

Laboratório de Psicologia Experimental
Departamento de Psicologia – UFSJ
Disciplina: Método de Pesquisa Quantitativa

TEXTO 9. DEFINIÇÃO DAS VARIÁVEIS E MÉTODOS DE COLETA DE DADOS

AUTORA: Prof. Marina Bandeira, Ph.D.

Após realizada a planificação operacional da pesquisa é preciso ainda fazer a classificação e definição das variáveis, é preciso também determinar qual será o método de coleta de dados a ser empregado e, finalmente, é preciso escolher o instrumento de medida a ser utilizado.

Todo projeto de pesquisa inclui, na formulação de seu problema, alguns conceitos teóricos. Estes conceitos devem ser claramente definidos. Para isto, torna-se necessário operacionalizar as definições, ou seja, passar as definições do mundo conceitual para o mundo empírico.

As variáveis a serem definidas são as variáveis independente e dependente, assim como as variáveis que serão controladas. Isto dará uma idéia melhor dos pontos fortes da pesquisa. Além disso, os vieses ou variáveis espúrias que podem estar interferindo com os efeitos da variável independente devem ser identificados e explicados. Assim, podemos ter a exata noção dos pontos fortes e fracos da pesquisa, colocando em perspectiva o alcance dos resultados.

A definição operacional das variáveis se faz através da indicação pormenorizada e empírica dos procedimentos necessários para medir a variável. Trata-se, portanto, de especificar como cada variável será medida. Por exemplo, a variável idade será definida como a diferença entre a data do nascimento e a data da coleta de dados. Se estivermos estudando o efeito de um tratamento do alcoolismo, podemos querer incluir a variável concernente à quantidade de bebida ingerida pelos sujeitos. Neste caso, a variável “quantidade de bebida ingerida” será definida operacionalmente em termos de “ml de bebida ingerida por dia”.

Se realizarmos uma pesquisa para verificar o efeito da ingestão de álcool no número de acidentes de carro, é possível imaginar várias definições operacionais diferentes da variável independente (ingestão de álcool). Por exemplo, temos as possibilidades abaixo, descritas por Contandriopoulos et al. (1994):

1. Podemos definir operacionalmente ingestão de álcool como sendo o número de copos de álcool consumidos nas 4 horas que precederam o acidente. Neste caso, só podemos medir esta variável perguntando ao motorista e preenchendo um questionário a respeito da ingestão de álcool.
2. Nossa definição operacional de ingestão de álcool pode também ser a concentração de álcool no corpo, tal como medida pelo teste do bafômetro.

3. Uma outra definição operacional da nossa variável independente poderia ser a capacidade do sujeito de andar sobre uma linha reta. Neste caso, utilizaríamos a observação e submeteríamos os sujeitos a este tipo de tarefa, tendo assim uma medida do seu comportamento.

4. Finalmente, poderíamos definir operacionalmente nossa variável independente pela taxa de álcool no sangue, através de um exame de laboratório. Teríamos assim uma medida de laboratório.

Podemos verificar que os diferentes exemplos acima determinam definições operacionais que implicam em diferentes métodos de coleta de dados e em diferentes níveis de qualidade dos instrumentos de medida. Assim, um exame laboratorial da concentração de álcool no sangue constitui uma medida de maior qualidade do que o primeiro exemplo, que implica em perguntar ao motorista o quanto ele bebeu. Em geral, duas considerações determinam o tipo de definição operacional a ser adotada: a qualidade do instrumento de medida e a viabilidade do método de coleta de dados.

Ao decidirmos sobre nosso instrumento de medida e o método de coleta de dados, devemos adotar sempre, dentre as medidas viáveis, aquela que implica em menor grau de reatividade, maior validade e maior fidedignidade. Assim, garantimos a qualidade das medidas que adotarmos. É preciso ainda indicar quais foram as pesquisas que validaram estes instrumentos e os seus resultados de fidedignidade e validade.

MÉTODOS DE COLETA DE DADOS

No projeto de pesquisa, o pesquisador deverá descrever detalhadamente o método que usará para coletar seus dados. Basicamente, ele pode adotar como método de coleta de dados a utilização de documentos, a observação de comportamentos ou a informação dada pelo próprio sujeito, seja oralmente (entrevistas) ou de forma escrita (questionários auto administrados). Abaixo discutiremos estes métodos de coleta de dados.

1. UTILIZAÇÃO DE DOCUMENTOS:

Trata-se de documentos escritos, oficiais ou pessoais. No caso de documentos oficiais, podemos utilizar, por exemplo, os prontuários de hospitais psiquiátricos se quisermos pesquisar as características clínicas dos pacientes, que freqüentaram a emergência no último ano ou que foram internados. Podemos ainda consultar as fichas de registro de acidentes em alguma organização do governo que registra este tipo de dados. Este tipo de documentos é muito utilizado em pesquisas que estudam o funcionamento de organizações. No caso de documentos pessoais, podem ser utilizados, por exemplo, os diários ou correspondências dos sujeitos. Mencionamos documentos que se resumem a textos, mas os documentos podem também ser visuais, tais como desenhos, pinturas ou documentos sonoros, tais como cassetes ou discos.

