

Princípios de elaboração de escalas psicológicas

[Luiz Pasquali](#) ¹

RESUMO

A teoria e o modelo de elaboração de escalas psicológicas descritas aqui são aplicáveis à construção de testes psicológicos de aptidão, de inventários de personalidade, de escalas psicométricas de atitude e do diferencial semântico. O modelo baseia-se nos três grandes polos ou procedimentos, que chamaremos de procedimentos teóricos, procedimentos empíricos (experimentais) e procedimentos analíticos (estatísticos). O primeiro pólo enfoca a questão da teoria, que deve fundamentar qualquer empreendimento científico, no caso a explicitação da teoria sobre o construto ou objeto psicológico para o qual se quer desenvolver um instrumento de medida, bem como a operacionalização do construto em itens. Esse polo explicita a teoria do traço latente, bem como a explicitação dos tipos, categorias, de comportamentos que constituem uma representação adequada do mesmo traço. A operacionalização do construto segue 12 regras de construção de itens. O polo empírico ou experimental define as etapas e técnicas da aplicação do instrumento piloto e da coleta da informação para proceder à avaliação da qualidade psicométrica do instrumento. O polo analítico estabelece os procedimentos de análises estatísticas a serem efetuadas sobre os dados para levar a um instrumento válido, preciso e, se for o caso, normatizado.

Unitermos: Construção de Escalas; Procedimentos Teóricos; Procedimentos Experimentais; Procedimentos Analíticos

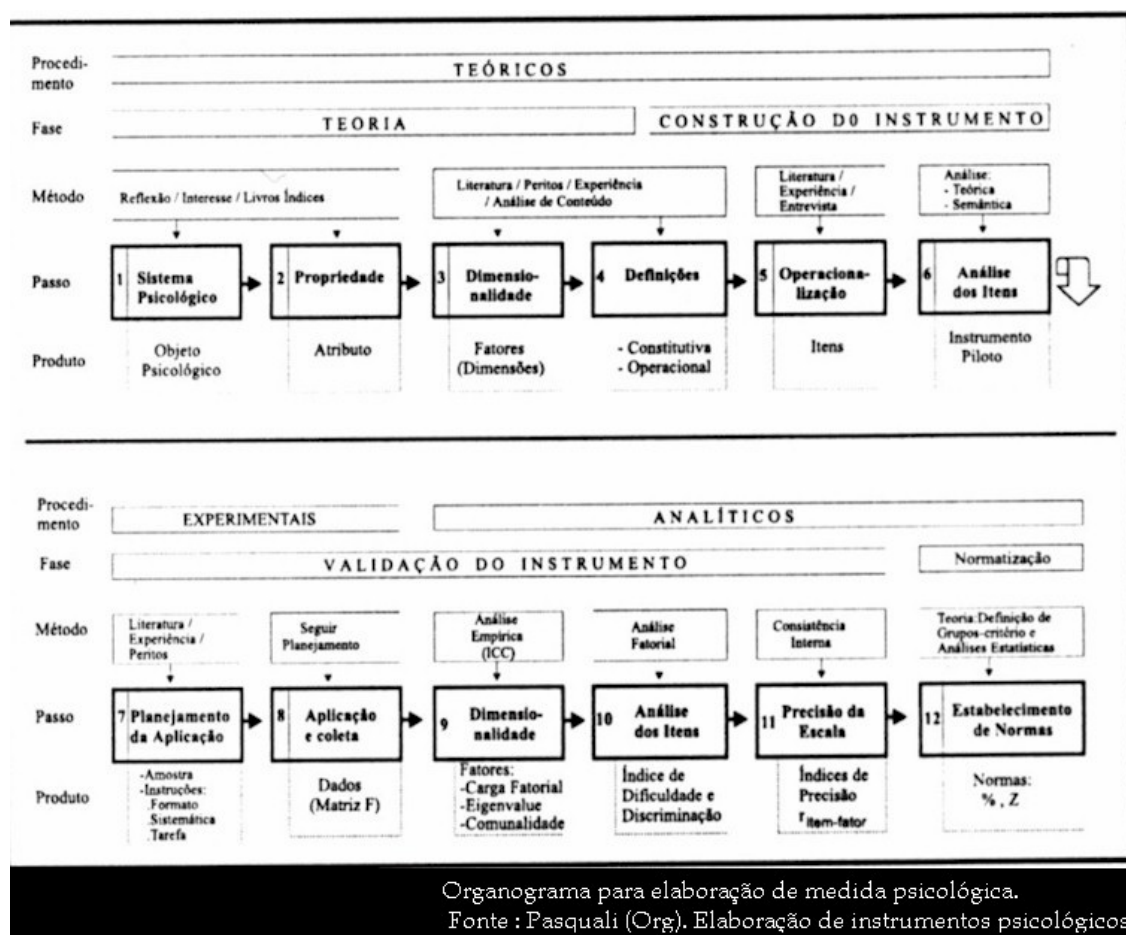
ABSTRACT

Principles of Elaboration of Psychological Scales

The theory and the model of psychological scale construction presented in this paper are applicable to psychological instruments, such as aptitude tests, personality inventories, attitude scales, and semantic differential. The model is based upon three major pillars or procedures, which we may call the theoretical pole, the empirical (experimental) pole, and the analytical (statistical) pole. The theoretical procedures deal with the theory, which should bases any scientific endeavor; in the present case, this means the explication of the theory of the constructs or the psychological object, for which one is willing to build a measurement instrument. This pole will elaborate the theory of the constructs, as well as the behavior categories in which such constructs manifest themselves. Twelve rules of thumb will guide the construction of the items in this phase. The experimental procedures will define the steps needed for the adequate gathering of the empirical evidence necessary to demonstrate the psychometric qualities of the measuring instrument. The analytical procedures, in turn, will define the appropriate statistical analyses needed to statistical demonstrate these psychometric properties of the instrument, as well as the eventual standardization of it.

