localizam no extremo pró-democrata do *continuum* são consideradas favoráveis aos Democratas. Da mesma forma, pessoas que consideram as afirmações

pró-democratas “as mais questionáveis” são consideradas desfavoráveis aos candidatos democratas, enquanto pessoas que consideram as afirmações pró-republicanas as “mais questionáveis” são consideradas desfavoráveis aos Republicanos.

Os resultados desse estudo indicam, ainda, que a aceitação e rejeição são, normalmente, imagens em espelho uma da outra. Poucas pessoas gostam ou desgostam de ambos os partidos igualmente.

As descrições verbais usadas para identificar as posições na escala podem variar, desde afirmações simples que indicam apenas graus do atributo, até descrições complexas que incluem exemplos do comportamento adequado para a categoria. Os dois exemplos acima apresentados contêm definições e ilustrações relativamente minuciosas. No outro extremo, Proshansky (1943), na análise de descrições de figuras num teste projetivo de atitude com relação a operários, instruiu seus juizes para que avaliassem as descrições através de uma escala de cinco pontos, na qual as duas categorias extremas eram definidas apenas como muito favorável”, e “muito desfavorável”, enquanto a categoria do meio indicava que era impossível classificar a pessoa como favorável ou desfavorável: a segunda e a quarta posições não tinham definições específicas, a não ser dizer que deviam ser usadas para pessoas que estivessem entre a posição extrema e a posição intermediária.

De modo geral, quanto mais claramente definidas as categorias, mais precisas tendem a ser as avaliações. A quantidade de definição necessária depende da natureza do material do conhecimento que os codificadores têm dos conceitos incluídos no estudo, e da sutileza das distinções necessárias para o estudo.

Os problemas existentes na construção e aplicação de escalas de avaliação com itens são, em sua maioria, semelhantes aos que caracterizam outros tipos de escalas de

avaliação. Tais problemas são tratados mais adiante nesta seção.

Escalas comparativas

Ao usar escalas de avaliação, gráficas ou com itens, o avaliador faz seu julgamento do indivíduo, sem referência direta às posições de outros indivíduos ou grupos dos quais poderia ser comparado. De outro lado, as escalas comparativas de avaliação — como se vê por seu nome — supõem claramente esses julgamentos relativos. As posições na escala de avaliação são expressamente definidas através de determinada população ou determinado grupo, ou ainda através de pessoas de características conhecidas. Por exemplo, um questionário usado para a seleção de candidatos à admissão em um curso de pós-graduação pode solicitar, do avaliador, uma estimativa da capacidade de determinado candidato para realizar trabalho de pós-graduação, “em comparação com o grupo total de estudantes de pós-graduação que você conhece” — será que este candidato é mais capaz que 10 por cento deles? 20 por cento? 30 por cento? O avaliador pode ser solicitado a indicar, por exemplo, se a habilidade de liderança de um indivíduo se assemelha mais à da pessoa A, da pessoa B ou da pessoa C (todos conhecidos pelo avaliador e todos avaliados através de sua habilidade de liderança). No primeiro exemplo dado, a fim de fazer uma avaliação válida, o juiz precisa ter uma clara concepção da amplitude e distribuição das capacidades do grupo total de estudantes pós-graduados. Frequentemente é difícil construir as escalas do segundo tipo, pois pode não haver suficiente variação no comportamento de lideranças (ou qualquer outra característica que esteja sendo avaliada) entre as pessoas conhecidas pelos juizes e que possam servir como exemplos para os vários pontos na escala.

Outro procedimento comparativo ou relativo de avaliação é a escala de ordem de

postos. Aqui, o juiz deve ordenar os indivíduos, especificamente, através da relação entre eles; indicar qual a pessoa colocada no posto mais elevado, em função da característica que está sendo medida, qual a colocada em segundo lugar, e assim por diante, até a que está colocada na posição mais baixa. A ordenação desse modo é usada apenas quando o pesquisador se interessa por um grupo limitado de indivíduos. A avaliação que um indivíduo recebe indica apenas sua posição ou seu posto relativo no grupo que está sendo estudado; não é necessário que tenha qualquer utilidade fora do grupo específico, cujos membros estão sendo comparados.

Auto-avaliações e avaliações feitas por outros

Todos esses tipos de escala podem ser usados para se obter a avaliação do indivíduo por si mesmo, ou a avaliação dele por outra pessoa. Parece razoável supor que o indivíduo está, freqüentemente, em melhor posição para observar e descrever seus sentimentos, suas crenças, e seus medos que qualquer outra pessoa. No entanto, essa suposição só é válida se o indivíduo está consciente de sua crença e seus sentimentos, e se está disposto a revelá-los aos outros. Se o indivíduo não está consciente, por exemplo, de que tem sentimentos hostis com relação a determinado grupo minoritário, ou se tem consciência de tais sentimentos mas teme as conseqüências de sua revelação então o procedimento de auto-avaliação tem pouco valor. Outra dificuldade decorre do fato de que, mesmo que o indivíduo seja capaz de descrever objetivamente seus sentimentos e crenças, sua concepção quanto ao que constitui uma posição extemada ou moderada pode ser muito diversa da de outros que fazem auto-avaliações comparáveis.

Apesar das dificuldades aí existentes, verificou-se que as auto-avaliações são úteis na mensuração de atitudes sociais. No ca-

so de sentimentos e crenças de que, segundo se pode esperar, o indivíduo está consciente e disposto a relatar, por exemplo, atitudes com relação a determinados programas de televisão —, as auto-avaliações constituem uma útil fonte de informação. Com relação a certas características das atitudes — por exemplo, intensidade e importância — verificou-se até agora que as auto-avaliações constituem a única fonte satisfatória de informação. A atenção dada pelo pesquisador a algumas questões — por exemplo, clara especificação da dimensão a ser avaliada e definição do quadro de referência ou dos padrões dentro dos quais devem ser feitas as avaliações — podem contribuir para reduzir as possibilidades de deformação nas auto-avaliações e para aumentar sua utilidade.

Construção e uso de escalas: algumas precauções

Como no uso de escalas de avaliação existe uma grande parte de julgamento, é muito possível que o viés pessoal do avaliador ou dos avaliadores introduza erros sistemáticos. Um erro sistemático relativamente comum é o *efeito de halo*. Se há necessidade de julgar mais de uma característica de uma pessoa, freqüentemente os avaliadores transferem, de uma avaliação para outra, uma impressão generalizada da pessoa, ou tentam tornar coerentes as suas avaliações.*

Assim, se um avaliador considera que uma pessoa é acanhada, e se acredita que as pessoas acanhadas são mal-ajustadas, tende a avaliar a pessoa como mal-ajustada, além de acanhada. É evidente que o efeito de halo reduz a validade das avalia-

* A tendência para tornar coerentes as avaliações não é diversa do que Newcomb (1931) descreveu como um “erro lógico”, isto é, os juizes freqüentemente apresentam avaliações semelhantes dos traços que lhes parecem logicamente relacionados. Ver também a discussão do efeito de halo e de outras tendências em Phillips (1971, pp. 12-49).

ções de alguns traços e introduz um grau espúrio de correlação positiva entre os traços avaliados.

Outro tipo freqüente de erro sistemático é o erro da *generosidade*. Aqui, a tendência do avaliador é superestimar as qualidades desejáveis das pessoas de quem gosta mais. Foram identificados, também, outros erros freqüentes. Por exemplo, os avaliadores tendem a evitar julgamentos extremos e a distribuir os indivíduos pelas categorias mais moderadas. Murray *et al.* (1938) identificaram o erro de *contraste*, no qual existe uma tendência do avaliador para ver, em determinado traço, os outros como opostos a ele. Verificaram, por exemplo, que os avaliadores muito organizados avaliavam os outros como relativamente desorganizados, enquanto os avaliadores menos organizados tendiam a ver os outros como mais organizados.