2. OBSERVAÇÃO:

Há vários tipos de observação. Se adotarmos este método de coleta de dados, é preciso especificar qual o tipo de observação adotaremos em nossa pesquisa.

2.1. Observação Sistemática: Este tipo de observação gera dados mais fidedignos, pois implica em uma distância maior entre o pesquisador e o fenômeno a ser observado. Neste caso, o instrumento de medida consiste em uma grade de observação, na qual estarão indicados quais elementos devemos observar. A grade de observação permite observar os comportamentos de todos os sujeitos da mesma maneira. Estas características garantem a padronização do procedimento de coleta e, conseqüentemente, a reprodutibilidade e a objetividade dos dados. Por exemplo, em uma pesquisa que foi realizada com o objetivo de avaliar a segurança ao volante de motoristas no seu primeiro ano de conduta, descrito por Contradopoulos et al. (1990), a observação sistemática foi feita, de forma padronizada para todos, da seguinte maneira:

- o observador se sentava no banco detrás do carro
- o circuito era semelhante para todos os motoristas observados
- as condições do tráfico eram pré- determinadas
- as condições climáticas eram pré- determinadas
- as medidas tomadas foram:
 - utilização dos retrovisores
 - utilização das setas luminosas
 - utilização dos acessórios de segurança
 - respeito dos sinais.

A observação sistemática pode ser realizada em diferentes contextos: natural, quase-experimental ou experimental. Estas três condições variam em grau de reatividade da medida, em ordem crescente.

O contexto natural produz a medida menos reativa, pois podemos observar os sujeitos sem que eles sejam perturbados ou até sem que eles saibam que estão sendo observados. Por exemplo, se queremos observar a frequência do uso do cinto de segurança em carros, podemos colocar observadores em lugar estratégicos para isso.

O contexto quase-experimental implica na observação de sujeitos cujo tratamento foi feito por outros. Já implica em medidas mais reativas que a mencionada acima.

O contexto experimental, finalmente, implica na observação de sujeitos que o próprio observador trata ou investiga. Este tipo de medida pode incluir ainda mais vieses, pois a expectativa do experimentador pode enviesar os dados obtidos.

2.2. Observação em entrevista: O pesquisador, neste tipo de observação, está menos distante dos sujeitos estudados. Neste caso, no decorrer da entrevista, ele pode observar diversos aspectos do comportamento humano, dependendo dos interesses e objetivos da pesquisa. Ele pode observar, por exemplo, a aparência física (cuidados com o corpo e a higiene), comportamentos não-verbais (ex. expressões faciais, gestos) ou paralingüísticos (tom de voz, entonação, latência das respostas), comportamentos verbais (estrutura da fala, formulação das frases). Pode observar ainda elementos interacionais, tais como as reações do sujeito às intervenções do entrevistador. É usada, por exemplo, em pesquisas onde há necessidade de fazer diagnósticos clínicos.

2.3. Observação Participante: Este é o tipo de observação que implica em dados mais reativos porque o pesquisador está menos distante dos seus sujeitos. Ele interage plenamente com os

sujeitos, sem que haja distinção entre sujeito-pesquisador. Não há uma grade de observação, mas apenas pontos gerais que guiam a atenção do pesquisador. Ele não anota seus dados no momento da observação, mas apenas posteriormente, usando de sua memória. Ele focaliza mais os aspectos qualitativos do que quantitativos referentes ao fenômeno sob estudo. Por exemplo, nos estudos antropológicos, este tipo de observação é muito utilizado. Neste caso, o pesquisador pode, por exemplo, desejar observar os hábitos de uma comunidade e para isto ele se integra na vida cotidiana desta comunidade.

2.4. Observação Livre: Quando o pesquisador aborda a situação, sem nenhuma grade de observação, ele está fazendo uma observação livre. É utilizada, como o método acima, pelos sociólogos e antropólogos.

Nestes dois últimos tipos de observação, a única maneira de controlar a reatividade das medidas é quando os sujeitos não tomam conhecimento de que estão sendo observados. Entretanto, esta prática coloca um problema ético delicado.