Key words: Scale Construction; Theoretical Procedures; Experimental Procedures; Analytical Procedures

A teoria e os procedimentos de construção de escalas psicológicas estão ilustrados na [figura 1](#).



O modelo, que é detalhado na figura 1, baseia-se nos três grandes polos, que chamaremos de procedimentos teóricos, procedimentos empíricos (experimentais) e procedimentos analíticos (estatísticos). O presente artigo irá focalizar unicamente os procedimentos teóricos, que são pouco abordados na literatura da avaliação psicológica, enquanto existem abundantes tratados sobre os procedimentos tanto experimentais e mais ainda sobre os procedimentos analíticos (em livro a ser brevemente lançado, todos esses procedimentos serão amplamente abordados, ver Pasquali, L. org.).

O polo teórico na elaboração de instrumentos psicológicos de medida enfoca a questão da teoria que deve fundamentar qualquer empreendimento científico, no caso a explicitação da teoria sobre o construto ou objeto psicológico para o qual se quer desenvolver um instrumento de medida, bem como a operacionalização do construto em itens. Esse polo explicita a teoria do traço latente, bem como os tipos, categorias, de comportamentos que constituem uma representação adequada do mesmo traço. Os procedimentos teóricos devem ser elaborados para cada instrumento, dependendo, portanto, da literatura existente sobre o construto psicológico que o instrumento pretende medir.

A teoria é, infelizmente ainda, a parte mais fraca da pesquisa e do conhecimento psicológicos, o que tem como consequência a precariedade dos atuais instrumentos psicométricos de medida nessa área. Tal ocorrência explica, em parte, porquê os psicometristas sistematicamente fogem da explicitação de uma teoria preliminar e iniciam a construção do instrumento pela coleta intuitiva e mais ou menos aleatória de uma amostra de itens, que

dizem possuir "face validity", isto é, que parecem cobrir o traço para o qual eles querem elaborar o instrumento de medida. Embora isto não pareça muito científico, infelizmente é o que ocorre com mais frequência na construção de instrumentos psicológicos. A inexistência de teorias sólidas sobre um construto não deve ser desculpa para o psicometrista fugir de toda a especulação teórica sobre ele. É obrigação dele levantar, pelo menos, toda a evidência empírica sobre o construto e procurar sistematizá-la e, assim, chegar a uma miniteoria sobre ele, que o possa guiar na elaboração de um instrumento de medida para o tal construto.

Este teorizar implica em resolver algumas questões básicas que permitem, então, se poder enveredar para a construção adequada de um instrumento de medida dos construtos assim elaborados. Especificamente, precisa-se estabelecer a dimensionalidade do construto, definir constitutiva e operacionalmente o mesmo e, no final, operacionalizá-lo em tarefas comportamentais.

A DIMENSIONALIDADE

A dimensionalidade do atributo diz respeito à sua estrutura interna, semântica. O atributo constitui uma unidade semântica única ou é ele uma síntese de componentes distintos ou até independentes? Deve ele ser concebido como uma dimensão homogênea ou deve-se nele distinguir aspectos diferenciados? A resposta a este problema obviamente deve vir ou da teoria sobre o construto e/ou dos dados empíricos disponíveis sobre ele, sobretudo dados de pesquisas que utilizaram a análise fatorial na verificação dos dados, pois o que está em jogo aqui é a questão de decidir se o construto é uni ou multifatorial. Os fatores que compõem o construto (o atributo) são o produto deste passo. Um exemplo: seja o processo cognitivo o objeto psicológico a ser pesquisado; a propriedade deste objeto psicológico a ser estudado é a inteligência verbal. Pergunta-se: é esta inteligência verbal um construto único ou deve-se distinguir nele componentes diferentes? Os dados empíricos disponíveis mostram que a inteligência verbal é composta por, pelo menos, dois fatores bem distintos e praticamente independentes, a saber: compreensão e fluência verbais. Consequentemente, se quiser pesquisar a inteligência verbal e construir para tal um instrumento de medida, não se poderá prescindir de conhecer e levar em conta o fato de que essa inteligência apresenta dois fatores distintos, cuja medida de ambos exige instrumentos diferentes. Claro, o pesquisador pode decidir-se por estudar somente a inteligência verbal compreendida sob seu aspecto de compreensão verbal e prescindir de se preocupar com a fluência verbal. Mas, neste, caso o atributo de interesse de estudo não é mais a inteligência verbal e sim a compreensão verbal.