Evidentemente, uma forma de reduzir os erros sistemáticos como os acima descritos é treinar cuidadosamente os avaliadores e, sobretudo, torná-los conscientes da possibilidade de tais vieses. É possível dar passos específicos para reduzir a probabilidade de tipos específicos de erro. Por exemplo, a tendência para evitar o emprego de posições extremas pode ser contrabalançada através da apresentação de títulos menos extremados para tais posições. As pessoas têm maior tendência para assinalar. "Estou satisfeito com meu trabalho", que: "Estou inteiramente satisfeito com meu trabalho"; no outro extremo, têm maior tendência para assinalar: "Existem muitas coisas de que não gosto em meu trabalho", que: "Não gosto de coisa alguma em meu trabalho". O "erro de generosidade" pode ser reduzido através do emprego de termos descritivos relativamente neutros, em vez de termos valorativos; por exemplo: "Não aceita facilmente maneiras novas de fazer as coisas ou novas opiniões", em vez de "rígido". É possível reduzir ou eliminar inteiramente os efeitos de halo, através de várias avaliações independentes de

determinada pessoa — seja por diferentes avaliadores, seja pelo mesmo avaliador, em momentos diferentes, sem ele saber que está avaliando a mesma pessoa. Evidentemente, esta última condição só pode ser satisfeita quando as avaliações são feitas através de material registrado — por exemplo, respostas e perguntas de entrevista, descrições de comportamento ou outras fontes das quais a informação identificadora pode ser suprimida.

Evidentemente, os erros sistemáticos reduzem a validade das avaliações. Existem também erros casuais que reduzem sua precisão. Uma fonte freqüente de imprecisão, entre diferentes avaliadores, é o fato de, em qualquer avaliação, estar implícito um quadro de referência; diferentes avaliadores podem usar diferentes quadros de referência ao descrever os indivíduos através de determinada característica. Por exemplo, a avaliação de uma pessoa como "conservadora" ou "radical" tem sentido a partir dos grupos de referência do avaliador — as normas de grupo em que pensa ao fazer sua avaliação. A falta de correspondência entre avaliações de diferentes observadores se deve, freqüentemente, ao fato de fazerem avaliações a partir de diferentes grupos de referência.

É possível aumentar a *precisão* ou *fidedignidade*, não apenas por *treinamento* cuidadoso dos avaliadores, mas também pela atenção à construção da escala de avaliação. *Definições claras* da característica medida e das várias posições na escala, bem como a *clara especificação* do grupo de referência, ajudam a reduzir a imprecisão. Sempre que possível, as definições dos pontos da escala devem *incluir* ilustrações *concretas* de respostas a perguntas, de tipos de comportamento, de conteúdo de comunicação. Deve-se considerar cuidadosamente a distinção entre posições adjacentes na escala; por exemplo, a diferença entre "favorável" e "muito favorável".

Ao construir uma escala de avaliação, é preciso decidir quantas posições ou cate-

gorias devem ser usadas, a não ser que se empregue uma escala gráfica na qual o avaliador possa analisar qualquer ponto numa linha contínua. Não existe regra simples para a determinação do número ótimo de posições. Evidentemente, uma consideração básica é o grau de diferenciação que se deseja na medida. No entanto, qualquer que seja a exigência do problema de pesquisa, outros fatores devem ser considerados: 1) a capacidade de discriminação dos juízes ou avaliadores, considerando-se, também, até que ponto têm treinamento e experiência; 2) o tipo de características que devem ser julgadas; por exemplo, se são atributos "internos" e complexos, ou se são atributos externos, mais manifestos; 3) as condições em que devem ser feitas as avaliações; por exemplo, se são baseadas em grandes quantidades de dados (longos períodos de observação da pessoa, ou muito conteúdo de comunicação) ou em poucos dados (observação rápida ou conteúdo limitado de comunicação).

Tais fatores interagem em seu efeito no grau de refinamento possível na escala de avaliação. Para avaliar comportamentos relativamente manifestos, é possível usar uma escala mais diferenciada, mesmo com juízes relativamente pouco sofisticados, desde que adequadamente treinados. Do outro lado, se é preciso julgar atributos complexos e não existem juízes experimentados, pode ser necessário empregar uma escala menos diferenciada a fim de conseguir avaliações fidedignas. Em qualquer caso, todos esses fatores devem ser considerados ao determinar o número de distinções que devem ser incluídas na escala de avaliação. Muito freqüentemente, há necessidade de teste preliminar antes de uma decisão final.

Geralmente, a fidedignidade das avaliações aumenta consideravelmente quando se faz com que vários avaliadores trabalhem em equipe — façam julgamentos independentes, comparem suas avaliações e discutam suas discrepâncias, façam outros julgamen-

tos independentes para reunir aos primeiros, ou dos quais se tira uma média, a fim de chegar a um escore final (ver Murray *et al.*, 1938). Muitas pesquisas demonstraram a superioridade da média, ou do consenso, dos julgamentos de várias pessoas, com relação ao julgamento de um só indivíduo (vide Murphy, Murphy e Newcomb, 1937) porque erros usuais tendem a ser eliminados quando se combinam várias avaliações independentes.

Seria relativamente inútil tentar discutir minuciosamente o que constitui um bom avaliador, pois existem poucas pesquisas definitivas a respeito. No entanto, algumas conclusões gerais parecem justificadas. Em primeiro lugar, não apenas os indivíduos diferem em sua capacidade como avaliadores, mas o mesmo indivíduo difere em sua capacidade para avaliar diferentes características. Um bom avaliador de determinada característica pode ser mais avaliador para outra. Em outras palavras, a capacidade para julgar não é geral; é uma função muito especializada no indivíduo. Em segundo lugar — ao contrário do que diria a expectativa geral —, a familiaridade com a pessoa a ser avaliada não provoca aumento da *validade* das avaliações. Guilford (1954) indicou que o conhecimento prolongado freqüentemente provoca grandes erros de generosidade. Em terceiro lugar, a confiança no próprio julgamento não é, necessariamente, indicação de sua *validade*. Pelo menos uma pesquisa (Kelly e Fiske, 1950) mostrou relação inversa entre confiança e validade das predições. No entanto, Guilford (1954) citou um estudo anterior, realizado por Cady, que indicava que os julgamentos a respeito dos quais o avaliador está seguro são muito mais fidedignos que aqueles a respeito dos quais está menos seguro. Em quarto lugar, quase todos os estudos de validade de avaliação parecem indicar que as pessoas possuem em alto grau alguma característica considerada de modo geral não-desejável são maus juízes dessa característica, tanto em si mesmo como nos outros.

Questionários que formam escalas

Como já foi indicado, existe uma grande parte de julgamento quando um avaliador coloca um indivíduo numa escala de avaliação, a partir de seu comportamento observado, suas respostas a perguntas abertas, ou outras respostas. Para criar procedimentos que permitam a colocação de um indivíduo numa escala, com menos probabilidade de erro, têm sido construídos questionários padronizados. Nesta abordagem, o indivíduo não se descreve diretamente através de sua posição em determinada dimensão. Em vez disso, exprime seu acordo ou desacordo com algumas afirmações relevantes para tal dimensão: a partir dessas respostas, recebe um escore. No processo de padronização do questionário, o pesquisador estabelece uma base para interpretar esses escores como indicadores de posições na dimensão. Como essa técnica tem sido usada mais freqüentemente na mensuração de atitudes, nossa discussão focalizará as escalas de atitude.

As escalas de atitude diferem entre si quanto ao método de construção, método de resposta e base para interpretação dos escores. Nesta seção serão discutidos diferentes tipos de escalas de atitude.

Geralmente as perguntas ou itens separados de uma escala de atitude não têm interesse por si mesmos: ao contrário, o interesse está no resultado total ou nos sub-resultados que cada indivíduo obtém com a combinação de suas respostas a vários itens. A suposição aceita é a de que escore total irá indicar a posição de cada indivíduo em uma atitude mais geral. Com efeito, qualquer conjunto de itens será tão bom quanto qualquer outro, desde que dê os mesmos resultados finais na atitude específica que está sendo medida.