3. INFORMAÇÃO DADA PELO PRÓPRIO SUJEITO:

Quando a informação que queremos coletar não pode ser observada e só temos acesso a ela interrogando o próprio sujeito, podemos colher informações orais ou escritas dos sujeitos. No caso de colhermos informações orais, o faremos através de entrevistas e questionários preenchidos pelo experimentador. Este método de coleta de dados é mais indicado quando os sujeitos são impossibilitados de escrever de maneira fidedigna, tais como os analfabetos, pessoas idosas ou jovens demais, pessoas com necessidades especiais. Uma vantagem deste método oral é que a taxa de respostas dos sujeitos é mais alta.

No caso de colhermos informações escritas, utilizamos questionários auto-administráveis, ou seja, o sujeito lê e responde sozinho, sem precisar de um entrevistador. Podemos entregar os questionários pessoalmente aos sujeitos ou até enviá-los pelo correio, para que eles os retornem após terem sido preenchidos. Neste caso, podemos incluir, dentro da correspondência, um envelope pré-selado e já endereçado para facilitar o retorno. Este método pode ser usado quando se trata de sujeitos alfabetizados, capazes de responderem sozinhos e quando as respostas são simples. O problema deste método de coleta de dados é que, muitas vezes, os sujeitos não preenchem completamente ou não devolvem o questionário. Então, a taxa de respostas é baixa.

Quando utilizamos as entrevistas como método de coleta de dados, estas podem ser de vários tipos. Veremos a seguir dois tipos básicos de entrevistas: as entrevistas livres e as entrevistas dirigidas.

3.1. Entrevistas Livres:

São entrevistas menos estruturadas ou limitadas por parâmetros previamente estabelecidos. Elas podem ser de três tipos:

Entrevista de Exploração: Este tipo de entrevista é adequado quando se está iniciando um programa amplo de pesquisa, pois ela visa levantar todas as variáveis possíveis que possam ter alguma importância para o problema investigado. Neste tipo de entrevista, o pesquisador pede ao sujeito para se expressar livremente sobre a questão que está sendo

investigada (sua variável dependente). Ele estimula as verbalizações do sujeito para ter o máximo de informações e anota todos os dados. É utilizada, por exemplo, em pesquisas que visam fazer a adaptação transcultural de uma escala de medida ou teste, através de entrevistas de exploração com uma Comissão de Especialistas e com Grupos da população-alvo, levantando aspectos referentes ao tema avaliado pela escala (ex.depressão).

Entrevista com um Informante: Neste tipo de entrevista, um informante de um grupo que ocupa uma função chave é escolhido para dar informações sobre o grupo ou sobre uma pessoa. O informante pode ser mais objetivo em suas respostas, uma vez que ele não responderá sobre sua própria pessoa. Ele dará informações sobre uma pessoa ou sobre um grupo do qual ele faz parte. É utilizada com frequência em psiquiatria, quando se aplica, por exemplo, uma escala de medida do funcionamento de pacientes psiquiátricos, em entrevistas com a enfermeira que conhece mais o paciente.

Entrevista Clínica: Neste caso, o objetivo da entrevista é clínico, ou seja, focada nos sintomas dos sujeitos. Pode ser utilizada para verificar diagnósticos psiquiátricos, por exemplo, em pesquisas que visam validar uma escala (ex.escala de depressão de Beck). Os resultados da escala são comparados com os diagnósticos de um psiquiatra, feitos na entrevista. As pessoas que recebem escores da escala, acima de um determinado valor (ex. acima de 22) devem ser as mesma que receberam o diagnóstico dado pelo psiquiatra (de depressão)..

3.2 . Entrevistas Dirigidas:

Além das entrevistas livres, há ainda as entrevistas dirigidas, ou seja, a informação colhida durante a entrevista é toda definida previamente. Estas entrevistas dirigidas podem ser de dois tipos, descritos abaixo.

Entrevista semi-estruturada: Neste caso, há um roteiro de perguntas pré-formuladas, mas elas não são exclusivas. O entrevistador pode acrescentar novas perguntas, durante a entrevista, que ele achar necessárias para aprofundar mais as informações colhidas dos sujeitos. Isto ocorre em algumas escalas psiquiátricas (ex.escala de qualidade de vida para pacientes com esquizofrenia), onde o entrevistador faz perguntas adicionais ao paciente para colher informações sobre alguns aspectos de sua vida (ex. vida social) e, em seguida, com base nestas informações, ele dá uma cota de adequação do sujeito nestes aspectos (ex. adequação da vida social). Esta cota pode ser dada, por exemplo, em uma escala de 1 a 5, na qual 1 significa pouca adequação e 5 muita adequação.

Entrevista dirigida com questionário: Neste caso, as perguntas já estão decididas anteriormente, em um questionário estruturado. O entrevistador apenas lê as perguntas e anota as respostas dos sujeitos, sem acrescentar nenhuma nova pergunta, durante a entrevista.