A questão da dimensionalidade constitui, talvez, o ponto mais crítico na elaboração dos instrumentos psicológicos, porque toda esta parte resulta essencialmente da teoria psicológica, a qual concebe, define e estrutura os construtos psicológicos. A tarefa da construção da teoria psicológica não é tarefa específica do psicometrista e sim do psicólogo teórico. O psicometrista deveria poder contar com essa teoria e com base nela fundamentar a construção dos instrumentos de medida. A existência de teorias ou fantasias as mais variadas sobre praticamente qualquer construto em Psicologia, torna, a tarefa do psicometrista, quase uma tragédia quando quer construir instrumentos para medir construtos sobre os quais os psicólogos não se entendem. Dessa sorte, o psicometrista acaba decidindo-se em construir um instrumento para medir um construto concebido segundo algum psicólogo. E ali existe uma legião enorme de psicólogos teóricos, desde os behaviorista até os dialéticos, que falam linguagens quase totalmente estranhas um em relação ao outro. Infelizmente esta é a situação da teoria psicológica atual. Para caricaturar, imagine o seguinte: um físico vai construir um instrumento para medir o comprimento de objetos físicos. Mas, se para poder efetuar tal

empreendimento, ele tivesse que decidir sobre "bem, comprimento entendido segundo quem?" Tal pergunta careceria de sentido e seria ridícula fosse ela feita sobre comprimento ou outras propriedades da matéria (pelo menos, na sua grande maioria). Mas, no caso do psicometrista, tal pergunta infelizmente é corriqueira, qualquer que seja o construto que ele queira estudar e medir, o que vem mostrar o estado primitivo em que vive a teoria psicológica. Precariedade da teoria psicológica é a principal responsável pela fuga, por parte dos psicometristas, de basear a construção dos instrumentos psicológicos numa teoria prévia e testá-los em seguida através da metodologia científica. Essa fuga permite que o psicometrista parta de uma coleção atabalhoada de itens para, em seguida, ver que eles estão medindo alguma coisa psicologicamente relevante.

Esse estado de coisas deveria e deve obrigar o psicometrista a expor ou elaborar uma miniteoria sobre o que entende pelo construto que pretende medir. Felizmente, já existe razoável abundância de dados empíricos sobre muitos construtos psicológicos, com base nos quais o psicometrista poderá desenvolver uma miniteoria do construto, a qual irá guiar a construção do seu instrumento de medida. Os dados empíricos que serão coletados por meio do instrumento assim construído irá decidir se sua miniteoria tem ou não alguma consistência. Isso não é uma tragédia, é a própria lógica da pesquisa empírica, isto é, a verificação empírica que pode ou não confirmar a validade de uma teoria: a verdade científica é sempre relativa, nunca será um dogma, e portanto sempre reformável.

A DEFINIÇÃO DOS CONSTRUTOS

Decidida a dimensionalidade dos construtos, é preciso conceituá-los de forma detalhada, novamente baseando-se na literatura pertinente, nos peritos da área e na própria experiência. O problema deste passo é, portanto, a conceituação clara e precisa dos fatores para os quais se quer construir o instrumento de medida. A tarefa aqui é dupla, tendo como resultado dois produtos, a saber: as definições constitutivas e as definições operacionais dos construtos.

1. A Definição Constitutiva

Um construto definido por meio de outros construtos representa uma definição constitutiva. Nesse caso, o construto é concebido em termos de conceitos próprios da teoria em que ele se insere. Definição constitutiva é a que tipicamente aparece como definição de termos em dicionários e enciclopédias: os conceitos são ali definidos em termos de outros conceitos; isto é, os conceitos, que são realidades abstratas, são definidos em termos de realidades abstratas. Por exemplo, se a inteligência verbal for definida como a "capacidade de compreender a linguagem", estamos diante de uma definição constitutiva, porque capacidade de compreender constitui uma realidade abstrata, um construto, um conceito.

As definições constitutivas são de extrema importância no contexto da construção dos instrumentos de medida, porque elas situam o construto, exata e precisamente dentro da teoria desse construto, dando, portanto, as balizas e os limites que ele possui. Assim, se definida a assertividade, como a capacidade de dizer não, a capacidade de expressar livremente sentimentos positivos e negativos, a capacidade de expor idéias sem receio, etc., estamos dando os limites semânticos que este conceito deve respeitar dentro da teoria de assertividade que será utilizada para a construção do instrumento que o vai medir. Definições dessa natureza impõem limitações definidas sobre o que se deve explorar quando o construto estiver sendo medido, limitações não somente em termos de fronteiras que não podem ser ultrapassadas, porém, mais ainda, em termos de fronteiras que devem ser atingidas. De fato,

normalmente um instrumento que mede um construto não chega a cobrir toda a amplitude semântica de um conceito. Assim, boas definições constitutivas vão permitir em seguida avaliar a qualidade do instrumento, que mede o construto em termos do quanto de sua extensão semântica é coberta pelo instrumento, surgindo daí instrumentos melhores e piores à medida que medem mais ou menos da extensão conceitual do construto, extensão essa delimitada pela definição constitutiva desse mesmo construto.