Ao selecionar os itens para uma escala, geralmente se usam dois critérios. Em primeiro lugar, os itens devem provocar respostas psicologicamente relacionadas com a atitude que está sendo medida. Por exem-

plo, numa escala que mede anti-semitismo, o item seguinte tem relação *manifesta* com a atitude que está sendo medida: "Todos que empregam muitas pessoas, devem ter cuidado para não empregar uma grande porcentagem de judeus" (Adorno *et al.*, 1950). No entanto, a relação não precisa, necessariamente, ser evidente. Na realidade, existe grande vantagem em usar itens que, superficialmente, não têm relação com a atitude que se mede. Isso pode evitar que a pessoa esconda ou deforme sua atitude. Assim, em seu estudo da ideologia antidemocrática, Adorno e seus colaboradores usaram muitos itens que não têm relação aparente com essa atitude — por exemplo: "Quando uma pessoa tem um problema ou uma preocupação, é melhor que não pense nisso e se ocupe com coisas mais alegres". Esse item é um dentre vários que indicam a preparação ou falta da preparação da pessoa para adotar uma visão psicologicamente compreensiva de outras pessoas e de si mesma. Segundo a teoria, as pessoas que não têm compreensão e autoconhecimento psicológico têm uma estrutura de personalidade (por exemplo, maior hostilidade reprimida ego mais fraco etc.) que as predispõe a uma ideologia antidemocrática*.

O segundo critério exige que a escala diferencie entre pessoas que estão em diferentes pontos da dimensão que está sendo medida. Para discriminar não apenas en-

* Evidentemente, tais itens indiretos não podem ser usados como medidas da atitude que está sendo estudada, apenas a partir de suposições teóricas sobre sua relação com a atitude. Antes de serem aceitos como medidas adequadas, é preciso demonstrar sua relação com a atitude. (Essa afirmação é também correta para itens que parecem ter uma relação *manifesta* com a atitude que está sendo estudada.) Assim, no estudo de ideologia antidemocrática, a hipótese de que alguns itens, como o citado, estavam ligados à ideologia antidemocrática foi testada — e revelada — por análise da diferença entre respostas a tais itens de pessoas conhecidas, por outras razões, por sua ideologia democrática, e as respostas de pessoas que, segundo se sabia, tinham uma ideologia antidemocrática.

tre extremos opostos quanto à atitude, mas também entre indivíduos que têm pequenas diferenças, geralmente se incluem itens que discriminem em diferentes pontos da escala. Assim, um teste de opiniões sobre práticas de educar crianças, na dimensão "permissivo-restritivo, deve conter não apenas itens que representem uma posição extremamente restritiva e outros que representem uma posição extremamente permissiva, mas também itens intermediários, que representem restritividade moderada, permissividade moderada, e algum ponto intermediário. No entanto, alguns tipos de escala permitem a identificação de posições moderadas, através da expressão de vários graus de acordo ou desacordo com itens extremos, e não pela inclusão de itens intermediários.

A maneira pela qual uma escala discrimina entre indivíduos depende da construção da escala e do método de atribuição dos escores. Em algumas escalas, os itens formam uma gradação de tal natureza que o indivíduo irá provavelmente concordar com apenas um ou dois, os quais correspondem à sua posição na dimensão que está sendo medida, e irá provavelmente discordar das afirmações que estiveram acima ou abaixo dos itens que escolher. Tais escalas, nas quais a resposta de uma pessoa localiza sua posição, são às vezes denominadas *escalas diferenciais*. Em outras escalas, o indivíduo indica seu acordo ou desacordo com cada item, e seu resultado total é calculado pela soma de resultados parciais atribuídos a suas respostas em todos os itens separados; tais escalas são às vezes denominadas *escalas somatórias*. Outras são estabelecidas de forma que os itens formem uma *série cumulativa*: teoricamente, um indivíduo cuja atitude se coloca em certo ponto da dimensão que está sendo medida responderá favoravelmente a todos os itens de um lado desse ponto, e desfavoravelmente a todos os do outro lado. Nos parágrafos seguintes serão discutidos todos esses tipos de escala. (Ver também

as discussões minuciosas sobre escalas de atitude em Dawes, Singer e Lemons, 1972; Edwards, 1957; Horst, 1968; e Scott, 1968.)

Escalas diferenciais

As escalas diferenciais para a mensuração de atitudes estão muito ligadas ao nome de L. L. Thurstone. Os métodos que criou representam tentativas para aproximar-se de escalas intervalares. Uma escala intervalar é aquela em que são conhecidas as distâncias entre os pontos do instrumento de medida, e na qual as distâncias numéricas iguais representam distâncias iguais no contínuo que está sendo medido. Tais escalas permitem a comparação de diferenças ou mudanças na atitude, pois a diferença entre um escore de 3 e um escore de 7 é equivalente à diferença entre um escore de 6 e um escore de 10, bem como à diferença entre dois outros escores quaisquer que estejam a quatro pontos de distância.

Uma escala diferencial consiste de alguns itens cuja posição na escala foi verificada por algum tipo de ordenação ou avaliação realizada por juízes. Têm sido, usados vários métodos para a obtenção de julgamentos de posição na escala: o *método de comparações aos pares* (vide Thurstone, 1927, 1928); o *método de intervalos de aparência igual* (vide Thurstone, 1929, 1931, e Thurstone e Chave, 1929); o *método de intervalos sucessivos* (vide Saffir, 1937). Está fora dos objetivos deste livro dar os pormenores de tais procedimentos; apresentaremos apenas o método de *intervalos de aparência igual*, de Thurstone, que é o mais comumente usado.

Ao selecionar os itens para a escala e ao atribuir-lhes valores, segue-se o seguinte procedimento: 1) o pesquisador reúne várias centenas de afirmações que, segundo supõe, estão ligadas à atitude que está sendo pesquisada; 2) um grande número de juízes — geralmente de 50 a 300 — trabalha independentemente e classifica tais ati-

tudes em onze grupos. No primeiro grupo, o juiz coloca as afirmações que considera mais favoráveis ao objeto; no segundo, as que considera logo abaixo; no décimo primeiro grupo, as afirmações que considera mais desfavoráveis. A sexta posição, "neutra", é definida como o ponto em que não existe "favorabilidade", nem "desfavorabilidade"; 3) o valor de escala de uma afirmação é calculado como a posição mediana (ou grupo mediano) em que foi colocada pelos juízes. São eliminadas, como ambíguas ou irrelevantes, as afirmações que tenham dispersão muito grande. 4) Faz-se a seleção final, considerando-se os itens que se distribuem uniformemente pela escala, de uma posição extrema a outra. É freqüentemente possível construir formas equivalentes da escala, a partir de itens que não foram usados na forma original.

A escala resultante *tipo Thurstone* consiste de uma série de afirmações, geralmente em número de vinte; a posição de cada afirmação, numa escala de atitude favorável-desfavorável com relação ao objeto, determinada pela classificação dos juízes. Os sujeitos, ao preencher o questionário, devem assinalar cada afirmação com que concordam, ou assinalar os dois ou três itens que estão mais próximos de sua posição.

* Em toda esta seção, com o objetivo de simplificação, a discussão é apresentada através de escalas que medem a dimensão favorável-desfavorável com relação a algum objeto. Evidentemente, uma escala pode estar voltada para outra dimensão, por exemplo, liberalismo-conservantismo de opiniões sociais, políticas ou econômicas; sermissividade-restritividade se assim por diante de opiniões a respeito de educação de crianças. Ao criar escalas tipo Thurstone, as instruções para os juízes especificam a dimensão na qual os itens devem ser colocados. Ao criar uma escala para a medida de liberalismo-conservantismo, os juízes, seriam instruídos para colocar no primeiro grupo os itens que consideram mais liberais, no décimo primeiro os que consideram mais conservadores. Os mesmos procedimentos e princípios se aplicam quando a dimensão a ser medida é favorável-desfavorável, ou qualquer outra.

A tabela 9.2 é um exemplo de itens de uma escala tipo Thurstone, e foi retirado de um estudo de MacCrone sobre atitudes com relação aos nativos da África do Sul (1937).