QUALIDADE DOS INSTRUMENTOS DE MEDIDA

Os instrumentos de medida escolhidos para a coleta de dados devem ser especificados e suas qualidades psicométricas devem ser descritas. As qualidades psicométricas de um instrumento de medida se referem ao seu grau de fidedignidade e à sua validade.

FIDEDIGNIDADE

É a capacidade do instrumento de medida de medir fielmente um fenômeno, ou seja, se refere à reprodutividade dos seus resultados. Ou seja, é a sua capacidade de reproduzir o mesmo resultado, de forma consistente no tempo e no espaço. Por exemplo, no caso de dois observadores utilizarem grades de comportamento para fazer observações de um sujeito, a fidedignidade deste instrumento de medida é a capacidade de se obter dados semelhantes coletados por eles sobre os mesmos comportamentos. Outro exemplo de fidedignidade seria a capacidade de se obter dados semelhantes coletados pelo mesmo observador em momentos diferentes, ou seja, os resultados serão os mesmos. A reprodutividade se refere também à homogeneidade dos itens do instrumento, ou seja, eles medem o mesmo fenômeno. A fidedignidade de um instrumento de medida inclui três qualidades psicométricas: 1. Estabilidade Temporal, 2. Acordo entre juizes ou observadores e 3. Consistência interna. Estas propriedades podem ser avaliadas através de três métodos básicos descritos abaixo:

1. Estabilidade Temporal: Esta qualidade psicométrica pode ser avaliada a partir do procedimento denominado Teste-Reteste, no qual se aplica um mesmo instrumento de medida em dois momentos diferentes, em uma mesma amostra de sujeitos e verifica-se o grau de concordância entre os resultados destas duas aplicações. Geralmente, um intervalo de 2 a 3 semanas é adotado entre o teste e o reteste do instrumento, ou seja, entre a primeira e a segunda aplicações do mesmo teste. O intervalo não deve ser longo demais, para evitar que ocorram eventos que possam afetar o comportamento avaliado. O intervalo também não deve ser curto demais, para evitar que o sujeito responda em função da sua lembrança do que respondeu na primeira aplicação. Para verificar o grau de concordância entre estas duas aplicações, utiliza-se uma análise estatística de correlação entre os escores do teste e do re-teste. Se os resultados do teste e de re-teste apresentarem uma correlação significativa, temos uma indicação de que este teste é fidedigno, em termos de sua estabilidade temporal. Ou seja, ele mede a mesma coisa quando aplicado em dois momentos diferentes em uma mesma amostra de sujeitos. Seus resultados, portanto, são replicáveis. A análise correlacional considerada mais pertinente neste caso é a correlação intra-classe (intraclass correlation coefficient – ICC).

2. Acordo inter-juizes: A qualidade psicométrica de acordo inter-juizes consiste na reprodutividade dos resultados de uma escala quando aplicada por dois ou mais observadores ou juizes. Após a aplicação da escala pelos juizes, os escores são comparados, verificando-se sua equivalência. Caso haja um alto grau de acordo entre as medidas dos observadores, podemos considerar que o instrumento de medida é fidedigno.

Neste caso, um critério deve ser adotado para se determinar o grau de acordo obtido entre os observadores. Um dos critérios é a porcentagem de acordo obtido entre eles, tendo sido adotado um mínimo de 80% à 85% como necessário para se considerar que a fidedignidade do

instrumento é satisfatória. Outra análise estatística também utilizada é o cálculo da correlação entres os escores dos observadores, devendo-se obter uma correlação significativa. Finalmente, tem sido usado o coeficiente Kappa para se analisar estatisticamente o grau de acordo entre os escores dos observadores. A vantagem do coeficiente kappa é que esta análise leva em consideração a possibilidade de haver um acordo entre os escores, baseado no acaso. Portanto, ao utilizarmos este coeficiente podemos estar mais seguros de que o acordo encontrado não é contaminado pela ação do simples acaso, mas sim um acordo real entre os observadores.

3.Consistência Interna: Quando o instrumento é composto de vários elementos (ex. uma escala composta de várias questões ou itens), pode-se medir a consistência ou homogeneidade de seus elementos, verificando-se o grau de concordância entre eles. Há três métodos utilizados para verificar a consistência interna de uma escala de medida. Um deles é calcular o grau de correlação entre as duas metades de uma mesma escala, devendo haver uma correlação significativa entre os escores médios das duas metades. Um segundo método é calcular a correlação entre os itens ímpares e os itens pares de uma mesma escala, devendo haver uma correlação significativa entre os escores médios destes dois grupos de itens. O terceiro método, que é o mais utilizado, consiste em fazer uma análise estatística através do coeficiente alfa de Cronbach. Este método serve para calcular o grau de homogeneidade ou semelhança dos diversos itens ou questões de uma mesma escala, através de um programa informatizado de estatística (ex.SPSS). Quando a escala avalia apenas uma única dimensão ou fator de um fenômeno, obtemos a consistência interna da escala como um todo. Quando a escala avalia diversas dimensões ou fatores de um fenômeno, calculamos a consistência interna das questões que medem cada dimensão ou fator, separadamente. O valor do coeficiente alfa varia de 0 a 1,0. Uma consistência interna adequada se situa entre 0,75 e 0,85, pois este valor indica que as questões são semelhantes ou homogêneas sem serem redundantes. O valor de alfa mais adequado depende do número de itens de cada escala. Veja a tabela em anexo. Uma escala possui consistência interna, portanto, quando suas questões ou itens são homogêneos.