2. A Definição Operacional

Com as definições constitutivas estamos ainda no terreno da teoria, do abstrato. Um instrumento de medida já é uma operação concreta, empírica. A passagem do terreno abstrato para o concreto é precisamente viabilizada pelas definições operacionais dos construtos. Este é, talvez, o momento mais crítico na construção de medidas psicológicas, pois é aqui que se fundamenta a validade desses instrumentos; é aqui que se baseia a legitimidade da representação empírica, comportamental, dos traços latentes (os construtos). Duas preocupações são relevantes e decisivas neste momento: 1) as definições operacionais dos construtos devem ser realmente operacionais e 2) devem ser o mais abrangente possível.

Primeiramente, as definições operacionais devem ser realmente operacionais. Esta tautologia é proposital, porque se peca demais neste particular. Uma definição de um construto é operacional quando o mesmo é definido, não mais em termos de outros construtos, mas em termos de operações concretas, isto é, de comportamentos físicos através dos quais o tal construto se expressa. Assim, se inteligência verbal for definida como a "capacidade de compreender uma frase" ou, mesmo "compreender uma frase", estamos diante de uma definição constitutiva e não operacional. Isto porque compreender não é um comportamento, mas um construto. Seria uma definição operacional de compreensão da frase, reproduzir a frase com outras palavras. Mager (1981) dá uma fórmula simples e perfeita para decidir se a definição é ou não operacional. Ela é operacional se você puder dizer à pessoa: "vá e faça...". Assim, se inteligência verbal for definida como compreender uma frase, o que é que se deve pedir à pessoa para fazer, pois "vá e compreenda..." não lhe diz nada que ele possa fazer. Ao passo que dizer "vá e reproduza a frase" indica claramente o que a pessoa deve fazer, como deve se comportar, e, portanto, esta última é uma definição operacional, pois ela define comportamentos que devem ocorrer, enquanto compreender a frase não indica nenhum comportamento concreto específico a ser exibido por parte da pessoa.

Em segundo lugar, a definição operacional deve ser o mais abrangente possível do construto. Nenhuma definição operacional esgota a amplitude semântica de um construto; assim, podem haver definições operacionais mais ou menos abrangentes do mesmo construto e essa grandeza de abrangência, evidentemente, fala da boa, má ou pior qualidade da definição operacional, o que vai obviamente repercutir sobre o instrumento de medida do construto, que será baseado nesta definição operacional. Aliás, uma definição operacional pode ser perfeitamente operacional e também equivocada ou errada, quando esta não cobrir nada do espaço semântico próprio do construto. Assim definir inteligência verbal como "desenhar círculos na areia" constitui uma definição perfeitamente operacional, pois todo o mundo entende quando se manda desenhar círculos na areia; contudo, apesar de operacional, ela é uma definição perfeitamente equivocada de inteligência verbal, pois o comportamento de desenhar círculos na areia não tem nada a ver com o construto em questão. Disto segue que as definições operacionais podem representar um construto numa escala que expressa uma proporção de coincidência entre construto e definição operacional que vai de 0 a 1; sendo 0, quando a definição não cobre nada do construto e 1, quando ela cobre 100% do espaço

semântico do construto. Como já dissemos, cobrir 100% do construto nenhuma definição operacional será capaz, mas quanto maior covariância existir entre construto e definição operacional, maior qualidade deve-se atribuir a esta definição do construto e, por consequência, maior chance terá o instrumento, que de tal definição resulta ser superior em qualidade. Dizemos maior chance, porque a qualidade do instrumento não depende unicamente de boas definições operacionais, embora sem a boa qualidade destas o instrumento já comece, de saída, a ser inferior.

Para garantir melhor cobertura do construto, as definições operacionais deverão especificar e elencar aquelas categorias de comportamentos, que seriam a representação comportamental do construto. Quanto melhor e mais completa for esta especificação, melhor será a garantia de que o instrumento que resultar para a medida do construto será válido e útil. Por exemplo, quais seriam as categorias de comportamentos que expressariam comportamentalmente a compreensão verbal? Seriam tais como: reproduzir texto, dar sinônimos e antônimos, explicar o texto, sublinhar alternativas, etc. Quanto mais completa esta listagem de categorias comportamentais, mais próximos estamos da construção do instrumento, porque o próximo passo será simplesmente expressar essas categorias em tarefas unitárias e específicas (os itens) e o instrumento piloto está construído. Por isso, nunca é demais gastar tempo na implementação detalhada das definições operacionais do construto.

Onde se inspirar para realizar adequadamente esta tarefa? Novamente, os métodos a serem utilizados para resolver o problema deste passo da construção de medidas psicológicas são a literatura pertinente sobre o construto, a opinião de peritos na área, a experiência do próprio pesquisador, bem como a análise de conteúdo do construto.

É bom lembrar neste contexto de que os instrumentos de medida psicológica visam medir traços latentes. Mas como medir traços latentes que são impérvios à observação empírica, que é o método da ciência? Estamos aqui nos defrontando com o problema da representação: qual é a maneira adequada de se representar esses atributos latentes para que possam ser cientificamente abordados? Embora o problema pareça, e é na verdade, grave, ele não é específico da Psicometria; ocorre na própria física com a teoria quântica, por exemplo. Como o comportamento representa estes traços latentes? É precisamente o problema que as definições operacionais precisam resolver.