Tabela 9.2. Atitude com relação aos nativos da África do Sul.

| Valor na Escala | Item n.º | |
|-----------------|----------|---|
| 0,8 | 11. | Preferiria ver o povo branco perder sua posição neste país a mantê-la à custa da injustiça para com o nativo. |
| 3,1 | 3. | Acho que o homem branco, ao colocar restrições tais como o "Impedimento da Cor", para o nativo, procura, na realidade, explorá-lo economicamente. |
| 3,8 | 22. | Acho que a comunidade branca deste país tem uma dívida real de gratidão para com os missionários, pela maneira com que tentaram erguer o nativo. |
| 8,4 | 17. | Penso que o nativo é tão infantil e tão irresponsável que não se pode esperar que saiba o que é melhor para ele. |
| 8,6 | 15. | Acho que não se pode colocar o nativo em posição de confiança ou de responsabilidade. |
| 10,2 | 2. | A idéia de contato com a pele negra ou escura do nativo provoca horror e repugnância. |
| 10,3 | 1. | Acho que o nativo serve apenas para o trabalho "sujo" da comunidade branca. |

Evidentemente, os valores de escala não são apresentados no questionário, e os itens geralmente são dispostos em ordem casual, e não na ordem de seu valor de escala. A média (ou mediana)* dos valores de escala

* Thurstone, a partir da suposição de que as escalas construídas por esse método eram verdadeiras escalas de intervalo, defendeu o emprego de estatísticas adequadas para escalas de intervalo — a média e o desvio-padrão. Outros pesquisadores, a partir da su-

dos itens que o indivíduo assinala é interpretada como indicação de sua posição numa escala de atitude favorável-desfavorável com relação ao objeto.

Teoricamente, se uma escala do tipo Thurstone é inteiramente fidedigna e se a escala mede uma única atitude e não um conjunto de atitudes, um indivíduo deveria assinalar apenas itens imediatamente contíguos no valor de escala — por exemplo, os itens 15 e 17 no caso acima apresentado. Se as respostas de um indivíduo se dispersam amplamente por itens não-contíguos, seu escore tende a não ter o mesmo sentido de um escore com pouca dispersão. As respostas dispersas podem indicar que o sujeito não tem atitude ou que sua atitude não se organiza na forma suposta pela escala. Não existe razão *a priori* para esperar que todas as pessoas tenham atitudes com relação às mesmas coisas ou que as dimensões das atitudes sejam as mesmas para todas as pessoas.

Foram levantadas diversas objeções à escala tipo Thurstone. Primeiro, muitos fazem restrições à quantidade de trabalho necessária para a sua construção. Sem dúvida, o procedimento é trabalhoso, e essa é uma das razões pelas quais não tem sido tão empregado como o era inicialmente.

No entanto, Edwards (1957) diz que, diante dos desenvolvimentos recentes em técnicas para economizar tempo, o tempo e o trabalho necessário para construir uma escala pelo método de intervalos de aparência igual não são muito diversos dos necessários para construir uma escala somatória. De qualquer modo, é duvidoso que algum dia sejam criados métodos simples para a construção rigorosa de escalas. Talvez a mensuração exata de atitudes seja um problema inevitavelmente complexo.

Uma segunda crítica diz que, como o escore de um indivíduo é a média ou mediana dos valores de escala dos vários itens que

posição mais prudente de que os intervalos não são realmente iguais, defenderam o emprego da mediana como adequada para escalas ordinais.

assinala, padrões fundamentalmente diversos de atitudes podem exprimir-se no mesmo escore. Por exemplo, na escala de atitudes com relação a nativos da África do Sul, antes apresentada, um indivíduo que assinala os dois itens moderadamente "anti", 15 e 17, obtém um resultado de 8,5 (a mediana de seus valores de escala). Outro indivíduo, que assinala os itens 1, 15, 17 e 22 (talvez porque o 22 tenha, para ele, um sentido diferente do que tem para os juízes), também recebe um escore de 8,5 (a mediana dos valores de escala desses itens). Os dois indivíduos são avaliados como se tivessem o *mesmo grau de preconceito, embora o segundo, ao contrário do primeiro, tenha assinalado o item mais desfavorável da escala.*

Um problema ainda mais sério consiste em saber até que ponto os valores de escala atribuídos aos itens são influenciados pelas atitudes dos juízes. Será que a formação e as atitudes dos juízes influem na posição dos vários itens na escala? Evidentemente, essa é uma pergunta que deve ser respondida por pesquisa experimental. Alguns estudos iniciais deram apoio à opinião de que os valores de escalas atribuídos não dependiam da atitude dos juízes (Hinckley, 1932; MacGrone, 1937; Beyle, 1932; Ferguson, 1935; Pintner e Forlano, 1937; Eysenck e Crown, 1949). Essa conclusão foi questionada a partir dos resultados de alguns estudos posteriores (Hovland e Sherif, 1952).

Trabalhos recentes (Upshaw, 1965), entretanto, têm revelado que o efeito das atitudes dos juízes impõe apenas uma transformação linear nas avaliações. Portanto, a natureza dos intervalos é mantida. Pode-se concluir seguramente que o procedimento visando criar intervalos da aparência igual fornece, de fato, uma escala intervalar, dentro da amplitude principal dos valores de escala.

Escalas somatórias

Uma escala somatória, como uma escala diferencial, consiste em uma série de itens,

aos quais é solicitado a responder. No entanto, não se faz qualquer tentativa para encontrar itens que sejam igualmente distribuídos numa escala de favorável-desfavorável (ou qualquer dimensão que se queira medir). Ao contrário, são usados apenas itens que parecem *nitidamente favoráveis* ou *nitidamente desfavoráveis* ao objeto, e não os itens neutros ou “levemente” favoráveis ou desfavoráveis. Em vez de assinalar apenas as afirmações com que concorda, a pessoa indica seu acordo ou desacordo para *cada* item. A cada resposta é atribuído um valor numérico que indica o fato de ser favorável ou desfavorável; freqüentemente, as respostas favoráveis são codificadas com valores positivos, e as desfavoráveis com valores negativos. A soma algébrica dos escores das respostas do indivíduo a todos os itens separados dá seu escore total, que é interpretado como representação de sua posição numa escala de atitude favorável-desfavorável com relação ao objeto. O princípio em que se fundamenta o uso desses resultados totais como base para a colocação dos indivíduos numa escala, parece ser o seguinte: a probabilidade de concordar com qualquer item, numa série de itens favoráveis a respeito de um objeto, ou de discordar de qualquer item desfavorável varia diretamente com o grau de favorabilidade da atitude de um indivíduo. Por isso, pode-se esperar que um indivíduo com atitude favorável responda favoravelmente a muitos itens (isto é, concorde com muitos itens favoráveis ao objeto e discorde de muitos itens desfavoráveis); pode-se esperar que um indivíduo ambivalente responda a favor de alguns e contra outros; que um indivíduo com atitude desfavorável responda desfavoravelmente a muitos itens.

O tipo de escala somatória mais freqüentemente usado no estudo de atitudes sociais segue o modelo criado por Rensis Likert (1932) e é denominado *escala tipo Likert*.

Na realidade, a maioria das escalas normalmente utilizadas em pesquisas das ciên-

cias sociais são do tipo Likert. Nessa escala os sujeitos devem responder a cada item, em termos de vários graus de acordo ou desacordo, por exemplo, 1) aprovo inteiramente, 2) aprovo, 3) indeciso, 4) desaprovo, 5) desaprovo inteiramente*. Abaixo são apresentados vários itens de uma escala tipo Likert, a escala Mach IV, utilizada por Christie e Geis (1970) para medir o maquiavelismo, ou o desejo de manipular outras pessoas.