VALIDADE

A validade de um instrumento de medida se refere à sua capacidade de medir o fenômeno ou construto que ele pretende medir (e não outro construto). Por exemplo, uma escala de medida da depressão deve ser capaz de avaliar diferentes níveis deste construto (depressão) e não de outro construto (ex.fobia). A validade de um instrumento de medida depende da adequação entre o conteúdo dos seus itens e o construto teórico que o instrumento pretende medir. Esta adequação depende do grau de operacionalização do construto na elaboração dos itens e depende ainda da relação entre os resultados deste instrumento e variáveis empíricas relacionadas a ele, tal como previsto na teoria do construto avaliado. A validade de uma escala de medida só fica estabelecida após muitos anos de utilização deste instrumento, pois trata-se de algo difícil e sutil de ser avaliado. Há três tipos básicos de validade:

1.Validade de Conteúdo

Consiste na capacidade do instrumento de medir todas as facetas, componentes ou aspectos do construto ou conceito. Implica em avaliar a coerência aparente entre o conceito e o instrumento. Quando uma escala possui validade de conteúdo, suas questões abrangem todos os

aspectos do fenômeno estudado. Por exemplo, se uma escala está avaliando o grau de depressão, ela teria validade de conteúdo se incluísse questões que avaliam todos os aspectos da depressão: somáticos, cognitivos, comportamentais, etc. Ou seja, a escala seria completa, abrangente. Além disso, ela não pode abranger conteúdos que não pertencem ao seu construto. O procedimento para se avaliar a validade de conteúdo de um teste envolve:

1. Consulta a uma Comissão de Especialistas: neste caso, o pesquisador deve reunir estes especialistas em grupo e submeter o teste à sua avaliação. Eles discutirão sobre as questões do teste, avaliando se os itens representam adequadamente todos os aspectos do construto e levantando, caso necessário, aspectos que ainda não tenham sido incluídos ou eliminando outros considerados irrelevantes. Aspectos do fenômeno que estejam teoricamente relacionados entre si e que não tivessem sido abordados são sugeridos e anotados pelo pesquisador para serem acrescentados ao teste. Por exemplo, no caso da depressão, psiquiatras e psicólogos formariam parte deste grupo de especialistas.
2. Consulta a um grupo da população-alvo: o pesquisador deve ainda reunir um grupo de pessoas que possuem a vivência do fenômeno que está sendo medido e realizar uma sessão de grupo focal. Por exemplo, no caso da validação de uma escala de depressão, deve-se formar um grupo de pessoas que sofrem de depressão. Nesta sessão, o pesquisador deve estimular os participantes a falarem de sua experiência de modo abrangente, incluindo todos os aspectos do fenômeno. Os aspectos levantados pelos participantes que ainda não tivessem sido abordados no teste, passariam então a ser incluídos sobre forma de novas questões elaboradas pelo pesquisador.

2. Validade de Critério

Consiste na capacidade do instrumento de medida em prever o resultado a ser obtido em um desempenho ou comportamento específico dos indivíduos, presente ou futuro. O desempenho do indivíduo torna-se o critério com relação ao qual a escala será comparada. É avaliado através do cálculo da correlação entre os escores do teste e uma outra medida, presente ou futura, de desempenho dos sujeitos, que chamamos de critério. Diferentes tipos de desempenho são utilizados: desempenho acadêmico, desempenho profissional, diagnóstico psiquiátrico, avaliações de pares. Quando se trata da correlação entre o instrumento de medida e um comportamento futuro, trata-se da validade preditiva. Por exemplo, a correlação entre um teste ou prova apresentada no vestibular e o sucesso no desempenho acadêmico das pessoas posteriormente na universidade. Outro exemplo: a correlação entre os escores da escala de otimismo (antes de entrar na universidade) e o desempenho acadêmico posterior. Ou: a correlação entre a escala de otimismo (aplicada antes de uma cirurgia) e a recuperação de pacientes após a cirurgia.

Quando se trata de uma correlação entre o instrumento de medida e um comportamento ou critério prático contemporâneo, chamamos este tipo de validade de validade concorrente ou simultânea. Por exemplo, a correlação entre os resultados de uma escala sobre a saúde mental de um grupo de pessoas e os resultados obtidos por estas mesmas pessoas em entrevistas clínicas realizadas por um psiquiatra para avaliar sua saúde mental.