A OPERACIONALIZAÇÃO DO CONSTRUTO

Este é o passo da construção dos itens, que são a expressão da representação comportamental do construto, a saber: as tarefas (os itens do instrumento) que as pessoas deverão executar para que se possa avaliar a magnitude de presença do construto (atributo).

1. Fontes dos Itens

Se os passos até aqui discutidos forem adequadamente resolvidos, estaremos agora diante das categorias comportamentais que expressam o construto de interesse, as quais dão praticamente a resposta à construção dos itens. Além disso, podemos apelar para outras duas fontes de itens: a entrevista e outros testes que medem o mesmo construto. A entrevista consiste em pedir as pessoas representantes da população para a qual se deseja construir o instrumento para opinarem em que tipo de comportamentos tal construto se manifesta. Por exemplo, se o desejo é construir um instrumento sobre assertividade, podemos nos dirigir a representantes da população e pedir "como é para você uma pessoa assertiva"? De uma

pesquisa dessa natureza pode surgir uma grande riqueza de comportamentos que expressam assertividade e que podem ser aproveitados como itens do instrumento. Ademais, podemos nos inspirar em itens que compõem outros instrumentos disponíveis no mercado e que medem o mesmo construto. Assim, temos três fontes preciosas para a construção dos itens:

- **literatura**: outros testes que medem o construto;
- **entrevista**: levantamento junto à população-meta;
- **categorias comportamentais**: definidas no passo das definições operacionais.

É importante notar que no processo de elaboração do instrumento como o temos exposto, os itens não são mais coletados a esmo ou "chutados", mas eles são elaborados ou, pelo menos, selecionados em função das definições operacionais de um construto, que foi exaustivamente analisado em seus fundamentos teóricos e nas evidências (dados) empíricas disponíveis. Então, não é qualquer item que pareça medir o construto que é aceito, mas somente aquele que corresponde às definições teóricas (constitutivas) e às suas definições operacionais. Não é mais a malfadada face validity que impera na seleção dos itens e sim a sua pertinência (nessa altura, obviamente, ainda teórica) ao contexto teórico do construto. Aliás, os itens não são selecionados ou "pescados", são construídos para representar comportamentalmente o construto de interesse.

2. Regras de Construção de Itens

Dadas as fontes que baseiam a construção dos itens, é preciso dar agora algumas regras ou critérios fundamentais para a elaboração adequada dos próprios itens. Essas regras aplicam-se, em parte, à construção de cada item individualmente, e em parte ao conjunto dos itens que medem um mesmo construto. Além disso, dependendo do tipo de traço a ser medido, se de aptidão ou de personalidade, algumas das regras aplicam-se e outras não.

a) Critérios para a construção dos itens

1. Critério comportamental: o item deve expressar um comportamento, não uma abstração ou construto. Segundo Mager (1981), o item deve poder permitir à pessoa uma ação clara e precisa, de sorte que se possa dizer a ele vá e faça. Assim reproduzir um texto é um item comportamental (vá e reproduza...), ao passo que compreender um texto não o é, pois a pessoa não sabe o que fazer com vá e compreenda....

2. Critério de objetividade ou de desejabilidade: para o caso de escalas de aptidão, os itens devem cobrir comportamentos de fato, permitindo uma resposta certa ou errada. O respondente deve poder mostrar se conhece a resposta ou se é capaz de executar a tarefa proposta. Ao contrário, para o caso das atitudes e da personalidade em geral, os itens devem cobrir comportamentos desejáveis (atitude) ou característicos (personalidade). Ele, neste caso, deve poder concordar ou discordar ou opinar sobre se tal comportamento convém ou não para ele, isto é, os itens devem expressar desejabilidade ou preferência. Não existem neste caso respostas certas ou erradas; existem sim diferentes gostos, preferências, sentimentos e modos de ser.

3. Critério da simplicidade: um item deve expressar uma única ideia. Itens que introduzem explicações de termos ou oferecem razões ou justificativas são normalmente confusos porque

introduzem ideias variadas e confundem o respondente. Por exemplo: Gosto de feijão porque é saudável. A pessoa pode de fato gostar de feijão mas não porque seja saudável; assim, ele não saberia como reagir a tal item: se porque o feijão é gostoso ou porque é saudável. O item exprime duas ideias.

4. Critério da clareza: o item deve ser inteligível até para o estrato mais baixo da população-meta; daí, utilizar frases curtas, com expressões simples e inequívocas. Frases longas e negativas incorrem facilmente na falta de clareza. Com referência às frases negativas: normalmente elas são mais confusas que as positivas; conseqüentemente, é melhor afirmar a negatividade do que negar uma afirmação. Por exemplo: fica mais inteligível dizer detesto ser interrompido do que não gosto de ser interrompido ou em vez de não me sinto feliz é melhor dizer sinto-me infeliz. Neste contexto, é preciso também não utilizar gírias, porque estas não são em geral inteligíveis para todos os membros de uma população-meta do instrumento, além de tipicamente ofender o estrato mais sofisticado da mesma população, o que pecaria contra o critério número 10. Contudo, o linguajar típico da população-meta deve ser utilizado na formulação dos itens; assim, são admissíveis e são mais apropriadas expressões conhecidas por tal população, ainda que elas possam parecer linguisticamente menos castiças. A preocupação aqui é a compreensão das frases (que representam tarefas a serem entendidas e se possível resolvidas), não sua elegância artística.