Os sujeitos foram instruídos para que indicassem o grau de concordância ou discordância com cada um dos itens da escala Mach IV, assinalando um número que corresponde a um total acordo, acordo, sem posição, desacordo ou total desacordo com cada item. Dentre os itens que se mostraram ser fortemente relacionados com o escore total da escala (isto é, aqueles que discriminaram bem nos testes iniciais entre altos e baixos resultados em maquiavelismo) estão os seguintes:

2. A melhor maneira de manipular pessoas é dizer-lhes o que querem ouvir. (+)

10. Quando você pede a alguém que lhe faça algo, é melhor dar as reais razões para se querer fazer, do que dar razões que tenham maior peso. (-)

18. É difícil progredir sem cortar caminhos aqui e ali. (+)

17. Barnum estava muito enganado quando disse que nasce um trouxa a cada minuto. (-)

15. É prudente bajular pessoas importantes. (+)

9. De qualquer modo, é melhor ser humilde e honesto do que ser importante e desonesto. (-)

O procedimento para a construção de uma escala tipo Likert é o seguinte: 1) o pesquisador reúne um grande número de itens considerados relevantes para a atitude pesquisada e que sejam claramente favoráveis ou desfavoráveis. 2) Tais itens são aplicados num grupo de pessoas representativas daquelas com as quais o questionário

* Embora Likert tenha usado cinco categorias de acordo-desacordo, alguns pesquisadores usaram número menor, enquanto outros empregaram número maior de categorias. Muitas escalas somatórias exigem apenas uma expressão de acordo ou desacordo, sem indicação de grau.

deve ser usado. As pessoas indicam sua resposta a cada item, assinalando uma das categorias de acordo-desacordo. 3) As respostas aos vários itens são codificadas de modo que uma resposta que indica a atitude mais favorável recebe o escore mais elevado. Não importa que 5 seja elevado e 1 baixo, ou vice-versa. O importante é que as respostas sejam avaliadas coerentemente, em função da orientação de atitude que indicam. O fato de “aprovo” ou “desaprovo” ser a resposta favorável a um item depende, naturalmente, de seu conteúdo e de sua redação. 4) O escore total de cada indivíduo é calculado pela soma de seus escores nos itens. 5) As respostas são analisadas para verificar quais os itens que discriminam mais claramente entre os que obtêm resultados elevados e os que obtêm resultados baixos na escala total. Por exemplo, as respostas daqueles cujos resultados totais estão no quartil superior e as dos que estão no quartil inferior podem ser analisadas, a fim de verificar, para cada item, até que ponto diferem as respostas desses grupos de critérios. Os itens que não apresentam uma alta correlação com o resultado total, ou que não provocam respostas diferentes dos que apresentam resultados elevados e baixos no resultado total, são eliminados para assegurar que o questionário seja “internamente coerente” — isto é, que cada item esteja relacionado com a mesma atitude geral.

A escala tipo Likert, como a escala Thurstone, tem sido usada amplamente em estudos de moral, de atitudes com relação a negros, de atitudes com relação a internacionalismos, e em muitos outros estudos. Mas como foi apontado por Triandis (1971), só porque uma escala é denominada “tipo Likert”, não se pode garantir que ela tenha sido construída através de uma análise de itens, como descrito anteriormente. Muitos pesquisadores usam “atalhos” e se referem às suas escalas como sendo “escalas tipo Likert”, quando apenas o formato é comum àquele tipo. Mas

uma escala do tipo Likert construída adequadamente tem várias vantagens sobre a escala Thurstone. Em primeiro lugar, permite o emprego de itens que não estão explicitamente ligados à atitude estudada. No método de Thurstone, a necessidade de acordo entre os juízes tende a limitar os itens ao conteúdo que está obviamente ligado à atitude estudada; no método de Likert, pode-se incluir qualquer item que se verifique, empiricamente, ser coerente com o escore total. Em segundo lugar, geralmente se considera que a escala tipo Likert é de construção mais simples. Em terceiro lugar, tende a ser mais fidedigna que uma escala Thurstone com o mesmo número de itens (Title e Hill, 1967). Dentro de certos limites, a fidedignidade de uma escala aumenta na medida em que se aumenta o número de possíveis respostas alternativas; a escala de tipo Likert permite a expressão de vários (geralmente cinco) graus de acordo-desacordo, enquanto o item de escala Thurstone permite apenas uma escolha entre duas respostas alternativas. Em quarto lugar, a amplitude de respostas permitidas a determinado item, numa escala tipo Likert, apresenta, na realidade, informação mais precisa sobre a opinião do indivíduo a respeito da questão apresentada por este item.

A escala de tipo Likert não pretende ser mais que uma *escala ordinal*: vale dizer, permite a ordenação de indivíduos, através da favorabilidade de sua atitude com relação a determinado objeto, mas não apresenta uma base para dizer *quanto* um indivíduo é mais favorável que outro, nem para medir a *quantidade* de mudança depois de certa experiência. Do ponto de vista do nível de mensuração que gostaríamos de obter com nossos instrumentos, isso é, naturalmente, uma desvantagem.

Outra desvantagem da escala tipo Likert é o fato de freqüentemente o escore total de um indivíduo não apresentar um significado muito claro, pois diferentes padrões de resposta aos vários itens podem conduzir ao mesmo escore.

Esta desvantagem é menos grave à medida que uma cuidadosa análise de itens garante que apenas serão incluídos, na determinação do escore final, aqueles itens que discriminam entre os escores totais altos e baixos da escala.

Já assinalamos que as escalas tipo Thurstone também estão sujeitas a este tipo de crítica, mas este se aplica ainda mais nitidamente às escalas tipo Likert, porque essas escalas permitem um número maior de diferentes possibilidades de resposta. Parece razoável supor que dois escores totais semelhantes, mas que estão baseados em diferentes combinações de respostas aos diferentes itens da escala, poderiam estar refletindo diferentes "atitudes". Assim, pode-se questionar seriamente se a escala tipo Likert satisfaz realmente os requisitos de uma escala ordinal, e muito menos ainda, os de uma escala intervalar. No entanto, apesar das deficiências da fundamentação teórica para a construção desse tipo de escala, os escores obtidos através do questionário tipo Likert freqüentemente fornecem, em termos pragmáticos, a base para uma ordenação aproximativa de pessoas quanto à característica que está sendo medida.

Em última análise, o que está envolvido nessa discussão é a questão da validade da escala. Questões de validade sempre envolvem questões de fato, as quais não podem ser resolvidas meramente através da especulação. O problema de saber se é possível atribuir, com significado, o mesmo escore a diferentes combinações de respostas é um problema que se coloca para a investigação empírica.

Escalas cumulativas

As escalas cumulativas, como as escalas diferenciais e somatórias, são constituídas por uma série de itens, diante dos quais as pessoas indicam acordo ou desacordo. Numa escala cumulativa, os itens estão relacionados entre si de forma que, idealmen-

te, um indivíduo que responde favoravelmente ao item 2, também responde favoravelmente ao item 1; quem responde favoravelmente ao item 3 também responde favoravelmente aos itens 1 e 2, e assim por diante. Assim, todos os indivíduos que respondem favoravelmente a determinado item devem ter escores mais elevados na escala total que os que respondem desfavoravelmente a esse item. O escore do indivíduo é calculado através da contagem do número de itens a que responde favoravelmente. Esse resultado o coloca na escala de atitude favorável-desfavorável, fornecida pela relação entre os itens.

Às vezes, os itens são apresentados na escala de acordo com uma ordem de favorabilidade; às vezes são apresentados de maneira casual. Geralmente, não se faz qualquer tentativa para verificar se os intervalos entre os itens são iguais; por isso, na prática as escalas cumulativas são escalas ordinais.

Uma das primeiras escalas usadas na mensuração de atitudes, a escala Bogardus de distância social (vide Bogardus, 1925, 1928, 1933), pretendia ser uma escala de tipo cumulativo. A escala de distância social, que se tornou uma técnica clássica para a mensuração de atitudes quanto a grupos étnicos, apresenta certo número de relações às quais os membros do grupo poderiam ser admitidos. A pessoa deve indicar, para nacionalidades ou grupos étnicos especificados, as relações às quais estaria disposta a admitir membros de cada grupo. Sua atitude é medida pela proximidade de relações que está disposta a aceitar. A escala de tipo Bogardus é exemplificada na tabela 9.3.

Os itens usados na escala de Bogardus (isto é, os títulos das colunas na tabela 9.3) foram escolhidos com fundamentos lógicos. Parece razoável esperar que um indivíduo que assinala 4 com relação aos chineses, indicando assim que estaria disposto a aceitá-los em emprego na sua profissão, geralmente também assinalaria 5, e não assinalaria 6 ou 7. (Aqui, como em ou-

Tabela 9.3. Escala Bogardus de distância social.