3. Validade de Construto

Consiste na capacidade de relacionar adequadamente a variável operacional (ex. os itens da escala) com o construto ou conceito teórico (definição do conceito, com base na teoria). Consideramos que nosso instrumento de medida tem validade de construto quando está realmente medindo o conceito teórico que pretendemos medir. Os testes e escalas validados que são utilizados em Psicologia para medir inteligência, aptidões, interesses ou traços de personalidade, são instrumentos de medida que foram analisados para se estabelecer sua validade. Há diversas maneiras de analisar a validade de construto de um instrumento.

a) Validade nomológica ou teórica (Teste de Hipótese). Quando avaliamos a capacidade de um instrumento de medida de comprovar uma hipótese que se baseia em uma teoria, chamamos esta validade de validade nomológica ou teórica. Em certos casos, a hipótese pode prever que dois grupos distintos de indivíduos deveriam teoricamente ter resultados diferentes em uma determinada escala. Por exemplo, uma escala que avalia a Qualidade de Vida de pacientes psiquiátricos deveria teoricamente dar resultados diferentes, quando são aplicadas em dois grupos de pacientes: um grupo que recebe um tratamento melhor, multidisciplinar, global e integrado, comparativamente a um grupo que recebe apenas um tratamento farmacológico. Em outro exemplo, citado por Contandriopoulos et al. (1994), uma escala de capacidade pulmonar que realmente medisse esta capacidade, ou seja, que tivesse validade de construto, deveria diferenciar significativamente grupos que variam com relação ao fenômeno medido, por exemplo, grupos que diferem com relação à capacidade dos pulmões, como no caso de um grupo de fumantes comparativamente a um grupo de não-fumantes (seus escores na escala deveriam ser significativamente diferentes, pelo teste estatístico “t” de Student). Outra hipótese a ser testada se refere à validade de testes cujo construto depende de mudanças no desenvolvimento cognitivo, em função da idade. Neste caso, testa-se a hipótese de que os escores do teste vão ser significativamente diferentes quando aplicado-a em indivíduos de diferentes faixas etárias.

Quando testamos a hipótese de que a escala deve correlacionar com outras variáveis com as quais o construto deve estar teoricamente relacionado (ex. habilidades sociais e assertividade), devemos encontrar uma correlação significativa entre estas variáveis (validade convergente). Por outro lado, quando testamos a hipótese de que a escala não deve correlacionar com variáveis com as quais o construto não está teoricamente relacionado, devemos encontrar uma correlação que não seja significativa (validade discriminante). Pasquali (2003) denomina este tipo de validade como convergente-discriminante.

b) Validade de traço. Quando avaliamos a capacidade do instrumento de medida de correlacionar com um outro instrumento de medida que avalia um conceito teoricamente semelhante e que já foi validado anteriormente, chamamos este tipo de Validade de Traço (Contandriopoulos et al., 1994). Alguns autores a denominam Validade Concomitante. Neste caso, aplicamos os dois instrumentos de medida em uma mesma amostra de sujeitos e calculamos o grau de correlação dos dados obtidos. Se os resultados obtidos com o novo instrumento apresentarem uma correlação significativa com o antigo instrumento já validado, podemos considerar que o novo instrumento possui validade de traço ou validade concomitante. Ou seja, aumentamos a confiança de que ele estaria medindo o mesmo conceito medido pelo teste antigo, devido à correlação encontrada entre eles.

c) Validade fatorial. Um outro método utilizado para se avaliar a validade de construto de um instrumento de medida consiste em identificar sua estrutura dimensional, ou seja, identificar quantas dimensões do construto a escala está avaliando com os seus itens. A identificação da estrutura dimensional se faz estatisticamente através de uma Análise Fatorial. A

Análise Fatorial avalia em que grau os itens ou questões de uma escala estão reagrupados em torno de um único fator ou subescala (ou dimensão) ou em vários fatores ou subescalas (ou dimensões). Por exemplo, em que grau as questões da Escala de Satisfação dos Pacientes com os Serviços de Saúde Mental (Escala Satis-Br) estão medindo um único fator ou várias sub-dimensões ou sub-escalas da satisfação com os serviços. Por exemplo: satisfação com a estrutura física do serviço, satisfação com o tratamento, satisfação com a equipe, etc. Estas seriam então as sub-dimensões do construto global *satisfação com os serviços de saúde mental*. Dizemos que a escala é unifatorial ou multifatorial, dependendo do número de fatores ou dimensões que possui. As escalas multifatoriais constituem instrumentos mais sensíveis, capazes de identificar diferentes aspectos da satisfação e de correlacionar com outras variáveis mais facilmente, do que uma escala unifatorial.