5. Critério da relevância (pertinência, saturação, unidimensionalidade, correspondência): a expressão (frase) deve ser consistente com o traço (atributo, fator, propriedade psicológica) definido e com as outras frases que cobrem o mesmo atributo. Isto é, o item não deve insinuar atributo diferente do definido. O critério diz respeito à saturação que o item tem com o construto, representada pela carga fatorial na análise fatorial e que constitui a covariância (correlação) entre o item e o fator (traço).

6. Critério da precisão: o item deve possuir uma posição definida no contínuo do atributo e ser distinto dos demais itens que cobrem o mesmo contínuo. Esse critério supõe que o item pode ser localizado em uma escala de estímulos; em termos de Thurstone, diríamos que o item deve ter uma posição escalar modal definida e um desvio-padrão reduzido. Em termos da Teoria da Resposta ao Item (TRI), este critério representa os parâmetros "b" (dificuldade) e "a" (discriminação) e pode realmente ser avaliado de forma definitiva apenas após coleta de dados empíricos sobre os itens.

7. Critério da variedade: dois aspectos especificam este critério:

I. variar a linguagem: uso dos mesmos termos em todos os itens confunde as frases e dificulta diferenciá-las, além de provocar monotonia, cansaço e aborrecimento. Exemplo: o EPPS (Edwards Personal Preference Schedule) começa quase todas as suas 500 frases com a expressão "I like...". Depois de tantos "I like", qualquer pessoa deve se sentir saturado!

II. no caso de escalas de preferências: formular a metade dos itens em termos favoráveis e metade em termos desfavoráveis, para evitar erro da resposta estereotipada à esquerda ou à direita da escala de resposta. É a recomendação que Likert já dava em 1932.

8. Critério da modalidade: formular frases com expressões de reação modal, isto é, não utilizar expressões extremadas, como excelente, miserável, etc. Assim, ninguém é infinitamente inteligente, mas a maioria é bastante inteligente. A intensidade da reação da pessoa é dada na escala de resposta. Se o próprio item já vem apresentado em forma

extremada, a resposta na escala de respostas já está viciada. Assim, se perguntar a pessoa se está pouco ou muito de acordo (numa escala, por exemplo, de 7 pontos que vai de desacordo total a acordo total), um item formulado extremado, tal como "meus pais são a melhor coisa do mundo", dificilmente receberia resposta 7 (totalmente de acordo) por parte da maioria das pessoas da população-meta, simplesmente porque a formulação é exagerada. Se em lugar dela fosse utilizada uma expressão mais modal, tal como "eu gosto dos meus pais", as chances de respostas mais variadas e inclusive extremadas (resposta 7) seriam de se esperar.

9. Critério da tipicidade: formar frases com expressões condizentes (típicas, próprias, inerentes) com o atributo. Assim, a beleza não é pesada, nem grossa, nem nojenta.

10. Critério da credibilidade (face validity): o item deve ser formulado de modo que não apareça como ridículo, despropositado ou infantil. Itens com esta última caracterização fazem o adulto sentir-se ofendido, irritado ou coisa similar. Enfim, a formulação do item pode contribuir e contribui (Nevo, 1985; Nevo e Sfez, 1985) para uma atitude desfavorável para com o teste e assim aumentar os erros (vieses) de resposta. Este tema, às vezes, é discutido sob o que se chama de validade aparente (face validity), que não tem nada a ver com a validade objetiva do teste, mas pode afetar negativamente a resposta ao teste, ao afetar o indivíduo respondente e, assim, indiretamente afetar a própria validade psicométrica do teste.

b) Critérios referentes ao conjunto dos itens (o instrumento todo)

11. Critério da amplitude: este critério afirma que o conjunto dos itens referentes ao mesmo atributo deve cobrir toda a extensão de magnitude do contínuo desse atributo. Critério novamente satisfeito pela análise da distribuição dos parâmetros "b" da TRI. A razão disso é que um instrumento deve poder discriminar entre indivíduos de diferentes níveis de magnitude do traço latente, inclusive diferenciar entre si os que possuem um traço alto, quanto entre os que possuem um traço pequeno, e não somente entre os de traço alto em relação aos de traço baixo.

12. Critério do equilíbrio: os itens do mesmo contínuo devem cobrir igual ou proporcionalmente todos os segmentos (setores) do contínuo, devendo haver, portanto, itens fáceis, e médios difíceis (para aptidões) ou fracos, moderados e extremos (no caso das atitudes). De fato, os itens devem distribuir-se sobre o contínuo numa disposição que se assemelha à da curva normal: maior parte dos itens de dificuldade mediana e diminuindo progressivamente em direção às caudas (itens fáceis e itens difíceis em número menor). A razão deste critério encontra-se no fato de que a grande maioria dos traços latentes se distribuem entre a população mais ou menos dentro da curva normal, isto é, a maioria das pessoas possuem magnitudes medianas dos traços latentes, sendo que uns poucos possuem magnitudes grandes e outros magnitudes pequenas.