Instruções: Para cada raça ou nacionalidade abaixo indicadas, assinale cada uma das classificações em que você estaria disposto a admitir a pessoa média dessa raça ou nacionalidade (nem as melhores pessoas que você conheceu nesse grupo, nem as piores). Responda de acordo com as suas primeiras reações.

| | A parentesco por casamento | A meu clube como amigos | A minha rua como vizinhos | A empregos em minha profissão | A cidadania em meu país | Só como visitantes do meu país | Excliria do meu país |
|------------|----------------------------|-------------------------|---------------------------|-------------------------------|-------------------------|--------------------------------|----------------------|
| Inglês | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Negro | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Francês | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Chinês | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Russo etc. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

tras escalas, o conteúdo do item deve ser considerado ao decidir se uma resposta "Sim" deve ser avaliada como favorável ou desfavorável. Como 6 e 7 são, fundamentalmente, afirmações de exclusão, a ausência de sinal constitui a resposta favorável a esses dois itens. Assim, nem 6 nem 7 devem ser assinalados para determinado grupo, se qualquer dos outros números estiver assinalado). Se o indivíduo não assinalou 3 (disposto a admitir em minha rua, como vizinhos), deve-se esperar, pela lógica, que também não assinalaria 2 ou 1.

No conjunto, verificou-se que esses itens formam uma escala cumulativa. No entanto, na prática ocorrem algumas inversões. Por exemplo, alguns indivíduos que não gostariam de viver com porto-riquenhos no mesmo edifício, não teriam objeções a eles como membros de um clube informal (vide Deutsch e Collins, 1951). Embora não seja pouco freqüente a inversão de respostas de indivíduos à escala de distância social, é relativamente difícil encontrar inversões por um grupo todo. Por isso, a escala de distância social tem sido usada de maneira eficiente para comparar as atitudes

de diferentes grupos de pessoas com relação a várias nacionalidades. Pode-se notar que as inversões podem ser quase sempre interpretadas através da interferência de algum outro fator, além da atitude do indivíduo com relação ao grupo considerado — por exemplo, a imagem que os outros teriam dele, ao saber que vive em certo bairro, ou sua expectativa com relação ao impacto, nos valores de imóveis, na aceitação de indivíduos do grupo minoritário na sua rua.

No fim da década de 20 e no início da década de 30, com o aparecimento dos métodos de Thurstone e Likert, deixou-se de dar atenção às escalas cumulativas. No entanto, na década de 40 vimos reaparecer o interesse por técnicas destinadas a verificar se os itens de uma escala realmente têm uma relação cumulativa, pareçam ou não cumulativas nos termos do senso-comum; nessa época ocorreu também um rápido desenvolvimento dessas técnicas.

Esse interesse renovado estava ligado ao trabalho de Louis Guttman (1944) e outros, sobre o desenvolvimento de escalas unidimensionais — isto é, escalas formadas por

itens que não levantam questões ou que não envolvem fatores que sejam estranhos à característica que está sendo medida. Na verdade, o objetivo de Louis Guttman era desenvolver uma técnica para determinar se o domínio de um estímulo era unidimensional. Coube a outros aplicar sua técnica como um meio de construir escalas de atitude.

Alguns pesquisadores indicaram que as escalas de Likert e Thurstone, embora explicitamente destinadas a medir “uma atitude”, continham afirmações a respeito de vários aspectos do objeto considerado. Assim, Carter (1945) indicou que a forma A da escala de Peterson para atitudes com relação à guerra (uma escala tipo Thurstone) tinha, como afirmação mais favorável, a frase: “A guerra é gloriosa; sua afirmação mais desfavorável era: “Não existe justificativa concebível para a guerra”; como seu ponto médio, a frase: “Nunca penso na guerra e esta nunca me interessa”. Carter comenta que é difícil pensar nessas afirmações como pertencendo a uma única dimensão. Sugeriu que algumas afirmações, como: “Os benefícios da guerra raramente compensam suas perdas, mesmo para o vencedor”, e: “As guerras defensivas são justificáveis, mas as outras não”, pertencem a duas escalas diferentes, uma referente aos resultados econômicos da guerra, a outra à ética da atividade militar. Já se sustentou que a combinação de itens referentes a aspectos diferentes do objeto impediria a especificação exata do que a escala estava medindo, e também explicava a dispersão de respostas que dificultavam a atribuição de sentido claro ao escore baseado na mediana dos itens assinalados.

Apareceram vários processos para enfrentar esse problema. Aqui, discutiremos apenas a técnica criada por Guttman, geralmente denominada *análise de escalograma* ou *método de escalograma*. Um dos principais objetivos dessa técnica é verificar se a atitude ou característica que está sendo estudada (tecnicamente denomina-

da o “universo de conteúdo” ou o “universo de atributos”) realmente envolve apenas uma dimensão. No procedimento de Guttman, considera-se que um “universo de conteúdo é unidimensional apenas se permite uma escala cumulativa perfeita ou quase perfeita. — isto é, se é possível dispor todas as respostas, de qualquer número de pessoas que responderam, num padrão do seguinte tipo:

Tabela 9.4. Um exemplo de escala cumulativa.

| Escore | Diz | | “sim” ao item | | Diz | | “não” ao item | |
|--------|-----|---|---------------|---|-----|---|---------------|---|
| | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 |
| 4 | X | X | X | X | | | | |
| 3 | | X | X | X | X | | | |
| 2 | | | X | X | X | X | | |
| 1 | | | | X | X | X | X | |
| 0 | | | | | X | X | X | X |

O importante nesse padrão é que — quando ele ocorre — um determinado escore numa série de itens tem sempre o mesmo sentido; o fato de conhecer o escore de um indivíduo permite dizer, sem consultar seu questionário, quais foram exatamente os itens que aceitou. Consideremos, por exemplo, os itens seguintes, com os quais as pessoas que respondem devem concordar ou não:

1. Não penso que funcionários públicos se preocupem muito com aquilo que pessoas como eu pensam.
2. Votar é a única maneira pela qual pessoas como eu podem dizer alguma coisa sobre a atuação do governo.
3. Pessoas como eu não têm nada a dizer sobre aquilo que o governo faz.
4. Algumas vezes governo e política parecem coisas tão complicadas que pessoas como eu não podem realmente compreender o que ocorre.

Se se verifica que esses itens formam uma escala cumulativa perfeita, sabemos, por exemplo, que *todos* os indivíduos com um resultado de 2 na escala concordariam com os primeiros dois itens e discordariam com os dois últimos itens. Também saberíamos que uma pessoa com um resultado de 3 discordaria apenas do último item, o de que “política e governo parecem tão

complicados que uma pessoa como eu não pode realmente entender o que ocorre”, e uma pessoa com o resultado 1 concordaria com o primeiro item, “Não penso que os funcionários públicos se preocupem muito com aquilo que pessoas como eu pensam”.

Na prática, raramente ou nunca se encontram, na pesquisa social, escalas cumulativas perfeitas, ou unidimensionais, mas muitas vezes é possível criar escalas que disso se aproximam. A análise de escalograma usa vários critérios para decidir se uma determinada série de itens pode ou não ser encarada, com proveito, como próxima de uma escala unidimensional perfeita. O mais importante desses critérios é a reprodutibilidade das respostas — a proporção de respostas de um grande número de pessoas que na realidade está no padrão acima apresentado. Esse padrão contém todas as respostas a itens específicos que seriam previstas a partir de um conhecimento do resultado total de um indivíduo na série de itens (seu “tipo escalar”). Assim, a proporção de respostas reais que caem no padrão apresenta uma medida da extensão em que as respostas específicas são “reproduzíveis” a partir do resultado total. Guttman e seus colaboradores estabeleceram que 0,90 é a reprodutibilidade mínima necessária para que uma série de itens seja vista como próxima de uma escala perfeita. No estudo inicial, encontrou-se, nos itens precedentes, uma reprodutibilidade de 0,94 (Campbell *et al.*, 1954). Em Stouffer *et al.* (1950) são apresentados exemplos de tais escalas.