Quando várias subescalas são identificadas, é recomendado investigar as correlações que existem entre elas e as correlações entre cada uma das subescalas com a escala global. Espera-se encontrar uma correlação mais elevada entre cada sub-escala e a escala global do que a correlação observada entre as sub-escalas. Este resultado seria indicativo da presença de um construto comum subjacente a estas subescalas. Este resultado indicaria a presença de um construto básico referente à satisfação geral ou global com os serviços de saúde mental.

Há uma diferença entre um questionário e uma escala de medida. Somente as escalas possuem propriedades psicométricas de validade e fidedignidade estabelecidas e, portanto, somente elas podem ser utilizadas para se medir o grau de um construto em uma amostra de sujeitos. Geralmente, as escalas possuem itens quantitativos com alternativas de escala dispostas em forma de uma escala Likert de 1 a 5 pontos. Por exemplo, se temos escala de medida da satisfação, poderemos ter 5 alternativas de resposta, variando de 1=muito insatisfeito até 5=muito satisfeito. Podemos calcular a média do grau de satisfação de um grupo de sujeitos e podemos também comparar esta média com a de outro grupo que nos interesse. Esta média seria calculada somando-se os pontos obtidos em cada item da escala e dividindo pelo número de itens.

Quando utilizamos um questionário, este não foi submetido a uma avaliação de validade e fidedignidade, pois não estamos avaliando um construto. Neste caso, não podemos medir o grau de um construto. Podemos apenas calcular a porcentagem de pessoas que responderam positivamente às questões do questionário ou não. Por exemplo, quantos por cento responderam que estão muito satisfeitos ou quantos por cento estão insatisfeitos. Sem realizarmos uma análise das propriedades psicométricas de um instrumento de medida, não podemos saber o que as suas questões estão medindo. Portanto, não basta elaborar questões de um questionário e supor que elas estão medindo o que pretendemos medir.

ADAPTAÇÃO TRANSCULTURAL DE UMA ESCALA: Quando se pretende utilizar uma escala que foi elaborada em outro país, para se fazer avaliações no Brasil, é necessário que esta escala seja primeiro adaptada e, em seguida, validada para o contexto brasileiro. A vantagem de se fazer isto é que podemos, mais tarde, comparar os dados obtidos aqui com os obtidos em outros países, usando a mesma escala. Por exemplo, pode-se avaliar se a taxa de depressão pós-parto no Brasil é mais elevada do que em outros países, usando-se a mesma escala de avaliação que se usa em outros países e, em seguida, pesquisar os fatores que estão associados a este problema. Para isto, é necessário que a escala seja equivalente à sua versão original, em termos do seu conteúdo, porém contendo palavras e situações que sejam pertinentes para nossa cultura (Bunchaft e Cavas, 2000).

Para fazer a adaptação transcultural de uma escala, a literatura internacional apresenta os procedimentos mais recomendados, de modo a se obter uma nova versão da escala, que seja equivalente à versão original (Guillemin, Bombardier e Beaton, 1993; Vallerand, 1989). Esta equivalência deve ocorrer em relação ao conteúdo da escala (equivalência semântica), às situações usadas nas questões da escala (equivalência experiencial), aos aspectos culturais (equivalência cultural), às alternativas de resposta para cada questão da escala (equivalência técnica) (Jorge, 2000) e à forma de aplicação (ex. entrevista face a face ou auto-aplicada) (Reichenheim e Moraes, 2007). Os procedimentos recomendados por Guillemin et al. (1993) para fazer a adaptação transcultural de uma escala são:

1. **Tradução:** um tradutor bilíngüe, cuja língua materna é o português, deve traduzir a escala, obtendo-se assim uma primeira versão brasileira. Quando possível, recomenda-se duas traduções por dois tradutores..
2. **Retradução:** uma outra pessoa bilíngüe, cuja língua materna é o inglês (no caso da escala ser de língua inglesa), deve então passar a primeira versão brasileira de novo para o inglês. Recomenda-se igualmente, quando possível, duas versões retraduzidas.
3. **Comissão de especialistas:** as diversas versões da escala são, então, submetidas a uma comissão de especialistas, composta por profissionais que têm conhecimento do tema da escala e que trabalham com pessoas apresentando o problema em questão. Os membros da comissão realizam duas tarefas:
 - a. Comparam a versão retraduzida para o inglês (ou versões) com a versão original em inglês (ou versões), juntamente com os tradutores. Caso haja palavras muito diferentes nestas duas formas que mudam o sentido das frases, isto indica que a primeira tradução continha erros que devem ser corrigidos. Os erros são, então, corrigidos nesta fase, na escala escrita em português.
 - b. Em seguida, os membros da Comissão discutem as questões referentes à adaptação transcultural de frases, palavras, exemplos, expressões verbais e situações da escala original, que necessitam ser adaptadas para o contexto brasileiro. Por exemplo, se a escala faz uso de uma situação comum no país de origem e que não ocorre frequentemente no Brasil, uma outra situação que ocorra frequentemente aqui (portanto culturalmente pertinente no nosso contexto) deve ser escolhida para substituí-la, desde que seja equivalente ao conteúdo da original. Frases e expressões verbais cotidianas devem ser escolhidas para formar as frases. Obtemos então uma versão preliminar da escala brasileira.
4. **Estudo piloto (Pré-teste):** A versão preliminar brasileira vai ser, então, aplicada em uma pequena amostra de pessoas da população-alvo, afim de se testar se a escala é de fácil compreensão, se as palavras escolhidas são de uso cotidiano, se as instruções estão claras, se as alternativas de resposta são fáceis de serem discriminadas. Para cada pessoa que responde à escala, procura-se identificar termos ou frases que ela considera pouco usual ou de difícil compreensão e busca-se encontrar outros substitutos, de melhor aceitação. A escala é modificada e, em seguida, aplicada a uma outra pessoa da população-

alvo, até que não haja mais problemas de compreensão na escala. Ela deve ser fácil para qualquer pessoa, independente de seu nível sócio-econômico e de escolaridade. Este procedimento foi denominado Técnica de Sondagem (Probe Technique)

Uma outra maneira de se proceder, no estudo piloto, é introduzir uma avaliação escrita das questões, de forma que o respondente classifique cada pergunta da escala em termos do seu grau de compreensão, em uma escala de alternativas (ex. de 1 a 4, onde 1 é muito confusa e 4 muito clara). Calcula-se, então o grau de compreensão da media do grupo, eliminando-se os itens que obtiveram baixa cotação. .

Pode-se ainda aplicar a versão brasileira e a versão original em um mesmo grupo de pessoas bilíngües e verificar se as respostas destes sujeitos apresentam uma correlação elevada, o que deveria ocorrer se as duas formas são equivalentes. Este último procedimento, entretanto, não garante que a formulação das questões seja de fácil compreensão para todos e um procedimento específico para isto deve ser feito, tal como descrito acima, perguntando-se aos sujeitos o seu grau de compreensão.

Referência Bibliográfica:

Contandriopoulos, Ap., Champagne, F., Potvin, L., Denis, J.L. E Bouyle, P. (1990) Savoir préparer une recherche. Montréal. Les presses de l'Université de Montréal.

Campbell, D.T., Stanley, J.C. (1979) Delineamentos experimentais e quase-experimentais de pesquisa. Editora da Universidade de São Paulo, SP.

Selltiz, C., Wrightsman, L.S., Cook, S.W. (1987) Métodos de pesquisa nas relações sociais. Editora da Universidade de São Paulo, SP.

Bunchaft G, Cavas CST. Sob medida: um guia sobre a elaboração de medidas do comportamento e suas aplicações. São Paulo, Vetor, 2002.

Guillemin F, Bombardier C, Beaton D. Cross-Cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology*, 46 (12): 1417-1432, 1993.

Jorge MR. Adaptação transcultural de instrumentos de pesquisa em saúde mental. In: Gorestein C, Andrade LHS, Zuardi AW, editors. Escalas de avaliação clínica em psiquiatria e psicofarmacologia. Lemos Editorial, p. 53-58, São Paulo, 2000

Pasquali (2003). *Psicometria. Teoria dos testes na Psicologia e na Educação*. Petrópolis: Editora Vozes.

Vallerand RJ. Vers une méthodologie de validation trans-culturelle de questionnaires psychologiques: implications pour la recherche en langue française. *Canadian Psychology*. 30 (4), 1989.

ANEXO

TABELA DE VALORES DE ALPHA DE CRONBACH

A tabela abaixo mostra o valor mínimo necessário para que o coeficiente alpha de Cronbach seja considerado adequado como índice de consistência interna da escala. Este valor mínimo varia em função do número de itens que a escala possui.

Valor mínimo de alpha	N. de itens da escala
0,32 para uma escala de	2 itens
0,42 para uma escala de	3 itens
0,50 para uma escala de	4 itens
0,55 para uma escala de	5 itens
0,70 para uma escala de	10 itens e mais

Referência:

Tabela estabelecida por Gulliksen, H. (1950). *Theory of Mental Test*, New York: John Wiley

Tabela retomada depois por Martinez, A.R. (1995). *Psicometria: teoria de los tests psicológicos y educativos*. Madrid: Editorial Sintesis.

Tabela citada por Caron, J., Mercier, C. e Tempier, R. (1997). *Santé Mentale au Québec*, XXII, 2, 195-2117.