3. Quantidade de Itens

Para se cobrir a totalidade ou a maior parte ou, pelo menos, grande parte da extensão semântica do construto, explicitada nas definições constitutivas, normalmente exige-se, no instrumento final, um número razoável de itens. O que é um número razoável? O bom senso de quem trabalha nessa área sugere que um construto, para ser bem representado, necessita de cerca de 20 itens. Há, evidentemente, construtos muito simples que raras vezes necessitam de tal número de itens, sendo suficientes apenas uma meia dúzia ou menos deles. Por exemplo, satisfação com o salário. Quantas maneiras há de se verificar tal satisfação? Parece exagerado

perguntar 20 vezes à pessoa se está satisfeita com o seu salário. Pode-se, sim, perguntar se ele está contente com a quantia, com o poder de compra, com a pontualidade de entrega, e alguns aspectos mais. Mas parece difícil descobrir 20 maneiras de estar satisfeito com o salário. Entretanto, a grande maioria dos traços latentes normalmente possuem uma diversidade bem maior de aspectos e, por isso, exigem maior número de itens.

Se o número final de itens, isto é, depois que o instrumento passou por todas as fases de construção e validação, deve ser em torno de 20, pergunta-se com quantos itens é preciso começar para que no final possamos salvar 20? A resposta dada no contexto da psicometria tradicional positivista é a de que se deve começar com, pelo menos, o triplo de itens para se poder assegurar, no final, um terço deles. Esta resposta deve-se ao modo positivista de construir instrumentos psicológicos. Neste enfoque, os itens não são construídos a partir de uma teoria; eles são coletados ou selecionados de um "pool of items" que parecem medir um dado construto e, em seguida, analisados estatisticamente para ver quais deles se salvam. Quer dizer, os itens são aqui simplesmente "chutados"; são selecionados apenas porque parecem medir o que se quer medir.

Dentro da técnica de construção de instrumentos baseada na teoria dos traços latentes que estamos expondo, para se salvarem 20 itens no final de toda a elaboração e validação do instrumento, não é necessário iniciar com mais do que 10% de itens além dos 20 requeridos no instrumento final. Isto porque os itens incluídos no instrumento piloto são itens que possuem validade teórica real e não simplesmente parecem ter validade.

A ANÁLISE TEÓRICA DOS ITENS

Operacionalizado o construto através dos itens, estamos diante da hipótese de que estes representam adequadamente o tal construto. Essa é a nossa versão da hipótese a ser testada. Contudo, é importante avaliar tal hipótese em relação à opinião de outros para nos assegurarmos de que ela apresenta garantias de validade. Essa avaliação ou análise da hipótese (análise dos itens) é obviamente ainda teórica porque consiste simplesmente em pedir outras opiniões sobre a hipótese, sendo que esses outros, que a vão avaliar, ainda não são amostra representativa da população para a qual o instrumento foi construído. Essa análise teórica é feita por juízes e ela comporta dois tipos distintos de juízes, segundo se a análise incide sobre a compreensão dos itens (análise semântica) ou sobre a pertinência dos itens ao construto que representam (propriamente chamada de análise dos juízes). Assim, antes de partir para a validação final do instrumento piloto, este é submetido a uma análise teórica dos itens através da análise semântica e análise dos juízes.

1. Análise Semântica dos Itens

A análise semântica tem como objetivo precípua verificar se todos os itens são compreensíveis para todos os membros da população à qual o instrumento se destina. Nela, duas preocupações são relevantes: verificar se os itens são inteligíveis para o estrato mais baixo (de habilidade) da população-meta e, por isso, a amostra para essa análise deve ser feita com esse estrato; segundo, para evitar deselegância na formulação dos itens, a análise semântica deverá ser feita também com uma amostra mais sofisticada (de maior habilidade) da população-meta (para garantir a chamada "validade aparente" do teste). Entende-se por estrato mais baixo, aquele segmento da população-meta que apresenta menor nível de habilidades. Assim, por exemplo, se o teste se destina a uma população que congrega indivíduos do I grau de ensino até universitários, obviamente o estrato mais baixo neste