A técnica de Guttman e outras técnicas semelhantes representam contribuições fundamentais para a metodologia de construção e análise de questionários. No entanto, devemos lembrar duas limitações referentes ao emprego de escalas unidimensionais: 1) uma escala desse tipo pode não ser a base mais eficiente para medir atitudes com relação a objetos complexos ou para fazer predições quanto ao comportamento com relação a tais objetos; 2) determinada escala pode ser unidimensional pa-

ra um grupo de indivíduo, e não para outro. A mesma escala pode ser unidimensional em um momento para um grupo e não ser mais em outro momento para o mesmo grupo. Como já foi apontado anteriormente, a tendência na última década referente à construção de escalas de atitudes tem sido em direção ao desenvolvimento de escalas que medem várias dimensões.

Há uma limitação maior comum a todos os três procedimentos de construção das escalas anteriormente descritos: eles determinam um número — isto é, *um ponto*, para representar a posição do indivíduo. A teoria atual de atitude demonstra a importância da amplitude de pontos aceitáveis para a atitude em estudo. Sherif, Sherif e Nebergall (1965) referiram-se a *latitudes de aceitação e rejeição*, isto é, cada indivíduo tem não só apenas uma posição preferida em um *continuum* de atitude, mas também algumas outras posições que ele considera aceitável. Além disso, há uma posição de atitude que ele rejeita veementemente, mas há uma *extensão* de posição que ele considera passível de objeção. Duas pessoas com diferentes *latitudes de aceitação* poderão ter *latitudes de rejeição* similares.

Algumas modificações de técnicas de escalas

Nem todas as tentativas para quantificar as atitudes caem na classificação de escalas que acabamos de descrever. Sobre tudo nos últimos anos, criaram-se alguns procedimentos que utilizam alguns aspectos de técnica de escalas, mas que não podem ser descritos, corretamente, como escalas diferenciais, somatórias ou cumulativas. Discutiremos dois desses procedimentos: a “técnica-Q” (Q-Sort) e o *diferencial semântico*.

A técnica-Q

As operações exigidas numa técnica-Q são semelhantes às primeiras etapas na

construção de uma escala Thurstone pelo método de intervalos de aparência igual. A pessoa recebe um grande número de afirmações que, segundo se acredita, são significativas para o assunto pesquisado, e deve separá-la num certo número especificado de pilhas — geralmente nove ou onze — de acordo com algum critério. Na técnica de Thurstone, como lembramos, esse processo de escolha tem como objetivo a atribuição de valores de escala às afirmações: o critério de escolha não é o fato de o juiz concordar ou não com uma afirmação, mas seu julgamento quanto ao fato de esta ser favorável ou desfavorável ao objeto. Na técnica-Q, o objetivo da seleção é conseguir uma imagem da opinião do indivíduo sobre o objeto considerado ou de sua atitude com relação a este. O critério para a colocação das afirmações nas várias pilhas é a extensão de seu acordo com elas; o padrão de sua escolha das afirmações constitui os dados para a análise de sua posição. Na técnica-Q introduzimos uma restrição que não existe no processo de escolha de Thurstone: especificação do número de cartões que devem ser colocados em determinada pilha. A fim de simplificar a análise estatística, os números são geralmente especificados de modo que a escolha forme, grosseiramente, uma distribuição normal.

Consideremos um exemplo do emprego de uma técnica-Q no contexto em que tem sido mais freqüentemente empregada — o estudo da personalidade. Mais especificamente, esse estudo referia-se às adesões individuais ao candidato durante a campanha presidencial em 1968 do senador Eugene McCarthy (Brown e Ellithorp, 1970). Cada pessoa recebia um conjunto de 60 afirmações, escritas em cartões individuais. Seguem-se exemplos dessas afirmações, representando algumas articulações teoricamente hipotetizadas entre o adepto individual de McCarthy e o movimento McCarthy:

1. Creio que um homem como McCarthy, mesmo que tenha falhado, deveria fazer o país ficar orgulhoso. Ele era um homem de quem eu seria adepto.

2. O fenômeno McCarthy permitiu-me ter esperanças de que mudanças ocorrerão. Estou confiante de que em 1972 alguém surgirá para exercer a liderança que necessitamos.

3. Em última análise, talvez não tenhamos sido derrotados. Talvez McCarthy tenha feito o seu trabalho. Ele se livrou de L.B.J., não? Além disso, de qualquer maneira não está ainda acabado.

Solicitava-se a cada pessoa que desse seu ponto de vista sobre cada item, distribuindo os cartões ao longo de um *continuum* com numeração que variava de 5 (maior discordância), 0 (sem posição) a +5 (maior concordância), que estavam dispostos em um grande quadro. Os participantes (34 estudantes que apoiavam McCarthy) distribuíam cada um dos 60 itens, colocando-os em algum ponto ao longo do *continuum*. Os 34 conjuntos de ordenações eram, então, inter-correlacionados, e se procedia à análise, através da análise fatorial dos dados. Neste estudo específico, a análise fatorial revelou a existência de três dimensões nos dados. Verificou-se que os participantes estavam vinculados ao movimento McCarthy, com base em pelo menos uma dessas três dimensões.

As técnicas estatísticas apropriadas para a análise de dados da técnica-Q são um tanto controversas. Alguns autores sustentam que *análise fatorial** é essencial; outros discordam. No estudo sobre os adeptos de McCarthy empregou-se a análise fatorial.

Embora a técnica-Q tenha sido usada mais freqüentemente em estudos desse tipo, onde se enfatiza a auto-imagem e as imagens de outras pessoas, o método é aplicável ao estudo de outras atitudes. É possível apresentar, às pessoas, conjuntos de afirmações sobre métodos de educação de crianças so-

* A análise fatorial é um método de descoberta do elemento ou elementos em comuns, subjacentes a um conjunto de medidas. Para uma discussão da análise fatorial, ver Corsuch (1974), Harmon (1967) ou Rummel (1970).

bre relações patrões-empregados, ou sobre chineses e americanos e pedir-lhes que as distribuam segundo o grau de seu acordo ou desacordo com cada afirmação. Os dados assim obtidos poderiam ser usados, por exemplo, para comparar as opiniões de determinado indivíduo a respeito de diferentes grupos étnicos, ou para comparar as opiniões de diferentes indivíduos a respeito de determinado grupo ou objeto.

O diferencial semântico

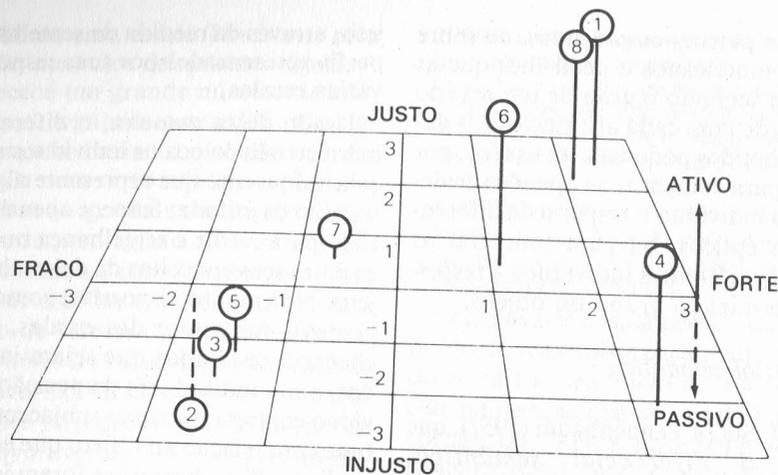
Osgood, Suci e Tannenbaum (1957), que criaram o *diferencial semântico*, descrevem-no como um método para medir o significado que um objeto tem para um indivíduo. O método pode ser também pensado como uma série de escalas de atitude. A pessoa deve avaliar determinado conceito (por exemplo, irlandês, “republicano”, “mulher”, “eu, tal como sou”) numa série de escalas bipolares de avaliação de sete pontos. Pode-se avaliar qualquer conceito — uma questão política, uma pessoa, uma instituição, uma obra de arte. As escalas de sete pontos incluem escalas bipolares tais como as seguintes: 1) justo-injusto, limpo-sujo, bom-mau, valioso-sem valor; 2) grande-pequeno, forte-fraco, pesado-leve; 3) ativo-passivo, rápido-lento, quente-frio.