contexto são aqueles do I grau e o mais sofisticado será representado pelos de nível universitário. De qualquer forma, a dificuldade na compreensão dos itens não deve se constituir em fator complicador na resposta dos indivíduos, dado que não se quer medir a compreensão deles (a não ser, óbvio, que o teste queira medir precisamente isto), mas sim a magnitude do atributo a que os itens se referem. Que técnica utilizar para fazer essa análise? Há várias maneiras eficientes para tal tarefa, como por exemplo, aplicar o instrumento a uma amostra de aproximadamente 30 pessoas da população-meta e em seguida discutir com eles as dúvidas que os itens suscitarem. Entretanto, uma técnica que se tem mostrado mais eficaz na avaliação da compreensão dos itens consiste em checá-los com pequenos grupos de pessoas (3 ou 4) numa situação de "brainstorming". Essa técnica funciona da seguinte forma: constitui-se um grupo de até 4 pessoas, iniciando com sujeitos do estrato mais baixo da população-meta, porque se supõe que se tal estrato compreende os itens, a fortiori o estrato mais sofisticado também os compreenderá. A este grupo é apresentado item por item, pedindo que ele seja reproduzido pelos membros do grupo. Se a reprodução do item não deixar nenhuma dúvida, o item é corretamente compreendido. Se surgirem divergências na reprodução do item ou se o pesquisador se perceber que ele está sendo entendido diferentemente do que ele, pesquisador, julga que deveria ser entendido, tal item tem problemas. Dada esta situação, o pesquisador então explica ao grupo o que ele pretendia dizer com tal item. Normalmente, neste caso, as próprias pessoas do grupo irão sugerir como se deveria formular o item para expressar o que o pesquisador quer dizer com ele; e aí está o item reformulado como deve ser. Quantos grupos são necessários para proceder a esta análise semântica? Bem, itens que não ofereceram nenhuma dificuldade de compreensão em uma, no máximo duas, sessões, não necessitam de verificação ulterior. Itens que continuam apresentando dificuldades após, digamos, no máximo cinco sessões, devem ser simplesmente descartados. Em seguida a essas sessões, é importante pelo menos uma sessão de verificação dos itens com um grupo de indivíduos mais sofisticados. O objetivo desta verificação consiste em evitar que os itens se apresentem demasiadamente primitivos para tais pessoas e assim perderem a validade aparente. É que os itens devem também dar a impressão de seriedade, como diz o ditado de que a mulher de César não somente deve ser honesta, mas deve também parecer honesta! (veja regra número 10 dos critérios de construção de itens).

2. Análise dos juízes

Esta análise é, às vezes, chamada de análise de conteúdo, mas propriamente deve ser chamada de análise de construto, dado que precisamente procura verificar a adequação da representação comportamental do(s) atributo(s) latente(s).

Nessa análise, os juízes devem ser peritos na área do construto, pois sua tarefa consiste em ajuizar se os itens estão se referindo ou não ao traço em questão. Uma tabela de dupla entrada, com os itens arrolados na margem esquerda e os traços no cabeçalho, serve para coletar essa informação. Uma concordância de, pelo menos, 80% entre os juízes pode servir de critério de decisão sobre a pertinência do item ao traço a que teoricamente se refere.

A técnica exige que se dê aos juízes duas tabelas: uma com as definições constitutivas dos construtos/fatores para os quais se criaram os itens e outra tabela de dupla entrada com os fatores e os itens, em que são avaliados os itens que medem os dois fatores de raciocínio verbal (compreensão verbal e fluência verbal). Normalmente, é necessária uma terceira tabela que elenca os itens, uma vez que a tabela de dupla entrada geralmente não comporta a expressão completa do conteúdo dos itens.

Com base nessas tabelas, a função dos juízes consiste em colocar um X para o item sob o fator ao qual o juiz julga o item se referir. Um número de 6 juízes será suficiente para realizar esta tarefa. Itens que não atingirem uma concordância de aplicação aos fatores (cerca de 80%) obviamente apresentam problemas e seria o caso de descartá-los do instrumento-piloto. Isso vale, contudo, se o construto, para o qual está sendo construído o teste, apresentar fatores (particularmente quando forem em maior quantidade) que se supõem ou se sabe que não são correlacionados. Quando se supõem que os fatores sejam correlacionados, acontece que uma mesma tarefa (item) pode se referir, certamente com níveis de saturação diferente; mas de fato se referir simultaneamente a mais de um fator, o que implicaria que os juízes iriam mostrar alguma discordância quanto à aplicação do item a este ou a aquele fator. Neste caso, esta discordância deve ser considerada como concordância. Uma outra solução seria instruir os juízes a marcarem, para cada item, não o fator mas aqueles fatores aos quais o item se refere. Entretanto, com tal dica, abre-se campo para muita divagação por parte dos juízes e, assim, arriscar perder-se a utilidade prática dessa análise. Seria melhor instruir os juízes para colocarem, se possível, cada item sob um fator somente.

Com o trabalho dos juízes, ficam completados os procedimentos teóricos na construção do instrumento de medida, os quais comportaram a explicitação da teoria do(s) construto(s) envolvido(s), bem como a elaboração do instrumento piloto, que constitui a representação comportamental desses mesmos construtos e que se põe como a hipótese a ser empiricamente testada (validação do instrumento), tarefa que será iniciada com os procedimentos experimentais, os quais consistem em coletar informação empírica válida e submetê-la às análises estatísticas pertinentes em Psicometria.

REFERÊNCIAS

Mager, R.F. – *Medindo os Objetivos de Ensino ou "Conseguiu um Par Adequado"*. Porto Alegre: Editora Globo, 1981.

Nevo, B. & Sfez, J. – Examinees' Feedback Questionnaires. *Assessment and Evaluation in Higher Education* 10:236-249,1985.

Nevo, B. – Face Validity Revisited. *Journal of Educational Measurement* 22:287-293,1985.

Pasquali, L. (org.) – *Elaboração de Instrumentos Psicológicos*. São Paulo: Casa do Psicólogo, no prelo.

PASQUALI, L. Princípios de elaboração de escalas psicológicas. *Revista de Psiquiatria Clínica, São Paulo*, v. 25, n. 5, p. 206-213, 1998. Edição Especial.