É possível usar as respostas do indivíduo para verificar se, para ele, dois conceitos ou objetos são semelhantes ou diferentes. Por exemplo, será que sua imagem de “eu, tal como sou” coincide com sua imagem de “eu, como gostaria de ser”, ou será que tais imagens são muito diferentes? Nesse emprego, obtém-se um “perfil” do significado que cada conceito tem para o indivíduo através de linhas entre os pontos assinalados, para cada conceito, em cada uma das escalas. Criou-se uma técnica para medir a extensão da semelhança entre os dois perfis. De forma semelhante, é possível comparar os conceitos que dois indivíduos têm a respeito de determinado ob-

jeto, através da medida da semelhança dos perfis apresentados por suas respostas nas várias escalas.

Usado dessa maneira, o diferencial semântico não coloca os indivíduos numa escala subjacente que represente alguma dimensão de atitude; fornece apenas um método para medir a semelhança ou diferença entre seus conceitos de determinado objeto. No entanto, é possível somar as respostas a subgrupos das escalas, a fim de chegar a resultados que sejam interpretados como indicadores da posição do indivíduo em três dimensões subjacentes de atitude com relação ao objeto que está sendo avaliado. Tais dimensões foram identificadas pelo uso de processos de análise fatorial no exame de respostas de muitos indivíduos com relação a muitos conceitos ou objetos. Verificou-se que, de maneira mais ou menos coerente, as escalas denominadas (1) acima, parecem agrupar-se; vale dizer, um indivíduo tende a colocar determinado objeto em posição semelhante em cada uma dessas escalas. De forma semelhante, as escalas denominadas (2) tendem a agrupar-se, e o mesmo ocorre com as escalas denominadas (3). O conteúdo manifesto de cada escala não indica claramente a dimensão subjacente de atitude para a qual é relevante, o que significa, que exemplo, avaliar “republicano” como quente ou frio, limpo ou sujo. No entanto, a partir de agrupamento coerente das escalas, Osgood e seus colaboradores inferiram que os três subgrupos medem as três dimensões seguintes de atitude: 1) a valoração pelo indivíduo, do objeto ou conceito que está sendo avaliado, o que corresponde à dimensão favorável-desfavorável das escalas mais tradicionais de atitude; 2) a percepção que o indivíduo tem da *potência* ou poder do objeto ou do conceito; 3) sua percepção da *atividade* do objeto ou conceito.

A figura 9.1. é uma adaptação de um diagrama de Osgood e seus colaboradores, e exemplifica o emprego do diferencial semântico na apresentação dos significados



Chave para a Figura

- | | |
|----------------------|------------------|
| 1. Taft | 5. Truman |
| 2. Política na China | 6. Bomba Atômica |
| 3. Socialismo | 7. Nações Unidas |
| 4. Stálin | 8. Eisenhower |

Figura 9.1. Modelo do Espaço Semântico para republicanos favoráveis a Taft. Adaptado de Osgood *et al.* (1957). Cada um dos oito conceitos é numerado, de acordo com a chave abaixo apresentada. O ponto em que começa a linha para determinado conceito mostra a avaliação desse conceito nas escalas *justo-injusto* e *forte-fraco*. O comprimento da linha mostra a avaliação na escala *passivo-ativo*; uma linha cheia indica uma avaliação com relação à extremidade ativo da escala, uma linha pontilhada mostra uma avaliação na extremidade *passivo* da escala. Pode-se ver, por exemplo, que Stalin (4) foi avaliado como muito injusto, muito forte e muito ativo; Truman (5), como um pouco injusto, um pouco fraco e um pouco ativo.

de vários e diferentes conceitos, para um grupo de republicanos num estudo realizado pouco antes da eleição presidencial de 1952 nos Estados Unidos da América. Pelo diagrama pode-se ver, por exemplo, que, para essas pessoas, o conceito "Truman" é muito diferente do conceito "Taft", mas muito semelhante ao conceito "socialismo". (O senador Robert A. Taft era um conservador, membro do Partido Republicano, e era um dos principais candidatos à indicação republicana para as eleições presidenciais, em 1952; o partido, entretanto, indicou Eisenhower.)

O conceito "Stálin" é muito semelhante ao conceito "Taft", nas dimensões de potência e atividade, mas *muito diferente dele na dimensão valorativa*.

Osgood e seus colaboradores sugeriram que o diferencial semântico permite a mensuração e a comparação por várias pessoas de vários objetos; supõem que o instrumento de medida não seja muito influenciado pela natureza do objeto que está sendo medido ou pelo tipo de pessoa que usa a escala. Se isso é realmente verdade, o diferencial semântico seria uma solução para muitos dos problemas da mensuração de atitude. Osgood, Suci e Tannenbaum apresentam algumas provas de que diferentes tipos de pessoas usam, de maneiras semelhantes, as escalas. Assim, indicam que, em suas reações, os republicanos favoráveis a Taft, os favoráveis a Eisenhower e os democratas usam as mesmas dimensões subjacentes nas suas reações; o mesmo ocorre

com normais e esquizofrênicos; o mesmo ocorre com estudantes coreanos e japoneses, que foram estudar nos Estados Unidos, e norte-americanos. Essa é uma prova notável de que as escalas são comparáveis quando se empregam diferentes tipos de pessoas, embora pareça possível que, se fossem comparadas pessoas com educação muito diversa, não aparecessem os mesmos resultados.

No entanto, Osgood e seus colaboradores apresentam muitas provas que indicam que o instrumento de medida não é inteiramente comparável em diferentes conceitos. Os sentidos das escalas e sua relação com outras escalas variam consideravelmente com o conceito que está sendo julgado. O que é bom, por exemplo, depende, em grande parte, do conceito que está sendo julgado; "forte" pode ser bom ao julgar atletas, mas não ao julgar aroma e odores. A consequência disso é que pode ser muito difícil criar escalas de avaliação que apresentem uma medida coerente das dimensões subjacentes, independentemente dos conceitos que estão sendo julgados.

Ao concluir nossa discussão de escalas de atitudes, deve-se indicar que o campo de construção e análise de escalas está em período de rápido desenvolvimento e mudança. Além dos trabalhos que discutimos, nos últimos anos apareceram contribuições fundamentais de Coombs (1964); Dawes, Singer e Lemons (1972); Lazarsfeld (1957), Loevinger (1947, 1948), Suppes e seus colaboradores no Applied Mathematics and Statistics Laboratory da Universidade de Stanford (Davidson, Suppes e Siegel, 1957), Torgerson (1958); e outros.

Resumo

Neste capítulo discutimos vários métodos de construção de escalas, isto é, de dis-

tinção entre objetos ou indivíduos, em função de seu grau de determinada característica. Aqui, tal como se disse no caso de outras técnicas de mensuração, apresentamos dúvidas quanto à fidedignidade e à validade. Embora muitos tenham pesquisado a fidedignidade e a validade de suas medidas, ainda é certo dizer — como o fez um autor (Ferguson, 1957) — que tem havido "mais mensuração que validação". Como já foi dito antes, existem razões compreensíveis para a falta de atenção à validade — sobretudo a dificuldade para determinar quais deveriam ser os critérios adequados para medidas de atributos complexos e para a obtenção dos dados necessários. No entanto, alguns pesquisadores têm demonstrado que, com um pouco de imaginação inventiva, é possível criar. Talvez se deva esperar que, em campos relativamente novos de pesquisa — como as ciências sociais — mais atenção se deva dar inicialmente à criação de diferentes técnicas de mensuração que à sua validação. No entanto, parece evidente que, agora, a pesquisa nas ciências sociais progrediria com maior atenção à validade dos instrumentos usados.

Os exemplos e discussões deste capítulo referiram-se às escalas de atitudes e não outras propriedades individuais ou de grupos. Disto não deve ser entendido que as técnicas de construção de escalas não são aplicáveis a outras características de indivíduos ou coletividades mas sim como uma maneira de se apresentar técnicas de construção de escalas de uma forma breve e clara. Há, neste capítulo, referências bibliográficas para o leitor, sobre mensuração e escalas, com muitos exemplos de colocação de outros objetos em escala, além de atitudes, bem como exemplos de técnicas de escalas mais sofisticadas.